The double life of probability

Dziurosz-Serafinowicz, Patryk

THE DOUBLE LIFE OF PROBABILITY

Cover designed by *studio Michał Sławiński
Cover illustration by Agnieszka Szewczuk
Printed in Poland

ISBN 978-90-367-9301-8
ISBN 978-90-367-9300-1 (e-book)

university of groningen

The Double Life of Probability

A Philosophical Study of Chance and Credence

PhD thesis

to obtain the degree of PhD at the
University of Groningen
on the authority of the
Rector Magnificus Prof. E. Sterken
and in accordance with
the decision by the College of Deans.

This thesis will be defended in public on

Thursday 12 January 2017 at 14.30 hours

by

Patryk Dziurosz-Serafinowicz

born on 31 October 1982
in Złotów, Poland

Supervisor
Prof. J.W. Romeijn

Co-supervisor
Prof. A.J.M. Peijnenburg

Assessment committee
Prof. R. Pettigrew
Prof. J. Williamson
Prof. L. Wroński

Contents

Acknowledgments

Until about five years ago, I had only the most cursory knowledge of both philosophical and mathematical aspects of probability. Then, I was fortunate to meet exceptionally talented and friendly people who taught me a great deal about how to think about probabilities and encouraged me to write this thesis. Help received from these people incurs an incalculable debt.

First of all, I am indebted to Jan-Willem Romeijn, my supervisor. I have been extremely lucky to have a supervisor who cared so much about my work and who was incredibly generous with his time and ideas. No doubt, the biggest pieces of progress that I have made as a philosopher are owed to him. I thank him for allowing me to pick his brain, for helping me to appreciate patience, rigour, scrupulous intellectual care, and for much else. I have learned far more from him than could be effectively conveyed here.

I owe special thanks to Jeanne Peijnenburg, who gave generously of her time, knowledge, and advice. This thesis would have been considerably less well thought out and less well written without her help.

I am grateful to Richard Pettigrew, Jon Williamson, and Leszek Wroński—the members of the assessment committee—for their invaluable comments and advice. I have been helped immensely by discussions with Richard Pettigrew, whose ideas about epistemic utility theory profoundly influenced this thesis and whose warm hospitality helped to make my visit to the University of Bristol a memorable occasion.

My special thanks go to Ronnie Hermens for numerous stimulating discussions, his invaluable help with the Dutch translation of the summary, and for being a marvellous office companion.

For the last five years I have benefited enormously from discussions during the talks I gave for the PCCP, the GROLOG, and the Chance Encounter seminars at the University of Groningen, the Erasmus Graduate Conference in Philosophy of Science in Rotterdam, the Graduate Conferences in Theoretical Philosophy in Groningen and Nijmegen, the First International Conference of the German Society for Philosophy of Science in Hannover, the EPSA 13 in Helsinki, the ECAP 8 in Bucharest, the EPSA 15 in Düsseldorf, the Entia et Nomina conference in Kraków, and the Wednesday Colloquium in Fribourg.

I also want to express my appreciation to the exceptionally talented people who have shared the intellectual adventure of doing philosophy in and outside Groningen's Faculty of Philosophy, and whose comments helped to improve this thesis, particularly Miloud Belkoniene, Benjamin Bewersdorf, Max Bialek, Job de Grefte, Coos Engelsma, Leon Geerdink, Leah Henderson, Pavel Janda, Franziska Köder, Pieter van der Kolk, Sebastian Kołodziejczyk, Karolina Krzyżanowska, Louwe B. Kuijer, Jonathan Payne, Patricia Rich, Gianfranco Soldati (special thanks for encouragement), Tom Sterkenburg, Marta Sznajder, Sander Verhaegh, and Christian Wallmann.

The Faculty of Philosophy at the University of Groningen has been a wonderfully supportive place to work, with friendly and helpful colleagues. Knowing what I know now, I could not imagine a better place to work. I want to thank all members of this faculty for their support.

Finally, I am grateful with all my heart to my wife Dominika for her love and for believing in me, and to my parents for everything they have done for me.

Chapter 1

Introduction

1.1 The Main Themes

Probability lives a double life. There are at least two distinct concepts of probability, the *physical* and the *subjective* or *epistemic* one, that are frequently used in our scientific and everyday talk.[1] Consider the following examples of our talk of probabilities:

(1) "The probability of canals in the Netherlands being frozen in February is 90%."

(2) "For me, the probability of canals in the Netherlands being frozen in February is very high, and so I can't wait to go ice skating."

In both examples, we employ the concept of probability, but in two different ways. In (1), probability is physical or statistical, quantifying certain aspects of the physical world that make the canals very likely to get frozen. This concept of

[1] A note on terminology. Often the notion of *objective* probability is used to refer to non-epistemic probability. However, this use is confusing, for the adjective "objective" is also used to characterize some types of epistemic probability. For example, it is used to characterize a particular Bayesian interpretation of probability called *objective Bayesianism*. But, on this view, probability is epistemic: it is a degree of belief that is normatively constrained by the evidence one has. Also, the so-called *logical interpretation of probability* is regarded as an objective interpretation, i.e. probability is understood as a partial entailment relation between propositions that is meant to be independent of an agent's epistemic state and of any fact about the physical world.

probability is often called *chance*. In (2), probability refers to what we expect or believe about the canals in the Netherlands, and so it is about our epistemic attitude to the proposition that the canals will be frozen. This concept of probability is often called *credence* or *degree of belief*.

This thesis is a philosophical study of some of the core questions concerning chance and credence. In this study, the thesis introduces and utilizes various principles or conditions relating chances to credences, prior (or present) chances to posterior (or later) chances, and prior credences to posterior credences. The primary aim is to show that focusing and elaborating on these principles gives us a fruitful way of theorizing about chance and credence. The secondary aim is to show how we can combine these principles with some well-established argumentative strategies to provide insight into both notions of probability. Although the questions to be discussed belong to familiar territory within philosophy of probability, I believe that progress concerning these issues requires that we sometimes step back and try to look at them from a different perspective. The thesis offers such a perspective.

What are the principles that drive the discussion in the coming chapters? The bulk of this thesis explores so-called *chance-credence principles*.[2] They tell us roughly that rational agents should set their credences equal to chances. Principles of this sort are commonly regarded as rationality constraints on credences, often together with such principles as the Principle of Indifference (Keynes 1921), the requirement of Regularity (Shimony 1970), or Reflection Principles (van Fraassen 1984). But there is also a different, less often recognized understanding of these principles: they provide constraints on chances. Roughly speaking, they tell us that a quantity can be legitimately called chance if it constrains the agent's credence. The thesis exploits both understandings of chance-credence principles. In so doing, it shows how we can use these principles to tackle the following questions:

- Why should we think that chances behave as probability functions over some set of propositions?
- How can we show that information about chance conveyed by statistical

[2]In the literature, these principles are sometimes called *probability coordination principles*; see, e.g. Strevens (1999).

evidence is valuable for epistemically rational agents? In particular, is information about chance epistemically valuable for judges and jurors in legal settings?

Other principles used extensively in this thesis relate prior and posterior chances. By analogy to chance-credence principles, they may be called *chance-chance principles*. The first principle of this sort to be explored says that the prior chance of some proposition A conditional on the proposition about some posterior chance of A should be set equal to that posterior chance of A. The second one says that the prior chance of some proposition A should be equal to the weighted average of possible posterior chances of A, where the weights are chances assigned by the prior chance function to propositions about A's possible posterior chances. A substantial part of this thesis shows that these two principles are an essential feature of *resilient* or *stable* chance functions, i.e. it is shown that chances obeying these principles maximize resiliency under variation of experimental factors. These principles are also used to provide a viable approach to the following question:

- How should we model the evolution of chance through time? Or, how should chances at different points in time be related to each other?

The thesis also exploits a principle that relates prior and posterior credences. This principle, which may be understood as a kind of *credence-credence principle*, says that one's prior credence in proposition A conditional on the proposition about A's posterior credence should be equal to that posterior credence. By focusing on this principle, the thesis provides some insight into the following questions:

- How should rational agents update their credences? Is there a universal updating rule for credences? What sort of considerations could underpin our choice of updating rules for credences?

It is important to emphasize at the outset that a third concept of probability is often distinguished in the philosophy of probability, viz. the concept of *logical probability*. We use this concept, for example, when we say that:

(3) "The probability that the canals in the Netherlands will be frozen in February, given present evidence, is quite high."

Here it seems that probability refers neither to a physical fact nor to one's degree of belief. Rather, it is a matter of evidential or logical relation between propositions: the proposition that the canals in the Netherlands will be frozen in February and the proposition about our present evidence. For many authors, logical probability quantifies the degree to which evidence supports or undermines the hypothesis. And this degree of support is supposed to capture some objective relation between the evidence and the hypothesis. Traditionally, an analysis of this relation is a subject matter for inductive logic (Keynes 1921; Carnap 1962). The question, however, of how this objective relation should be understood is a matter of substantial controversy. It is not the goal of this thesis to contribute to the debate about the nature of logical probability. Nor is the aim to provide an understanding of this concept by studying the notion of credence. In particular, the question of whether the degree of support can be understood via some kind of suitably constrained credence, as some Bayesians want to suggest, will not be pursued in this thesis.

In the remainder of this introduction, I first describe the main argumentative strategies used in this thesis. Second, I situate the main themes of this thesis within both contemporary philosophy of chance and Bayesianism. Third, I present the basic presuppositions that pervade the thesis and remain the background against which the main ideas will be canvassed. Finally, I give an overview of the chapters.

1.2 The Main Argumentative Strategies

Along with the various principles aforementioned, the thesis also attempts to show how one can utilize some argumentative strategies either to argue for these principles or to argue from these principles in order to justify certain features of chance and credence. Two such strategies are extensively explored in this thesis. One is the argumentative strategy that was first used by Bruno de Finetti (De Finetti 1970) and then developed by James M. Joyce (1998), Hannes Leitgeb and Richard Pettigrew (2010b; 2010a), and Richard Pettigrew (2012; 2013c).

The basic methodological commitment behind this strategy is that we can justify various constraints (e.g. the norm of Probabilism, Bayes's rule, the Principle of Indifference, the Principal Principle) on the agent's credences by appealing to a fundamental norm for credences called *inaccuracy minimization*. With a suitable account of inaccuracy measure, this justification proceeds by arguing that (i) if the agent's credences fail to satisfy a given constraint, then they can be *dominated* by other credences that satisfy this constraint and that are closer to any truth-value assignment (or are less inaccurate), and (ii) if the agent's credences satisfy that constraint, then they cannot be so dominated. It is then concluded that credences violating that constraint are inadmissible with respect to the inaccuracy measure.

I develop a similar argumentative strategy in the case of chance, and argue that certain constraints or principles for chance follow from some fundamental norms governing chances, to wit, the norm of *weak predictive accuracy* and the norm of *maximizing resiliency*. More specifically, I show that (i) if a chance function fails to satisfy the axioms of finitely additive probability, then it can be dominated by a chance function that satisfies these axioms and is closer to any possible truth-value distribution, and (ii) if a chance function satisfies these axioms, then it cannot be so dominated. I then conclude that a chance function violating these axioms cannot be weakly predictively accurate, and so it cannot play the role of an expert for an agent's credences.

Similarly, I argue for two chance-chance principles by showing that (i) if a chance function does not obey them, then there is another chance function that satisfies them and is more resilient under variation of experimental factors, and (ii) if a chance function obeys them, then there is no other chance function that is more resilient in that way. I then conclude that any chance function which violates these principles is inadmissible with respect to a resiliency measure.

Another argumentative strategy that orients this thesis appeals to the well-known *value of learning theorem*, advanced by Leonard Savage (1954) and Irving J. Good (1967).[3] Generally speaking, this result says that the prior expected utility of making an informed decision is always at least as great as that of making

[3] In fact, a sketch of this result was first given by Frank P. Ramsey in his note entitled "Weight or the Value of Knowledge", published in Ramsey (1990).

an uninformed decision. I utilize the argumentative strategy behind this result (i) to show when learning ruled by the principle of maximum relative entropy leads to new credences that are expected to be helpful, and never harmful in making decisions, and (ii) to show that the subjective expected accuracy of verdicts at a legal trial cannot decrease when credences are expected to match chances.

1.3 Positioning the Main Themes

With some exceptions, most notably de Finetti[4], the double life of probability has long been recognized by philosophers, mathematicians, and statisticians. Famously, Ian Hacking claimed metaphorically that probability is Janus-faced—it is like Janus, a Roman god depicted as having two faces, one looking to the future and the other looking to the past. Hacking explained this metaphor by saying that:

> On the one side it [probability] 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)

Interestingly, even if the reigning orthodoxy throughout the development of probability theory during the seventeenth, eighteenth, and nineteenth century was the subjective concept of probability, an idea of physical probability was also presupposed in many important results obtained in that time. For example, in "An Essay Towards Solving A Problem in the Doctrine of Chances" (written around 1749 and published posthumously in 1763), Thomas Bayes posed the famous problem: how to calculate the probability that the *chance* of an event on a single trial lies in a certain interval of values, given frequencies in which this event happened in a finite number of trials? Importantly, in construing this

[4]De Finetti argued that the concept of physical or statistical probability in science is questionable, and thus it is unnecessary. He argued that the role of chance in science can be discharged by the notion of degree of belief. De Finetti's ingenious thought was that if your degrees of belief are exchangeable, then there is a unique representation of your degrees-of-belief distribution over the trials as an expectation of possible objective probability distributions, according to which the trials are independent and identically distributed. That is, whether or not you believe in the reality of chances, your degrees of belief act *as if* they were degrees of belief about possible chance distributions.

problem, Bayes assigned probabilities over hypotheses about unknown probability of the event in question, called by him chance. Bayes's idea was that before any trial, we have degrees of belief over a set of hypotheses stating different values of that chance, where each value lies in a certain interval. As it has been stressed out by some commentators, it seems plausible to think that while probabilities over these hypotheses refer to one's degrees of belief about some objective fact of the world, the hypotheses themselves represent objective facts about the true chances.[5] Moreover, as it has been suggested in Earman (1992, p. 52) and in Uffink (2011, p. 32), in providing his solution to the aforementioned problem, Bayes used a chance-credence principle. That is, he needed to determine the subjective probability that in a series of n independent trials some event occurs k times, given the supposition that the physical probability has a certain value. And his solution was that the subjective probability should be set equal to this value of physical probability.

Traditionally, the two faces of probability are subjects of two closely related research fields: philosophy of chance and Bayesianism. In what follows, I will situate the ideas of this thesis within these two fields. However, before doing this, I will touch upon David Lewis's idea of chance-credence relation—the relation that animates the bulk of this thesis.

1.3.1 David Lewis's Legacy

David Lewis in his "A Subjectivist's Guide to Objective Chance", the gemlike self-contained essay expounding the double life of probability, suggested an interesting reorientation of our attempts to scrutinize the concept of chance, and provided a stripe of Bayesian epistemology of credence.

In stating and defending his version of chance-credence principle, to wit, the Principal Principle, Lewis argued that it yields interesting consequences for both chance and credence. As to the notion of chance, Lewis (1986, p. 86) claimed that it captures "all we know about chance". He explored this idea by showing how the Principal Principle could be informative about chance by "being rich in

[5] This interpretation of Bayes's problem is to be found in Uffink (2011). Interestingly, he says about Bayes's contribution to the philosophy of probability that "by introducing probabilities of probabilities, his paper is really the first to invite the notion of objective chance!" (Uffink 2011, p. 34).

consequences that are central to our ordinary ways of thinking about chance" (Lewis 1986, p. 109). In particular, Lewis went on to show how we can use the Principal Principle to derive certain consequences for chance: that chance obeys the axioms of probability theory, that chance evolves through time by some sort of conditionalization, or that credences of frequencies can be inferred from credences about chance.

As to the notion of credence, Lewis (1986, p. 110) claimed that the Principle Principle "tells us something about what makes an initial credence function be a reasonable one". And though it does not tell us all about rational credence, it supplies our understanding of it by requiring one's rational credences to line up with chances. Viewed this way, the Principal Principle is regarded as an important component of what John Earman (1992, chapter 2) called *modern Bayesianism*, together with the requirement that rational credences should obey the axioms of probability (known also as Probabilism) and the requirement that rational agents should not assign zero credence to epistemically possible propositions (known also as Regularity).

Much of what is presented in this thesis is both a critical assessment and a development of Lewis's idea about the role of chance-credence principles in theorizing about both chance and credence. One major moral that the thesis attempts to convey is that much more than the Principal Principle informs our understanding of chance. In so doing, it also has a significant revisionary component: it discusses and revises Lewis's arguments for the claim that chances are probability functions and the claim that chances evolve by conditionalizing on the intervening histories of categorical-property instantiations. Lewis argued for these claims by appealing to his Principal Principle. The thesis develops alternative ways to vindicate these two claims by either reorienting the use of the Principal Principle or by invoking some plausible constraints on chance that are different from the Principal Principle.

Another significant use of Lewis's legacy concerns his stripe of Bayesianism. Although the thesis does not provide a systematic defence of Lewis's Bayesianism, it shows how its salient features can be developed to handle the problem of "naked" statistical evidence in the context of legal proof processes. To this end, I offer a Bayesian model of credence with the allied decision-theoretic component in which two types of chance-credence principles play a pivotal role.

1.3.2 Philosophy of Chance

The concept of chance has been a matter of considerable discussion among philosophers for a long time. Traditionally, frequency and propensity theories are regarded as the two dominant philosophical theories of chance. These theories are meant to be metaphysical: they aim at providing an answer to the question of what chance is.[6] In so doing, they hold that the concept of chance can be analysed in terms of more graspable concepts that themselves make no reference to chance. As such, they embody a classical project of conceptual analysis: they show when a story told in a target vocabulary (a story about the analysandum) is made true by a story told in some more fundamental or base vocabulary (a story about the analysans) thought to be more graspable.[7]

The two dominant philosophical theories of chance differ on the nature of the analysans of chance. Whereas frequentists claim that the analysans of chance describes some sort of regularity, proportion in the arrangement of instantiations of some non-modal and non-dispositional properties pertaining to the outcomes of a certain type of experiment, propensity theorists take this analysans to describe some kind of dispositional and irreducible property of a chance set-up.[8] Propensity theorists claim that the notion of chance as posited in our scientific theories, stochastic laws, and models expresses a kind of modality that itself cannot be analysed away: in particular, it cannot be identified with the patterns among outcomes of some type of experiment. But in recent years, both these theories,

[6] For example, this point has been made forcefully in Strevens (2003, chapter 1). He distinguished between a metaphysics and a physics of probability. While the former concerns the question of what probability is, the latter focuses on the question of what physical underpinnings of probabilistic processes are. He then argued that traditional philosophical theories of chance, like frequency and propensity theories, are primarily concerned with the former question.

[7] By analogy, to conceptually analyse the notion "knowledge" is to reduce a story told in a vocabulary that uses "knowledge" to a story told in a vocabulary that uses the terms "true", "justified", and "belief".

[8] Various frequency theorists identify chance with either actual finite relative frequency (Venn 1866), limiting relative frequency (Reichenbach 1949; von Mises 1957), or with hypothetical limiting relative frequency (Kyburg 1974; van Fraassen 1979) of the occurrences of some event or property in a certain reference class. Propensity theorists identify chance with a property of some experimental or chance set-up to produce long-run relative frequencies (see, e.g. Popper 1959), or to produce certain outcomes on single trials (see, e.g. Giere 1973).

and the analytical project they embody, have fallen into disrepute. In particular, several arguments have been devised to show that frequency and propensity theories cannot provide an illuminating understanding of chance (see, e.g. Hájek 1996, 2009; Eagle 2004).

There is, however, another idea that drives current attempts to scrutinize the concept of chance. This idea can be expressed as follows: we can master the concept of chance by trying to understand what chance does, or what functional roles it plays. More specifically, by elevating the Principal Principle to the status of the sole constraint on chance, Lewis, in his "Humean Supervenience Debugged", advanced the idea that chance is whatever satisfies the Principal Principle:

> A feature of Reality deserves the name of chance to the extent that it occupies the definitive role of chance; and occupying the role means obeying the old Principle [the Principal Principle], applied as if information about present chances, and the complete theory of chance, were perfectly admissible. (Lewis 1994, p. 489)

A closely related view to Lewis's has been defended by Jonathan Schaffer (2007). He has argued that chance is what chance does, and a given probability function is chance if it plays a number of chance roles, not only the role expressed by Lewis's Principal Principle. For example, according to Schaffer, chance is whatever *grounds* realizing possibilities: if there is a non-zero chance that a flipped coin will land heads, and the proposition that the flipped coin landed heads is true, then there is a possible world in which the coin lands heads; this chance then is said to be grounded by that world, and not grounded by a possible world in which the flipped coin lands on its edge. Or, chance is whatever fits a lawfully projected quantity: if the chance that the coin lands heads is 0.5, then the laws of nature together with the history of the coin-landing outcomes entail that value. According to Schaffer, appreciating these "chance roles" provides a viable understanding of the concept of chance. That is, we can understand this concept by endorsing the following characterization:

> Chance is that probability function from propositions, worlds, and times onto the closed unit interval, which best satisfies: (i) the Principal Principle, (ii) the Realization Principle, (iii) the Futurity Principle, (iv) the Intrinsicness Requirement, (v) the Lawful Magnitude Principle, and (vi) the Causal Transition Constraint. (Schaffer 2007, p. 126)

Importantly, all the principles listed in Schaffer's characterization of chance are supposed to be platitudes we apparently have about chance, platitudes that capture important roles played by chance in our life.

In discussing the notion of chance, the thesis does not attempt to provide a conceptual analysis of chance in the spirit of frequency and propensity theories. Rather, it exemplifies the type of approach to chance proposed by Lewis and Schaffer. To give it a name, this approach may be called *chance functionalism*.

In subsequent chapters concerning the life of probability as chance, the general idea of chance functionalism is articulated by saying that chance is a quantity that satisfies certain plausible principles or conditions. These principles are meant to capture various roles we attribute to chance. Instead of providing an exhaustive list of these principles, the thesis explores only chance-credence principles and certain kind of chance-chance principles. This methodological choice may be considered as a shortcoming, serious enough to doubt whether the thesis should be added to the library. But the mission of the study of chance presented here is different: it aims to show how reorienting our way of thinking about chance could be worthwhile. In so doing, it invites the reader to judge the new perspective on chance by its fruits, not by whether it provides a complete picture of chance.

Focusing on various principles for chances helps in readdressing some old problems as well as in developing some relatively unexplored areas. One of the old chestnuts is the question of whether chance can be represented as the mathematical concept of probability, known as the question of *formal adequacy*. Traditionally, this question has been approached by providing representation theorems for various philosophical theories of chance. According to this approach, a conception of chance satisfies some axiomatization of probability just in case one can show that the characteristic kind of axiomatized structure of that conception is representable as a probability space. Typically, this approach proceeds, first, by laying down some mathematical structure (a chance structure) and axioms that are claimed to be characteristic of a given conception of chance, and then by proving, in the form of a *representation theorem*, that such chance structure is representable as a probability space (Suppes 1973, 1974; Eells 1983).

Whether or not this approach is cogent and even applicable depends to a large extent on what philosophical theory of chance is assumed. If such theory takes chance to be reducible to some sort of relative frequency or limiting relative frequency, then it is a straightforward exercise to show that a frequency structure is isomorphic to a finitely additive probability space. But things are different when we think of chance as some kind of real and irreducible propensity, or genuine causal tendency to produce a given outcome or a series of outcomes. When chances are irreducible propensities that cannot be defined in terms of some observable or detectable properties, but can only be elucidated by them, there seems to be no straightforward way to prove that they are representable as a probability function on some space. For if propensities cannot be so defined, it is not entirely clear how they behave, and so what axioms a propensity structure might satisfy. But despite this apparent obstacle, some authors have attempted to provide representation theorems for propensities. Most notably, Ronald Giere (1976) has provided representation theorems for a single-case propensity theory. This result, however, has been criticized on the grounds that the axiomatized structure of the single-case propensity, as being based on a Laplacean possibility model, is entirely formal, and as such it is not able to provide a satisfactory understanding of what propensities might be (e.g. Milne 1987).

Things are even more complicated with the representation theorem approach if we think of chance as a theoretical concept that cannot be reduced to non-theoretical terms. Elliott Sober's (2010) *no-theory theory of chance* is a recent attempt to understand chance in this way. If chances are just theoretical quantities posited by our scientific theories, then not only does the representation theorem approach lose its allure, but also the very question of formal adequacy seems superfluous. For, by and large, the characteristic axiomatized structures of those probabilistic theories (e.g. quantum mechanics, statistical mechanics, or evolutionary theory) are already modelled as probability spaces.

My aim is to address the question of formal adequacy from a different angle. To approach this question, we need not to decide whether chance is a sort of frequency, propensity, or theoretical entity. Consequently, we need not to lay down some characteristic axiomatized chance structure. Instead, I explore one particularly important functional role of chance—the expert role carved out by the Principal Principle—and then argue that this role demands chance to be a finitely additive probability function. This view is not only a viable alternative to the representation theorem approach, but it also avoids some serious shortcomings of Lewis's (1986) argument for chance's formal adequacy. Like the approach to be presented, Lewis's argument builds on the thought that the Principal Principle is a vehicle by which we could draw the conclusion that chance is a finitely additive probability function. But, controversially, it forces us to adopt the view that chance is a sort of credence, an *objectified* credence.

A relatively unexplored problem by philosophers is the question of how chances should change through time. An interesting approach to this problem was proposed by Lewis (1986). He argued that a particular kinematical model for chances follows from his Principal Principle. The model that follows from this principle is a form of Bayesian conditionalization: any later chance function is equal to an earlier chance function conditional on the intervening history of categorical-property instantiations in between.

But Lewis's argument for Bayesian kinematics of chance appears to be problematic: there are reasons to doubt whether the epistemic role of chance captured by the Principal Principle could be used to establish the way by which chances should change through time. Intuitively, if chances are attributable to the mind-independent world, the way they evolve through time might be quite independent of the way they constrain one's credences. But, as will be apparent, we can retain Bayesian kinematics of chance without appealing to the Principal Principle. More specifically, it can be shown that this kinematics of chance follows from a plausible principle relating prior and posterior chances. This sort of chance-chance principle turns out to be an essential feature of resilient chances, that is, it characterizes chances that maximize resiliency under variation of intervening histories.

In developing the idea of resilient chances, the thesis also sheds new light on a class of philosophical theories of chance called Humean accounts of chance (Lewis 1994; Loewer 2004; Hoefer 2007). These theories face a problem known as the Big Bad Bug, which has been a matter of considerable discussion in recent years.[9] As presented originally in Lewis (1994), the Big Bad Bug is a reductio which aims to show that Humean accounts of chance contradict the Principal Principle. The Big Bad Bug is regarded essentially as an *epistemological* argument, for it

[9]For example, an extensive discussion of this problem is to be found in Lewis (1994), Thau (1994), Hall (1994), Ismael (2008), and Briggs (2009a).

appeals to the Principal Principle that captures the epistemic role of chance in guiding one's credence. But, as suggested in Bigelow et al. (1993) and in Briggs (2009b), a version of the Big Bad Bug can be presented without making this epistemological detour. More specifically, the Big Bad Bug can be formulated as a genuine *metaphysical* argument by appealing to certain chance-chance principles relating prior and posterior chances. I explore this suggestion by linking the metaphysical reading of this argument with the idea of resilient chances. This exploration in turn leads to a more general observation that Humean chances do not necessarily maximize resiliency under variation of experimental factors.

1.3.3 Bayesianism

Bayesianism is a philosophical position centered on the interpretation of probability as credence. It is also one of the best examples of a research field in which the interaction between credence and chance is widely recognized. For one can advocate Bayesianism without endorsing the radical position that probabilities should always be credences.

Conceived as a philosophical position, Bayesianism has a wide scope of application: there are Bayesian positions in epistemology (in particular, in theories of quantitative belief), decision theory, philosophy of science (in particular, in theories of scientific confirmation and inductive reasoning), statistics, in legal theory (in particular, in theories of legal evidence). Despite this wide scope of application, there are some features that might be regarded as common to all Bayesian positions. They may be introduced as the following claims:

(i) An agent's doxastic attitudes come in degrees called credences, characterized by a function that assigns real numbers to propositions.

(ii) The rational agent's credences ought to satisfy the axioms of probability.

(iii) Bayes's rule or the rule of conditionalization is the standard way of updating the agent's credences.

There are many points of disagreement among Bayesians.[10] One substantial point

[10] For excellent surveys of the main points of disagreement among Bayesians, see Easwaran (2011) and Weisberg (2011).

of disagreement concerns the question of whether there are other constraints besides (ii) and (iii) that the rational agent's credences ought to satisfy.[11] Bayesians that endorse only the constraints (ii) and (iii) are traditionally called *strictly subjective* or *permissive.*

By imposing only these two constraints, strictly subjective Bayesians, like de Finetti and his followers, allow for a wide disagreement between agents' credences, even if they are formed on the basis of the same evidence. This is so because constraint (ii), traditionally called Probabilism, can be satisfied in a number of different ways, i.e. different degrees-of-belief assignments may satisfy this constraint. Strictly subjective Bayesians often add that when evidence accumulates, the disagreement between agents' initial or prior credences would be "washed out" in the long run, and so agents' credences would come to an agreement.

Less permissive subjective Bayesians, called sometimes *empirically-based subjective Bayesians* or *tempered personalists*[12], hold that the agent's credences ought to be proportional to her evidence. And some stripes of this sort of subjective Bayesianism, like Abner Shimony's tempered personalism, hold that agents should assign non-zero credences to any hypothesis seriously proposed by a member of a scientific community.

So-called *objective Bayesians* are even less permissive than tempered personalists: some, like Harold Jeffreys (1961), hold that there are constraints that collectively fix a unique rational credence, others, like Edwin Jaynes (1957) and Jon Williamson (2010, pp. 15-19), defend the view that if evidence determines a range of compatible credences, the agent should adopt those credences that equivocate between the propositions over which she has an opinion. A characteristic feature of many sorts of objective Bayesianism is that the constraints put on credences, like Jaynes's principle of maximum entropy, go beyond purely empirical ones.

Among additional constraints on credences endorsed by both empirically-based subjective Bayesians and objective Bayesians is the requirement according

[11]Constraint (i) is also a matter of considerable discussion in the literature. There is much controversy over the question of whether one's credence should by characterized by a *single* real value or by a *set* of such values. The second option gives rise to the idea of "vague" or "indeterminate" credence.

[12]For a more detailed characterization of tempered personalism, see Shimony (1970), Earman (1992, p. 35), and Williamson (2010, pp. 15-19).

to which the agent's credences should reflect all the available evidence. That is, the set of the agent's permissible credences should be narrowed down to the set of credences that are proportional to her evidence. When the agent's evidence is about chance, the way in which the agent's credences should be proportional to that evidence is captured by various chance-credence principles.

In a way, the perspective from which the concept of credence is studied in the thesis is a sort of empirically-based subjective Bayesianism. The thesis, however, does not focus on providing a systematic defence of the main tenets of that Bayesian position. Specifically, it is not concerned with providing a justification for chance-credence principles. Such justifications have been given elsewhere in the literature.[13] Instead, it concentrates on how these principles can illuminate certain questions concerning the epistemology of credence. Viewed this way, the thesis presents an application of this core component of empirically-based subjective Bayesianism in the context of legal proof process. It develops a Bayesian model of legal fact-finding covering both a fact-finder's credences over factual hypotheses disputed in courts of law and her verdicts. This particular application helps in addressing the problem of using "naked" statistical evidence in that context.

Together with a concept of accuracy suitable for the legal proof process, the thesis develops a close connection between chance-credence principles and the idea of credence and verdict accuracy maximization. In so doing, it subscribes to one of the most flowering research areas within Bayesian epistemology, viz. the project of linking chance-credence principles with the idea of accuracy maximization (see, e.g. Pettigrew 2012, 2013a). Also, by exploring the idea of accuracy maximization in the legal proof process, the thesis develops an approach that closely resembles Alvin Goldman's (1999, chapter 9) *veritistic* epistemology as applied to legal settings. As shown by Goldman (Goldman 2002), this approach allows us to defend Bayes's rule in legal settings as a veritistically good inferential practice, viz. practice that is likely to end up closer to the truth. Much like in Goldman's approach, it is argued that (i) credences satisfying certain chance-credence principles cannot lead to harmful verdicts: the subjective

[13] For a sampling literature focused on this issue, see Braithwaite (1966), Mellor (1982), Earman (1992, chapter 2), Strevens (1999), Howson and Urbach (2006, chapter 3), and Williamson (2010, pp. 39-42).

expected accuracy of verdicts cannot decrease when credences are expected to match chances, and (ii) the objective expected accuracy of credences satisfying those chance-credence principles can only increase.

Bayesians also disagree as to whether Bayes's rule is the only rational method of updating the agent's credences upon receipt of new evidence. Some Bayesians, like Richard Jeffrey (1983), propose a modified form of conditionalization called Jeffrey's rule. This rule is a generalization of Bayes's rule: while Bayes's rule applies only to a learning experience that makes the agent certain of a proposition, Jeffrey's rules applies also to a learning experience in which the agent redistributes her credences over some propositions without her becoming certain of any of them. But it is widely recognized that both Bayes's and Jeffrey's rule cannot be applied to every case in which we need to change our credences. Moreover, even if the two conditionalization rules are applicable, it appears that they sometimes lead to hardly acceptable results. Specifically, it is often claimed that learning experiences where one assigns a conditional posterior credence to some proposition given another proposition cannot be adequately modelled by Jeffrey's rule. One of the widely discussed examples of this sort is the famous Judy Benjamin Problem (van Fraassen 1981; van Fraassen et al. 1986).

The thesis contributes to the topic of credence updating by focusing on the principle of maximum relative entropy, which is typically endorsed by objective Bayesians. This principle allows us to cover a wide spectrum of learning experiences: in principle, it can be used to model a learning experience where one's expectation of a random variable, computed relative to one's posterior credence function, gets a certain value. But in comparison with the conditionalization rules, the question of how we can justify this principle as an updating method received relatively little attention in the literature.[14] Against this background, I explore the possibility of providing such justification by looking at the extent to which updating governed by this principle leads to new credences that are expected to be helpful, and never harmful in making decisions. Along this way,

[14]For various arguments in favour of the conditionalization rules, see Teller (1973), Brown (1976), Skyrms (1987a), van Fraassen (1989, pp. 331-337), Armendt (1993), van Fraassen (1999), Greaves and Wallace (2006), Leitgeb and Pettigrew (2010b), and Easwaran (2013). Attempts to provide a justification for the principle of maximum relative entropy are to be found in Jaynes (1957), Shore and Johnson (1980; 1981) and Grünwald (2000).

I show that recognition of the Judy Benjamin Problem may imperil such justification by showing that the value of learning theorem may not hold when one learns a conditional proposition.

1.4 Central Presuppositions of the Thesis

Given the variety of issues to be discussed, it is necessary to state some central presuppositions and regimentation of concepts if the discussion is to go forward. Below I describe the main background of the thesis together with a prolegomenon to the topic of expert functions.

1.4.1 Higher-Order Probabilities

Although each of the coming chapters is more or less self-contained, developing its own framework to tackle issues pertinent to chance or credence, there is an important presupposition that underlies and motivates discussions in each of the chapters—the idea that we can explore certain important relations between chance and credence, chances themselves, and credences themselves by imposing a framework of higher-order probabilities. This framework constitutes the entrance hall to the entire edifice. It allows us to speak about credences assigned to propositions about chances, prior chances assigned to propositions about posterior chances, and prior credences assigned to propositions about posterior credences.

The idea of higher-order probabilities has a long tradition in the philosophy of probability. Though initially contested as leading to inconsistency (Good 1950; Miller 1966) or to a problematic "endless hierarchy" (Savage 1954, p. 58), this idea gained currency as being legitimate and theoretically fruitful (Skyrms 1980b; Domotor 1981; Lewis 1986; Gaifman 1988; Peijnenburg and Atkinson 2012). In recent years, its formal fruitfulness has been proved by showing that higher-order probabilities (i) allow us to formulate various principles for credences like chance-credence principles (Lewis 1986) or Reflection principles (van Fraassen 1984), (ii) enable us to build a richer theory of credence updating that allows us to conditionalize on propositions about posterior credences (Skyrms 1980b;

Domotor 1981), or (iii) allow us to formulate important principles relating prior and posterior chances (Bigelow et al. 1993; Lange 2006).

The thesis does not develop a formal theory of higher-order probabilities, however. Instead, the various higher-order probabilities used in this thesis can be read through the lens of Haim Gaifman's (1988) framework, which underpins his theory of expert functions.[15] The guiding idea behind Gaifman's framework is that we can enlarge the original set of propositions over which a probability function is defined by adding propositions about probabilities that this probability function assigns over propositions in the original set. Consequently, we can effectively assign probabilities to all the propositions in the enlarged set. According to Gaifman, the propositions about probabilities may be understood as propositions describing various expert functions or expert assignments.

1.4.2 Expert Functions

Since the idea of expert functions underpins the principles used in this theses, it is useful to show at the outset how this idea works.

Let p and q be two probability functions over a finite set of propositions $\mathcal{A}$. Each function assigns a non-negative real number from $[0, 1]$ to every $A \in \mathcal{A}$. Let C_q be the proposition that the probability distribution over $\mathcal{A}$ is given by q. I assume that $C_q \in \mathcal{A}$. Then:

Expert Function: For all $A \in \mathcal{A}$ and all q,

$$p(A|C_q) = q(A),$$

providing $p(C_q) > 0$.

That is, q is an expert function for p concerning a set of propositions $\mathcal{A}$ just in case, for every proposition A in $\mathcal{A}$, the probability of A conditional on the proposition C_q, $p(A|C_q)$, is equal to the probability $q(A)$.

How could this abstract idea of expert function underpin the various principles discussed in this thesis? To give a concrete example, let us focus on chance-credence principles. If we replace p with an agent's credence function cr, and q

[15] A similar theory of expert functions has been developed in van Fraassen (1989, chapter 8).

with a chance function ch, we get the following form of chance-credence principle:

Chance-Credence Principle: For all $A \in \mathcal{A}$ and all ch,

$$cr(A|C_{ch}) = ch(A),$$

providing $cr(C_{ch}) > 0$.

This principle expresses the idea that a chance function is an expert function for the agent's credence function concerning a set of propositions $\mathcal{A}$.

Similarly, the idea of expert function underpins chance-chance and credence-credence principles used in this thesis. That is, if we think of p as a prior chance function and of q as a posterior chance function, we obtain a form of chance-chance principle: it expresses the idea that a posterior chance function is an expert function for a prior chance function concerning $\mathcal{A}$. And if we take p as standing for a prior credence function and q for a posterior credence function, we get an analogous formulation of credence-credence principle.

It is also useful to drive a wedge between two types of expert functions, viz. database-expert function and analyst-expert function.[16] Stated broadly, we defer to database-experts because they are better informed than we are, and we defer to analyst-experts because they are better than we are at analysing evidence they are given. So we might defer to an analyst-expert, even if we are better informed than she is. For example, a judge in a criminal court regards a mental health expert as an analyst-expert because the expert is good at analysing evidence at trial, no matter how much information about the crime she has. But the judge regards an eyewitness to the crime as a database-expert because the witness is better informed about the crime, and not because she is good at analysing the evidence at trial. Importantly, as argued in Hall (2004), q is a database-expert for an agent's credence function cr if the agent has no evidence that the expert lacks. After all, she defers to q because of q's evidence. But this is not so with analyst-experts: to regard q as an analyst-expert is to defer to q's opinions conditional on the agent's evidence.

With this in place, let us apply the notions of database- and analyst-expert

[16] This distinction has been introduced by Ned Hall (2004), and it dovetails with Elga's (2007) expert/guru distinction.

function to the chance-credence relation. Suppose that $E \wedge C_{ch}$ is the proposition about the agent's total evidence. Then:

Chance as Database-Expert Function: For all $A \in \mathcal{A}$ and all ch,

$$cr(A|E \wedge C_{ch}) = ch(A),$$

if E is admissible and $cr(E \wedge C_{ch}) > 0$.

That is, to defer to chance regarded as a database-expert is to set your credence in A equal to the *unconditional* chance of A, provided that your evidence E is admissible. And, roughly, E is admissible if it gives no information about the truth of A that does not go through the chance of A.[17] And:

Chance as Analyst-Expert Function: For all $A \in \mathcal{A}$ and all ch,

$$cr(A|E \wedge C_{ch}) = ch(A|E \wedge C_{ch}).$$

That is, to defer to chance regarded as an analyst-expert is to set your credence in A equal to the *conditional* chance of A, where the condition is your total evidence $E \wedge C_{ch}$.

In the coming chapters, the thesis puts to work chance-credence principles that characterize chance as both a database- and expert-function. In particular, it uses Lewis's Principal Principle and its various formulations as an instance of Chance as Database-Expert Function, the so-called New Principle as an instance of Chance as Analyst-Expert Function, and the End-Point Chance-Credence Principle as a special case of Chance as Analyst-Expert Function.

1.5 Synopsis of the Thesis

Our journey begins in chapter 2 with developing an answer to the question of whether chance satisfies a constraint known in the literature as the condition of

[17] For similar formulations of what admissible evidence is, see Lewis (1986), Loewer (2004), and Pettigrew (2012). A somewhat different characterization of admissible evidence is to be found in Thau (1994), Strevens (1995), and Meacham (2010).

formal adequacy, which requires chance to satisfy some axiomatization of probability. The primary aim of this chapter is to show how considerations concerning the chance-credence interaction can be used to vindicate the formal adequacy of chance. More precisely, this chapter introduces and motivates a framework in which it can be shown that, under fairly intuitive assumptions, the expert role codified by the Principal Principle demands chance to be a finitely additive probability function.

Chapter 3 focuses on two chance-chance principles relating prior and posterior chances. I first introduce, motivate, and make precise a resiliency-centered approach to chance whose basic idea is that any chance distribution should be maximally invariant under variation of experimental factors. Second, I provide resiliency-based arguments for the two principles: I show that any chance distribution that violates them can be replaced by another chance distribution that satisfies them and that is more resilient under variation of experimental factors. I then go on to show that these principles lead to hardly acceptable consequences in the case of Humean accounts of chance. Also, I show that considerations of the resiliency of chance have substantial repercussions on the question of whether these principles should be retained in that case.

The topic of chapter 4 concerns the kinematics of chance, to wit, the problem of how chances should change through time. First, the chapter investigates the conditions that any kinematical model for chance needs to satisfy to count as Bayesian kinematics of chance. Second, it presents and discusses Lewis's argument for Bayesian kinematics of chance, viz. it shows how this kinematical model for chances follows from Lewis's Principal Principle. Third, the chapter presents an alternative argument for Bayesian kinematics of chance that does not appeal to the Principal Principle, but to a principle that relates prior and posterior chance functions. This principle in turn is motivated by resiliency considerations similar to those presented in chapter 3.

Chapters 5 and 6 focus on the credence-side of probability. In chapter 5, I apply a simple Bayesian model to legal fact-finding to argue that statistical evidence in courts of law is conducive to the achievement of accuracy, which appears to be a fundamental objective of legal fact-finding. I present two accuracy-based arguments for the thesis that chances should constrain a fact-finder's credences about factual hypotheses discussed in courts of law. The first argument says that the fact-finder's credences informed by chances cannot lead to a decrease of subjective expected verdict accuracy. The second argument shows that the fact-finder's credences informed by chances maximize objective expected credence accuracy. The notions of subjective expected verdict accuracy and objective expected credence accuracy are precisely explained within a Bayesian model of legal fact-finding. This model also induces a particular chance-credence principle that captures the idea of using exclusively statistical evidence in courts of law.

While chapter 5 concentrates on issues pertaining to the statics of credence (it studies various constraints on credence at a given time of the agent's epistemic life), chapter 6 moves to the dynamics of credence (it studies various ways by which the agent should change her credences over time). It examines the possibility of justifying the principle of maximum relative entropy, considered as an updating rule, by looking at the value of learning theorem established in classical decision theory. This theorem captures an intuitive requirement for learning: learning should lead to new degrees of belief that are expected to be helpful, and never harmful in making decisions. I call this requirement *the value of learning*. I consider the extent to which learning ruled by the principle of maximum relative entropy could satisfy this requirement, and so could be a rational means for pursuing practical goals. In passing, I discuss a long standing controversy in the philosophy of credence, that is, the question of whether there could a universal or mechanical updating rule for credences.

Most of the chapters can be fully comprehended without reading the material in other chapters. The only principal interdependency concerns chapters 3 and 4. That is, the latter applies the idea of chance's resiliency that is fully explained and motivated in the former.

Chapter 2

Chance-Credence Principles and the Question of Formal Adequacy

The chapter explores a particular chance-credence relation in order to develop an argument for the thesis that chances are finitely additive probability functions. The chance-credence relation that plays a pivotal role in this chapter is captured by David Lewis's Principal Principle.

The argument to be given purports to show that the expert role of chance codified by the Principal Principle demands chance to be a finitely additive probability function. This result is established, first, by employing chance functionalism: the view that chance is whatever plays certain functional roles. Second, by exploring the expert role codified in the Principal Principle, it is claimed that a real-valued function ch over a finite set of propositions is chance if it plays the role of an expert for the epistemic agent. Third, it is argued that the function ch plays that role if it is weakly predictively accurate. This condition is taken to be fairly minimal: ch is weakly predictively accurate if and only if there is no other real-valued function over a finite set of propositions, known to the epistemic agent, that could match better the possible truth-value distributions over that set. Fourth, it is proved that the function *ch* is weakly predictively accurate if and only if it is a finitely additive probability function. Finally, it is concluded that the function *ch* is chance if it is a finitely additive probability function.

2.1 Introduction

Is chance a model for the mathematical theory of probability? Philosophical tradition has it that any satisfactory account of chance (physical or statistical probability) should meet certain fairly intuitive conditions (Salmon 1967; Suppes 1973, 1974; Eells 1983). One of the core conditions, which has become to be known in the literature as *formal adequacy*[1], says that chance should satisfy some axiomatization of probability.[2] This chapter shows how considerations concerning the chance-credence relation can be used to vindicate the formal adequacy of chance. More precisely, this chapter introduces and motivates a framework in which it can be shown that the expert role codified by the Principal Principle demands chance to be a finitely additive probability function.

The argument to be presented in this chapter can be outlined as follows:

1. A real-valued function *ch* over a finite set of propositions is chance if it plays the role of an expert for the epistemic agent whose evidence is admissible. That is, *ch* is chance if it satisfies the Principal Principle.

2. *ch* plays the role of an expert if it is weakly predictively accurate. That is, *ch* is an expert function if there is no other real-valued function ch', known to the epistemic agent, that could match better all the possible truth-value distributions over a finite set of propositions.

3. Theorem: *ch* is weakly predictively accurate if and only if it is a finitely additive probability function. This can be shown once we adopt the Brier

[1] Wesley Salmon (1967) called this condition *admissibility*.

[2] I understand the condition of formal adequacy broadly, in the sense that it does not require chance to satisfy a particular axiomatization of probability, e.g. Kolmogorov's axioms, Popper's axioms, or Rényi's axioms of probability. Rather, the condition says that chance should satisfy *some* axiomatization of probability.

score as a measure of the "distance" between the real-valued function *ch* and a truth-value distribution.

4. Therefore, *ch* is chance if it is a finitely additive probability function.

That is, starting from the premise that chance is an expert function for the epistemic agent, and that any expert function must be weakly predictively accurate, we get the conclusion that chance must be a finitely additive probability function.

The linchpin of this argument shows close affinity with a more familiar idea of James M. Joyce's justification of *probabilism*—the thesis that one's credence function should obey the axioms of probability. Inspired by Bruno de Finetti (1970), Joyce (1998) showed that a credence function that is incoherent (violates the axioms of probability) can be replaced by a coherent one which is closer to the truth values of propositions in every possible world.[3] Similarly, the argument just outlined can be read as follows: if *ch* is not a finitely additive probability function, then there is ch' that could match better all the possible truth-value distributions. Hence, by premise 2, *ch* is not an expert function and, by premise 1, is not a chance function.

The general similarity with Joyce's approach notwithstanding, there are also important differences. The first and most obvious difference is that this chapter aims to justify the probability axioms as a requirement for chances posited in the stochastic theories and referred to the objective features of a world, not as a rationality norm for one's doxastic attitudes understood as credences. Of course, it might be pointed out that this difference vanishes if one interprets chances as embodiments of expert opinions. According to such a view, chances would be credences of an expert believer, perhaps suitably constrained. But the account of chance defended in this paper does not commit us to accept that view. That is, I argue that chance is first and foremost a function from propositions to real numbers that plays certain roles like guiding one's credences in those propositions, and as such it need not be reducible to frequencies, propensities, or a believer's credences.

[3] In a similar spirit, Hannes Leitgeb and Richard Pettigrew (2010b; 2010a) have argued that conformity to the axioms of probability results in minimizing expected inaccuracy of credence functions.

The second difference is that the approach focuses on the predictive accuracy of chances, rather than on the closeness-to-the-truth of one's credences. While the former has to do with how accurately the predicted chances match possible random outcomes of a given chancy process, the latter concerns the question of how accurately credences *represent* a world. Moreover, I suggest to take the former as a key to explain why chances are expert functions, while the latter is often regarded as a mark of partial believers' epistemic success.

The third difference worth emphasizing is that while the accuracy of credences is taken to be *absolute*, the predictive accuracy of chances defended here is *comparative*. That is, whereas a credence is accurate if it is as close as possible to the truth (ideally having the value 1 for the truth and 0 for the falsehood), a chance is predictively accurate if there is no other chance that could match better the possible outcomes (it can have an intermediate value, say 0.5, insofar as no other chance could do better in reducing the distance to the truth about these outcomes). According to this view, no threshold to be exceeded in reducing this distance is required to call a chance predictively accurate. I argue that this *weak* notion of predictive accuracy not only does justice to our intuitions about chance, but more importantly it suffices to show that chance is a finitely additive probability function.

The possibility of using a chance-credence principle to tackle the issue of formal adequacy is perhaps mostly connected with David Lewis's philosophy of chance, launched in "A Subjectivist's Guide to Objective Chance" (Lewis 1986). Lewis showed that his Principal Principle, though primarily concerning the relationship between chance and credence, brings interesting consequences for chance alone. One of those consequences is that chance is a finitely additive probability function, and so satisfies formal adequacy. At the core of Lewis's argument lies a particular use of the Principal Principle, which defines chance as a suitably objectified credence. As will be apparent in this chapter, this use of the Principal Principle is controversial. Though the argument to be given also appeals to the Principal Principle, it does not commit us to adopt Lewis's controversial move. Rather, the understanding of the Principal Principle that inspires my argument is much closer to Lewis's idea introduced in his "Humean Supervenience Debugged" (Lewis 1994). In that paper, he advanced the idea that chance is whatever satisfies the Principal Principle. Elaborating on this idea, I argue that chance is whatever plays the expert role codified by the Principal Principle together with some other equally important roles.

Before I continue, let me say something more about why formal adequacy appears to be an important condition on chance. Clearly, we require chance to satisfy more than just the condition of formal adequacy. This is because objects that can hardly be called chances also satisfy formal adequacy. For example, if the surface of a table is divided into black and red measurable regions, the proportion of black regions of the table satisfies Kolmogorov's axioms of probability. But we are far from calling this proportion a chance. However, formal adequacy seems to place a substantive constraint on chance. Obviously, if chance does not satisfy formal adequacy, then it is not an interpretation of a mathematical concept of probability, but it is an interpretation of some other mathematical formalism. More importantly, if chance satisfies formal adequacy, then we know how to reason about it. For to reason about chance, we use axioms and theorems of a given mathematical theory of probability. Thus, the language and rules of a mathematical concept of probability transfer into the language and rules for chances.

It also needs to be emphasized that, unlike the other chapters of this thesis, this chapter uses the set operations $\cap$ and $\cup$. The reason for using these operations is that, in the framework to be developed in section 2.4, I give a fuller understanding of the propositions over which the function *ch* and the credence function are defined. That is, I will understand these propositions as certain subsets of a set of possible worlds. Since the two functions are to be related by means of the Principal Principle, I also use the set operations to formulate this principle.

The structure of this chapter is as follows. Section 2.2 discusses in more detail Lewis's Principal Principle-centered argument for formal adequacy. Here I explain why I find this argument unsatisfactory and why I think the argument to be defended is a viable alternative with none of the controversies surrounding Lewis's argument. Section 2.3 introduces and discusses the view that chance is whatever plays the expert role codified in the Principal Principle. The discussion is embedded in a more general account of chance called chance functionalism. Here I also explore a link between chance functionalism and Ramsey-Lewis method of conceptual analysis in order to show how we can provide a functional definition of chance. Section 2.4 provides a formal characterization of the real-valued function that is required to fill the expert role given by the Principal Principle—the candidate chance function. Section 2.5 introduces and defends a necessary condition for that function to fill the expert role. This condition is called weak predictive accuracy. This section provides a precise formulation of the predictive accuracy measure, and introduces a particular measure called the Brier score. This measure is then used to underpin the condition of weak predictive accuracy. Section 2.6 states the centrepiece of the argument: a theorem which shows that, relative to the Brier score, any finitely additive candidate chance function is predictively accurate in the weak sense. Section 2.7 provides motivation for using the Brier score to measure predictive accuracy. Section 2.8 generalizes the theorem established in section 2.6 by employing a scoring function represented as Bregman divergence. Section 2.9 provides some suggestions on how the predictive accuracy as measured by the Brier score could incorporate a particular measure of calibration between chance functions and relative frequencies. This in turn shows that the predictive accuracy of a chance function could also reflect "nearness" between a real-valued distribution it gives and the relative frequencies. Section 2.10 draws the pieces and concludes.

2.2 Lewis's Argument for Formal Adequacy

This section expounds and examines Lewis's argument for formal adequacy. There are at least two reasons for doing this. First, Lewis's argument is arguably the first argument incorporating the thought that the Principal Principle is a vehicle by which we could draw the conclusion that chance is a finitely additive probability function. Second, by pointing out some serious shortcomings of Lewis's argument, I suggest how to reorient the use of the Principal Principle in order to avoid these shortcomings and possibly to reach the conclusion that chance is a finitely additive probability function.

Lewis (1986, p. 98) explored a particular formulation of the Principal Principle in order to reach the conclusion that chance is a finitely additive probability function. His argument can be presented as follows:

1. Given a finite partition of histories and complete theories of chance $\{H_i \cap T_i\}$, the chance distribution at time t and world w over a family of propositions $\mathcal{A}$, ch_{tw}, comes from any "reasonable initial credence function"[4] over $\mathcal{A}$, cr, by *conditionalizing* on the member of the partition $H_i \cap T_i$ that holds true at w. Formally, for any time t, world $w \in H_i \cap T_i$, and proposition A in the domain of ch_{tw},

$$ch_{tw}(A) = cr(A|H_i \cap T_i).$$

2. cr is a probability distribution.

3. Whatever comes by conditionalizing from a probability distribution is itself a probability distribution.

4. Therefore, ch_{tw} is a probability distribution.

Premises 2 and 3 seem straightforward. Premise 2 might be justified by appealing to Dutch book arguments (Ramsey 1931) or epistemic inaccuracy arguments (Joyce 1998). Premise 3 is true, since any function updated by means of Bayes's rule is a conditional probability function. I will argue below that what makes Lewis's argument controversial is premise 1.

Premise 1 is a version of Lewis's second formulation of the Principal Principle called the Principal Principle Reformulated. What is striking about this principle is that it is able to give a semantics for statements of the form "The chance assignment over $\mathcal{A}$ is given by ch". To show this, consider first the notion of credence relativized to a partition $\{H_i \cap T_i\}$ which is a finite partition of histories together with complete theories of chance (a history-theory partition), i.e. a finite number of ways that the true determinant of A's chance might be. It may be defined as a random variable whose values are:

$$cr(A|H_1 \cap T_1) \text{ if } w \in H_1 \cap T_1,$$

$$\ldots,$$

[4] This is one's hypothetical credence function prior to accumulating any evidence. For simplicity, I will refer to it as one's credence function, always keeping in mind its special meaning.

$$cr(A|H_n \cap T_n) \text{ if } w \in H_n \cap T_n.$$

Each such value may be called a *version* of the credence of A conditional on $\{H_i \cap T_i\}$.[5] According to the Principal Principle Reformulated, in any world $w \in H_i \cap T_i$, a version $cr(A|H_i \cap T_i)$ gives the value of the chance of A which is constant for such worlds. Lewis called such version the *objectified credence*, i.e. the credence of an agent after her learning that the cell, $H_i \cap T_i$, of the partition holds true at w.

Importantly, each version or objectified credence gives truth conditions for statements of the form "The chance assignment over $\mathcal{A}$ is given by ch", which can be abbreviated as C_{ch}. That is, $C_{ch} \Leftrightarrow H_1 \cap T_1 \cup ... \cup H_n \cap T_n$, for all $H_i \cap T_i$ such that $cr(A|H_i \cap T_i) = ch(A)$. Moreover, we have that

$$\begin{aligned}
cr(A|C_{ch}) &= cr(A|H_1 \cap T_1 \cup ... \cup H_n \cap T_n) \\
&= \frac{cr(A \cap (H_1 \cap T_1 \cup ... \cup H_n \cap T_n))}{cr(H_1 \cap T_1 \cup ... \cup H_n \cap T_n)} \\
&= \frac{\sum_i cr(A \cap H_i \cap T_i)}{\sum_i cr(H_i \cap T_i)} \\
&= \frac{\sum_i cr(H_i \cap T_i) cr(A|H_i \cap T_i)}{\sum_i cr(H_i \cap T_i)} \\
&= \frac{\sum_i cr(H_i \cap T_i)}{\sum_i cr(H_i \cap T_i)} ch(A) \\
&= ch(A).
\end{aligned} \tag{2.1}$$

That is, the semantics for "The chance assignment over $\mathcal{A}$ is given by ch" given by objectified credences leads us to a rough formulation of Lewis's original Principle Principle. According to this formulation, one's credence in A conditional on C_{ch} should be set equal to $ch(A)$.

The crucial question is whether objectified credence can always be identified with the chance of A for worlds that are members of $H_i \cap T_i$, as it is required by the Principal Principle Reformulated. If this is not the case, then premise 1 should not be accepted, and so Lewis's argument pales. In what follows, I give

[5]For a similar account of credence relative to a partition, see most notably Skyrms (1984; 1991) and Jeffrey (1983, chapter 12). According to Skyrms, it is an essential part of what he calls a subjectivist or Bayesian theory of objective chance.

four reasons for doubting the positive answer to this question.

The first reason for doubting premise 1 is that the Principal Principle Reformulated commits us to accept the view that there might be no unique chance of A. For two different reasonable initial credence functions may lead to two different chances of A, even if they are conditioned on the same proposition $H_i \cap T_i$. And nothing in Lewis's characterization of reasonable initial credence implies that it is unique.[6] While it is acceptable to believe that two chances of A conditional on two different propositions may be different, it is hardly acceptable that two chances of A conditional on the same proposition could differ. If we were to think that one's objectified credence is an estimate of the chance of A, we would reasonably accept the view that two estimates of the chance of A conditional on the same proposition could be different. But the Principal Principle Reformulated does not say merely that one's objectified credence in A is one's estimate of A's chance: it says that this credence is the chance of A.

The next three reasons for doubting premise 1 concern Lewis's own metaphysical account of chance, to wit, his Humean or best-system theory of chance. According to this theory, the value of A's chance at time t and world w is fixed by the global regularity of A's coming out true, covered by a stochastic law of the best system of truths about w (i.e. by a history-to-chance conditional in the set T_i which is true at w), together with the history of A's coming out true up to t.

With this in mind, let me give the second reason against premise 1. Suppose that given a history-theory partition $\{H_i \cap T_i\}$, the value of A's chance at w is fixed by the version $cr(A|H_i \cap T_i)$, but given another coarser, yet equally legitimate by Lewis's own standards, partition $\{H_k \cap T_k\}$, this value is fixed by the version $cr(A|H_k \cap T_k)$ such that it disagrees with the former. Hence, the determination of A's chance also depends on what history-theory partition is assumed. But, then,

[6] According to Lewis (1986, p. 80), one's initial credence "is to be reasonable in the sense that if you started out with it as your initial credence function, and if you always learned from experience by conditionalizing on your total evidence, then no matter what course of experience you might undergo your beliefs would be reasonable for one who had undergone that course of experience". Moreover, even if, as Lewis assumes, one's initial credence function is a regular probability function, there is not just one way in which it can be so: there are different ways in which such a credence function gives a non-negative, normalized, and finitely additive assignment over some set of propositions.

A's chance so determined cannot be Humean, since it is fixed by something more than just the stochastic law and the history. More generally, but relatedly, in his "Postscripts" to "A Subjectivist Guide to Objective Chance", Lewis formulated certain conditions that any history-theory partition should satisfy, e.g. that a partition must be natural, not gerrymandered, or that it should be feasible to investigate which cell of that partition is true, and then concluded that there seems to be no unique partition that satisfies them. So unless other conditions are formulated, there are many competing partitions that are equally admissible. For this reason, he noticed that objectified credences are more like *counterfeit chances* that are "not the sort of thing we would want to find in our fundamental physical theories" (Lewis 1986, p. 121).

The third reason concerns a peculiar feature of Humean chances, i.e. the fact that the Humean chance at w of $H_i \cap T_i$ may not equal 1. That is, for any time t, there is a small but non-zero chance that the determinant of A's chance at w is different than $H_i \cap T_i$, and so $ch(H_i \cap T_i) \neq 1$. This is because Humean chances, understood as global regularities covered by stochastic laws of the best system of truths, are counterfactually dependent on any future course of a world's history, even a future that undermines these chances. Just consider the following example. Suppose that, at some initial time t at world w, $H_i \cap T_i$ says that ten independent coin tosses will take place, and half of them will yield the result "heads". So, $H_i \cap T_i$ determines that $ch(heads) = \frac{1}{2}$. Suppose further that F is the proposition about an *undermining future* which says that the coin will land heads on every toss. We have that, at time t, $ch(F) > 0$ (i.e. $ch(F) = \left(\frac{1}{2}\right)^{10} = 0.0009765625$). Notice that F determines a different Humean chance for heads than ch, i.e. $ch'(heads) = 1$, and so T'_i is true at w.[7] But since the chance function ch gives a non-zero chance to F, and thus to the possibility that T'_i is true at w, it must assign to $H_i \cap T_i$ a chance different from 1.

With this peculiarity in mind, suppose that given a set $U \in \{H_i \cap T_i\}$, which is a disjoint union $H_{i_1} \cap T_{i_1} \cup ... \cup H_{i_m} \cap T_{i_m}$, $m \leq n$, we want to calculate at w the credence $cr(A|U)$. Since, as conditional probability, such credence can be

[7]For a more elaborate analysis of the problem of undermining futures, see Lewis (1994), Ismael (1996), and Vranas (2002). This problem is also discussed in chapter 3.

defined by the ratio formula, we can use the following equivalent product rule:

$$cr(A \cap U) = \sum_m cr(A|H_{i_m} \cap T_{i_m})cr(H_{i_m} \cap T_{i_m}), \tag{2.2}$$

where the credences, $cr(H_{i_m} \cap T_{i_m})$, are the weights assigned by an agent to the disjoint members of U. Of course, after learning that a particular member $H_{i_m} \cap T_{i_m}$ holds true at w, the agent sets her weight assigned to $H_{i_m} \cap T_{i_m}$ equal to 1 while her weights for the other members of U equal to 0. But the weights given by the Humean chance at w may be different, due to the peculiarity mentioned above. That is, such chance may assign a value to $H_{i_m} \cap T_{i_m}$, which is close but not equal to 1, and some positive values to other members of U. Thus, by equating objectified credences with Humean chances, we do not give enough weight to some striking features of the metaphysics of Humean chances.

The fourth reason has been given by Carl Hoefer (2007). His main complaint against premise 1 of Lewis's argument is that the Principal Principle Reformulated gives a misleading picture of Humean chance in saying that the source of Humean chance at a given world is one's credence function. It is misleading because such a chance is rooted in the global regularity among categorical properties covered by stochastic laws of the best system of truths, not in one's credence, however constrained or objectified. It is true that one's credence about A, after enough objectification, may be equal to the Humean chance of A, but it is not clear-cut, and actually doubtful, that such credence is ontologically on a par with the Humean chance. Without a sufficient reason, one cannot simply take the semantics for chance statements given by objectified credences as dictating to the metaphysics of A's Humean chance.

What I have shown above is that we have good reasons to doubt Lewis's Principal Principle-centered argument for chance's formal adequacy. Although the argument to be given in this chapter hinges on the Principal Principal, it dispenses with its problematic reading given by the Principal Principle Reformulated, and fastens on the idea that the Principal Principle is an essential part of an account of chance which I call *chance functionalism*. The view to be developed is that it is constitutive to chance that it should be an expert function, and the Principal Principle is just one way to express this idea. And if chance is whatever satisfies the Principal Principle or, equivalently, whatever plays the expert role, we stay neutral as to whether objectified credence is really a chance. Perhaps it is once we show that it satisfies the Principal Principle. But it is equally possible that other things, usually claimed to be chances like propensities, frequencies or Humean best-system chances, satisfy the Principal Principle equally well. One important point about chance functionalism is that it allows for there being *multiple satisfiers* of the Principal Principle. Thus, by reorienting our understanding of the Principal Principle, we avoid the problematic claim that chance is an objectified credence.

2.3 The Principal Principle and Chance Functionalism

Lewis's Principal Principle connects chance with an agent's credence. But in doing so, the Principal Principle may be viewed as playing at least a twofold role. Its first role is to provide a principle of rationality for an agent's credence function. That is, the Principal Principle *normatively requires* the agent to set her credence about a proposition in accordance with the proposition's chance, given that the agent's evidence is admissible.

The second role of the Principal Principle is to put a requirement on any candidate for chance. According to Lewis (1994), chance is what satisfies the Principal Principle and thus any feature of reality that aspires to be chance must satisfy this principle. Lewis put this as follows:

> Don't call any alleged feature of reality "chance" unless you've already shown that you have something, knowledge of which could constrain rational credence. (Lewis 1994, p. 484)

To be clear, by highlighting this role of the Principal Principle Lewis suggested to think of chance as follows: take any feature of reality which you think can play the role of chance, whether it be a frequency, symmetry, propensity, or Humean best-system chance. But to be called chance properly, this feature must be a thing that would, if known, constrain an agent's credence.

Lewis's Principal Principle-centered view about chance can be make more precise as follows. Let *ch* be a real-valued function over a family of propositions

$\mathcal{A}$, which assigns to any proposition A in $\mathcal{A}$ a chance—a number in $[0, 1]$. Think of the function *ch* as representing a candidate for chance, e.g. frequency, strength of propensity, or Humean best-system chance. Assume that *cr* is an agent's credence function over $\mathcal{A}$ assigning credences—numbers in $[0, 1]$. Let C_{ch} be the proposition that the chance assignment over $\mathcal{A}$ is given by *ch*, and assume that $cr(C_{ch}) > 0$.[8] Then:

> **Chance:** The function *ch* is *chance* if for all propositions $A \in \mathcal{A}$,
>
> $$cr(A|E \cap C_{ch}) = ch(A),$$
>
> providing that evidence E is admissible with respect to C_{ch}.

In other words, any candidate for chance codified by the function *ch* is required to play the role of an *expert* function whose expert assignment over $\mathcal{A}$ constrains one's credence function over $\mathcal{A}$, provided that one has admissible evidence.[9] And evidence E is admissible with respect to C_{ch} if it gives no information about the truth of each A that does not go through the chance of A. We say that such evidence is *screened off* by chance.

Importantly, Lewis took the expert role given by his Principal Principle to be a condition for any adequate theory of chance, including his Humean theory of chance:

> If Humean Supervenience is true, then contingent truths about chance are in the same boat as all other contingent truths: they must be made true, somehow, by the spatiotemporal arrangement of local qualities. How might this be? Any satisfactory answer must meet a severe test. The Principal Principle requires that the chancemaking pattern in the arrangement of qualities must be something that would, if known, correspondingly constrain rational credence. Whatever makes it true that the chance of decay is 50% must also, if known, make it rational to believe to degree 50% that decay will occur. (Lewis 1994, p. 476)

I propose to think of Lewis's account as a functional account of chance (hereafter, *chance functionalism*). Broadly speaking, this is the view that chance is, first

[8] One of Lewis's assumptions is that chance is time-dependent, and so the proposition about chance should be written as C_{ch_t}. For the purposes of this chapter, I dispense with this assumption, for nothing to be presented here essentially hinges on it.

[9] For the classical account of experts, see Gaifman (1988). Like Gaifman, I understand the term "expert" very broadly, so that an expert may be a person, a stochastic theory yielding chance, or even a mechanical device, e.g. a thermometer.

and foremost, constituted in terms of its functional roles. As it is easy to see, Lewis suggests a particular sort of chance functionalism, since he defines only one role of chance, namely its expert role.

Could there be other roles that chance should play? As it has been shown in chapter 1, on a closely related view defended in Schaffer (2007), the answer is in the affirmative. Recall that Schaffer argues that chance is what chance does, and a given probability function is chance if it plays not only the expert role. In addition, to count as chance a probability function should play the roles described by principles that connect chance to possibility, lawhood, futurity, intrinsicness, and causation. For example, chance is whatever grounds realizing possibilities: if there is a non-zero chance that my ticket wins a lottery and the proposition that my ticket wins the lottery is true, then there is a possible world in which the ticket wins that lottery; this chance is said to be grounded by that world and not grounded by a possible world in which the ticket wins some other lottery.

Although I follow Lewis in claiming that the Principal Principle captures a central function of chance, I take it, *pace* Lewis, that the expert role given by the Principal Principle is not the only role that chance should play. Like Schaffer, I take it that chance is whatever plays the expert role together with some other roles. To the chance roles defined by Schaffer, let me add the chance roles codified in the following, fairly intuitive principles:

> **Explanatory Role:** The chance of some proposition A should help explain the frequency with which A comes out true in a sequence of trials conducted on some experimental set-up.
>
> **Frequency Connection:** The frequency with which A comes out true in a sequence of trials conducted on some experimental set-up provides evidence for the chance of this set-up resulting in A.
>
> **Frequency Tolerance:** In any finite sequence of trials on some experimental set-up, no matter how long, the frequency with which A comes out true in that sequence may diverge from the chance of that set-up resulting in A.

It is important, however, to emphasize one point of disagreement between Schaffer's chance functionalism and the version of chance functionalism defended here.

Unlike Schaffer, I do not assume that whatever plays these chance roles is already a probability function, and hence satisfies formal adequacy. Instead, my goal in this chapter is to show that formal adequacy follows from these chance roles. Specifically, I explore one particular chance role—the expert role—and show that whatever plays this chance role must also be formally adequate.

Continuing to chance functionalism itself, let me characterize some of its salient features. The first important feature is that, like many other types of functionalism (e.g. about the mental, colours or moral concepts), chance functionalism lends support for the *multiple realizability or satisfiability thesis*. The thesis says that there are multiple satisfiers of the chance roles. So, for example, a coin's 50% chance of landing heads on a trial may be realized by the coin's causal tendency or propensity to produce heads and tails, or by symmetries inherent to the coin's constitution, or by certain frequencies of heads and tails in a sequence of tosses of this coin. And, what determines whether those potential realizations are chances is not that they have certain intrinsic features, e.g. are frequencies or propensities of a certain sort, but whether they fill the chance roles, e.g. whether they constrain rational credence in the manner prescribed by the Principal Principle. In other words, once we accept the multiple realizability/satisfiability thesis, the temptation to think of chance as *reducible* to only one sort of satisfier loses its allure; as far as the functional role is concerned it is left open what the nature of the occupants of this role is.

The second feature worth mentioning is that our functional analysis of chance tells us what is required for having a mastery of the concept of chance, and as such is still silent on whether we can provide a *functional definition* of chance. But this possibility is not blocked. In what follows, I explore the well-known Ramsey-Lewis method for construing functional definitions.[10] This method has become a standard way of defining terms for functionalists of all varieties.

Consider the following sentence:

(F) Chance constrains credence, and chance explains relative frequencies, and chance fits a lawfully projected quantity, and ...

This sentence is a conjunction of sentences expressing various roles played by

[10]For a presentation of this method, see Lewis (1970).

chance. Now, we "Ramseify" (F) by replacing every "chance" in it by a variable ϕ, so that we obtain the following sentence:

(FR) $(\exists\phi)$(ϕ constrains credence, and ϕ explains relative frequencies, and ϕ fits a lawfully projected quantity, and...)

Next, we can use (FR) to define what is for chance to be assigned to an event:

(D) For any event e, we assign chance to $e = (\exists\phi)$(ϕ constrains credence in e, and ϕ explains relative frequencies with which e occurs, and ϕ fits a lawfully projected quantity assigned to e, and...) and we assign ϕ to e.

Essentially, this definition says that we assign chance to some event just in case there is some variable ϕ that fills those various roles specified by (F). From this perspective, the question of what entities in the world satisfy those chance roles is not important. Moreover, it might likely be that many different entities satisfy those roles. If so, the functionalist would count any of them as chance.

Before closing this section, let me briefly compare chance functionalism with the analyses of chance given by frequency and propensity theories. As indicated in chapter 1, these theories aim at analysing the concept of chance in terms of more fundamental or base concepts. Recall that various frequency theorists of chance identify chance with either actual finite relative frequency (Venn 1866), limiting relative frequency (Reichenbach 1949; von Mises 1957), or with hypothetical limiting relative frequency (Kyburg 1974; van Fraassen 1979) of the occurrences of some event in a certain reference class. Similarly, various Humean best-system theories of chance gives an analysis of the concept of chance in terms of regularities posited in stochastic laws of the best-system of truths, i.e. a system that receives the optimal balance between fit, informativeness, and simplicity. Propensity theorists postulate that a different kind of property may be used to analyse chance, to wit, disposition, propensity, or causal tendency. And it is typically taken to be a property of some experimental system to produce long-run relative frequencies, or to produce outcomes on a single trial.

In contrast to the theories of chance given above, the version of chance functionalism just given does not purport to provide an analysis of the concept of chance in terms of some more fundamental or graspable vocabulary. In particular, it does not identify chance with some specific property of an experimental system. There is also good reason to think that chance functionalism is a viable alternative to such theories of chance. For there are serious reasons to doubt whether an analysis of chance either in terms of propensities or frequencies can succeed (Hájek 1996, 2009; Eagle 2004), and chance functionalism simply bypasses all the familiar charges against such theories.

Given chance functionalism, how should one tackle the question of formal adequacy? There seems to be at least two strategies. First, one may add the condition of formal adequacy to the list of chance roles, and argue that chance is whatever satisfies this condition. As suggested above, this strategy has been employed by Schaffer. Second, one may try to show that formal adequacy follows from some or all of the other chance roles. In this chapter, I employ a version of the second strategy. In the next sections, I will show that whatever plays the expert role given by the Principal Principle must also be a finitely additive probability function.

2.4 The Candidate for Chance

In this section, I characterize more precisely the real-valued function *ch* that is required to fill the expert role codified by the Principal Principle. Here this function is regarded as an abbreviated description of a candidate for chance, i.e. any function that is required to satisfy the Principal Principle, or simply a *candidate chance function.* I introduce a framework which remains neutral as to whether this function represents a measure of propensity, relative frequency, or Humean (best-system) chance. I also argue that this framework is able to encompass both *reductive* and *non-reductive* metaphysical theories of chance. Roughly stated, while the former takes facts about chances to be reducible to some non-modal facts, the latter takes facts about chances to be ontologically on a par with these non-modal facts.

To start with, let a single outcome of some chancy process (e.g. tossing a coin) at time t, $t = 1, ..., N$, be denoted by the variable a_t taking one of the values in the set $\{0, 1\}$, where 1 indicates that an outcome occurs and 0 that it does not. Denote by $\{0, 1\}$ the set of all possible outcomes a_t at a given time t. Now consider $w = a_1a_2...a_N$, which is a finite N-long binary sequence consisting of time-indexed outcomes a_t, say a sequence of heads and tails in the coin tossing up to and including time $t = N$. For example, if 1 stands for "heads" and 0 for "tails", then the ordered sequence $w_4 = \langle 1100 \rangle$ indicates the fact that on the first two tosses the coin landed heads and on the next two tails. Let $\mathcal{W}^N$ be the set of all N-long binary sequences of time-indexed outcomes; it is the Cartesian product of N copies of $\{0, 1\}$. The set $\mathcal{W}^N$ has 2^N elements. We will think of the sequences w as possible worlds in a logical space $\mathcal{W}^N$.

We may now define sets within $\mathcal{W}^N$ called *cylinder sets*. The basic category of cylinder sets in $\mathcal{W}^N$ are the propositions denoted by A_t^a. Each such proposition is a set of all those sequences or worlds w that agree on the outcome a at time t for some t-th element $w(t)$ in a sequence w. More precisely:

$$A_t^a = \left\{ w \in \mathcal{W}^N : w(t) = a \right\}. \tag{2.3}$$

So, for example, the proposition A_4^1 means that at time $t = 4$ the outcome is $a = 1$, and is identified with all those worlds w in which at time $t = 4$ the outcome $a = 1$ occurs, that is, $A_4^1 = \{w : w(4) = 1\}$. Bear in mind that what happens before and after time $t = 4$ is irrelevant to identifying these sets at time $t = 4$, that is, before and after $t = 4$ these worlds may exhibit a different sequence of outcomes.

There is also a special category of propositions that can be defined as cylinder sets in our framework. These are the historical propositions that specify the "complete" history of outcomes up to some time t. We denote these propositions by $H_t^{\langle a_1 \dots a_t \rangle}$, where $\langle a_1 ... a_t \rangle$ refers to some finite history of outcomes up to time t, $t < N$. We define $H_t^{\langle a_1 \dots a_t \rangle}$ as sets of worlds in $\mathcal{W}^N$ that for any time $t' \leq t$ share the finite sequence of outcomes $\langle a_1 ... a_t \rangle$ whose t-th element is denoted by $w(t)$. More precisely:

$$H_t^{\langle a_1 \dots a_t \rangle} = \left\{ w \in \mathcal{W}^N : \forall t' \leq t \quad (w(t') = a_{t'}) \right\}. \tag{2.4}$$

For example, consider the historical proposition $H_2^{\langle 11 \rangle}$ at time $t = 2$ characterized by the finite sequence of outcomes $\langle 11 \rangle$. On our account, this proposition is identified with the set of worlds $w \in \mathcal{W}^N$ in which, for $t = 1$ and $t = 2$, $w(t_1) = 1$ and $w(t_2) = 1$. In due course I will omit, when unnecessary, the reference to a

and $\langle a_1...a_t \rangle$ in A_t^a and $H_t^{\langle a_1...a_t \rangle}$ respectively, writing just A_t and H_t.

Each H_t can be understood as intersections of the sets $A_{t'}$. More generally:

$$H_t = \bigcap_{t'=1}^{t} A_{t'}. \tag{2.5}$$

We take it that each history H_t up to some time t is extendable by adding propositions specifying subsequent outcomes A_{t+1}, thus $H_{t+1} = H_t \cap A_{t+1}$. This means, in fact, that after each extension by A_{t+1}, each previous history H_t is narrowed down to the smaller set, $H_{t+1} \subsetneq H_t \subsetneq H_{t-1}$.

We can also make sense of propositions in $\mathcal{W}^N$ that describe global histories, i.e. past, present, and future history of outcomes. Each such proposition, denoted by H_N, is given by:

$$H_N = \bigcap_{t=1}^{N} A_t. \tag{2.6}$$

Let $\mathcal{A}$ be a finite algebra generated by the cylinder sets A_t^a in $\mathcal{W}^N$. This is the smallest algebra containing the cylinder sets in $\mathcal{W}^N$.[11] We can define the function is required to fill the expert role as a set function ch over $\mathcal{A}$, $ch : \mathcal{A} \to [0,1]$. It follows that the function $ch : \mathcal{A} \to [0,1]$ uniquely determines the function $ch : \mathcal{W}^N \to [0,1]$, i.e. the chance function over all the elements ("thin" cylinders) of $\mathcal{W}^N$. This is a function defined over the most fine-grained partition of $\mathcal{W}^N$, i.e. the set of all singleton propositions $\{w\}$ for $w \in \mathcal{W}^N$. Does the converse hold? Can we say that the function $ch : \mathcal{W}^N \to [0,1]$ is uniquely extendible to all the subsets of $\mathcal{W}^N$? A positive answer to this question would require the following to hold:

$$\forall A_t \subset \mathcal{W}^N : \quad ch(A_t) = \sum_{w \in A_t} ch(\{w\}). \tag{2.7}$$

However, there is reason to suspect that condition (2.7) does not hold until the formal adequacy of ch is established. For it requires *finite additivity* to hold, and we do not know yet whether or not ch is finitely additive. As we will see later on, the function $ch : \mathcal{W}^N \to [0,1]$ plays a crucial role in providing a geometrical

[11] The algebra $\mathcal{A}$ over $\mathcal{W}^N$ is a set of subsets of $\mathcal{W}^N$ that contains $\mathcal{W}^N$, $\varnothing$, and is closed under complementation and union, i.e (i) if $A_t \in \mathcal{A}$, then $\mathcal{W}^N - A_t \in \mathcal{A}$ and (ii) if $A_t, B_t \in \mathcal{A}$, then $A_t \cup B_t \in \mathcal{A}$.

representation of ch which is needed for making the talk of "distances" between chance distributions and the truth-value distributions graspable. For this reason it seems prudent first to address the question of formal adequacy of the function $ch : \mathcal{W}^N \to [0,1]$. Once its formal adequacy is established, one can easily generalize it into the function $ch : \mathcal{A} \to [0,1]$ by using (2.7).

As advertised at the beginning of this section, the framework just described is capable of representing both reductive and non-reductive metaphysical accounts of chance. A reductionist about chance holds that chances are reducible to certain other, metaphysically prior non-modal features of the world. A non-reductionist about chance holds that they are not: that, ontologically, chances are on a par with other non-modal features of the world. Recall that within our framework, a set of these non-modal features at a given time t is given by the set $\{0,1\}$, that is, the set of chancy outcomes. But how should these non-modal features be understood? Lewis (1994), for example, characterised them as perfectly natural properties (non-gerrymandered), possessed intrinsically by space-time points or occupants thereof. Whether or not we accept this characterization, we should keep in mind that what matters for a reductionist about chance is that these properties cannot involve chances, dispositions, or causal tendencies.

Now, suppose that the function $ch : \mathcal{A} \to [0,1]$ represents a feature of the world that is reducible to these non-modal features. For concreteness, let's focus on one stripe of reductionism about chance, namely on Lewis's Humean theory of chance. Let C_{ch} be the proposition that the real-valued assignment over $\mathcal{W}^N$ is given by the function ch. C_{ch} is true at a world w if and only if $ch = ch_w$. That is, C_{ch} is true at those worlds at which the function ch and ch_w give the same real-valued assignment over the singleton propositions $\{w\}$ in $\mathcal{W}^N$. In every such world, Humean chance function ch is identical with the function ch_w, which is the so-called *initial* or *ur*-chance function at w, i.e. the chance function at the beginning of w's history. What are the worlds at which C_{ch} is true? The answer to this question is provided by the metaphysical doctrine of Humean Supervenience (HS). In general, HS states that modal concepts such as laws of nature, dispositions, counterfactuals, causation, or chances supervene on the global histories of non-modal properties. Applied to chances, HS can be formulated within the framework given above as follows:

HS$_{\textbf{chance}}$: For any worlds w and w' in $\mathcal{W}^N$, if w and w' have the same global (past, present, and future) histories of chancy outcomes, then w and w' have the same chance function ch.

For example, suppose that at worlds w and w' some chancy process takes place and instantiates, on each of its ten trials, one of the two possible non-modal properties denoted by 0 or 1. Suppose that the pattern of these instances is the same at these worlds, i.e. w and $w' \in H_N$. By the HS$_{chance}$, the worlds have the same ch. Given that a summary of the pattern is a theorem of the best system of truths, ch is fixed, at each of these worlds, by the same T, which is a set of history-to-chance conditionals that are true at w and w'. These conditionals have the status of stochastic laws of the best-system theory for w and w', i.e. the best summary of the true pattern at w and w'. Lewis characterised them as follows:

> (1) The consequent is a proposition about chance at a certain time. (2) The antecedent is a proposition about history up to that time; and further, it is a complete proposition about history up to that time, so that it either implies or else is incompatible with any other proposition about history up to that time. It fully specifies a segment, up to the given time, of some possible course of history. (3) The conditional is made from its consequent and antecedent not truth-functionally, but rather by means of a strong conditional operation of some sort. (Lewis 1986, p. 94)

Further, the chance function at time t, ch_t, comes from the chance function ch by conditionalizing on the history up to t, H_t; more precisely:

$$ch_t(-) = ch(-|H_t). \tag{2.8}$$

Therefore, ch_t is always fixed by the history H_t at t and the stochastic laws in T. Importantly, once ch is shown to be formally adequate, ch_t must also be formally adequate. Since ch_t comes from ch by conditionalizing on the history, formal adequacy is "carried over" to it from ch.

But the function $ch : \mathcal{W}^N \to [0, 1]$ might well represent a feature of the world that is not reducible to the non-modal features. Non-reductionists about chance, most notably propensity theorists, reject HS$_{chance}$ and claim that different chance distributions could give rise to the same global history of chancy outcomes. That is, chance distributions do not depend ontologically on such histories. At best, such histories provide evidence for these chance distributions.

2.5 Playing the Expert Role: A Necessary Condition

It seems rather uncontroversial to say that if a theory or a person is an expert for you, it will be incoherent if you also believed that this theory or person is not reliable. In this section, I argue in a similar spirit that if the function *ch* is as an expert for you, it must be the case that you believe it to be predictively accurate in some definite way. More specifically, it is shown that the agent sets her credence function over $\mathcal{W}^N$ equal to a real-valued assignment over $\mathcal{W}^N$ given by *ch* if she believes that *ch* is *weakly predictively accurate*, i.e. it is such that there is no other function ch' over $\mathcal{W}^N$, known to her, that could match better all the possible truth-value distributions over $\mathcal{W}^N$. Although the condition of weak predictive accuracy may not suffice to play the expert role, it is argued that it is necessary to play this role. It is shown that for *ch* to not ruin a priori the possibility of being an expert function, it must satisfy the condition of weak predictive accuracy. For if this function does not satisfy this condition, there is an epistemic sense in which your credences over $\mathcal{W}^N$ fixed by that function can be "outperformed" by credences fixed by some other candidate chance function that you might adopt instead.

This section is organized as follows. First, I explain what it means when we say that the function *ch* matches a possible truth-value distribution over $\mathcal{W}^N$. In so doing, the notion of *ch*'s predictive accuracy is introduced and explained. Second, I introduce the notion of weak predictive accuracy, and show why it is a necessary condition for *ch* to play the expert role. Third, I show why the condition of weak predictive accuracy in the case of chances is superior over other two intuitive conditions that might be formulated in terms of predictive accuracy.

2.5.1 Chance and Predictive Accuracy: Brier Score

For each time t, the set $\mathcal{W}^N$ contains singleton propositions describing possible outcomes of a chancy process. Given that the truth-value distributions over these propositions fully characterize the possible worlds in $\mathcal{W}^N$, we can ask how well the predicted values given by ch match these truth-value distributions. The extent to which the values predicted by ch accord with a given truth-value distribution at w will be called ch's *predictive accuracy* at w. I take the predictive accuracy of ch to be a measurable property. Mathematically, a measure of predictive accuracy is a function of ch and w whose value indicates how much ch diverges from w. I will refer to this measure as a *scoring function*.

Due to its pivotal role in my argument, let me describe the scoring function more precisely. If $\mathcal{C}$ is a finite set of candidate chance functions over $\mathcal{W}^N$ and $\mathcal{W}^N$ is a set of possible worlds, then our scoring function may be defined as a real-valued function $S : \mathcal{W}^N \times \mathcal{C} \to \mathbb{R}$ which, for each pair (w, ch) of $ch \in \mathcal{C}$ and $w \in \mathcal{W}^N$, gives the predictive accuracy score $S(w, ch)$. This score measures ch's predictive accuracy at w, which is the extent to which it diverges from the world w. If $S(w, ch) < S(w, ch')$, we say that ch is less divergent from the world w than ch', and thus is more predictively accurate at w than ch'.

Our scoring function S may be also characterized as follows. Given the characteristic function v_w of w, i.e. a function such that for each $\{w_j\}$ in $\mathcal{W}^N$,

$$v_w(\{w_j\}) = \begin{cases} 1 & \text{if } w \in \{w_j\}\,, \\ 0 & \text{if } w \notin \{w_j\}\,, \end{cases} \tag{2.9}$$

and the value $ch(\{w_j\}) \in [0, 1]$ for each singleton proposition $\{w_j\}$, let $s : \{0, 1\} \times [0, 1] \to \mathbb{R}$ be a *scoring rule*. Then, our scoring function is given by

$$S(w, ch) = \sum_{j=1}^{2^N} s\Big(v_w(\{w_j\}), ch(\{w_j\})\Big). \tag{2.10}$$

That is, whereas $s\Big(v_w(\{w_j\}), ch(\{w_j\})\Big)$ measures how the individual chance, $ch(\{w_j\})$, diverges from the individual truth value, $S(w, ch)$ measures how the whole function ch over $\mathcal{W}^N$ diverges from the whole truth-value distribution over $\mathcal{W}^N$. We say that $S(w, ch)$ is generated by the scoring rule s.

But how could the predictive accuracy score be calculated? From now on, I confine my attention to a particular predictive accuracy score called the *Brier* *score*:[12]

> **Brier Score:** Suppose that $S : \mathcal{W}^N \times \mathcal{C} \to \mathbb{R}$ is a scoring function for each pair (w, ch). Then, $S(w, ch)$ is said to be the Brier score if and only if for any world $w \in \mathcal{W}^N$ and any $ch \in \mathcal{C}$,
>
> $$S(w, ch) = \frac{1}{|\mathcal{W}^N|} \sum_{j=1}^{2^N} \Big(v_w(\{w_j\}) - ch(\{w_j\}) \Big)^2,$$
>
> where $|\mathcal{W}^N|$ stands for the number of elements in $\mathcal{W}^N$.

The Brier score thus construed gives a score in the form of a sum of squares of distances between the characteristic function $v_w(\{w_j\})$ of a singleton proposition $\{w_j\}$ in a world w and the value $ch(\{w_j\})$ assigned to this proposition, given an equal weight $\frac{1}{|\mathcal{W}^N|}$ assigned to every world $w \in \mathcal{W}^N$.

Proceeding *more geometrico*, each world $w \in \mathcal{W}^N$ where $\mathcal{W}^N = \{w_1, ..., w_{2^N}\}$, can be represented by the 2^N-vector of truth values δ_w having entries $v_w(\{w_j\}) = 1$ if $w \in \{w_j\}$ and 0 otherwise. The vector δ_w represents the truth-value distribution over the propositions $\{w_1\}, ..., \{w_{2^N}\}$ if the world $w \in \mathcal{W}^N$ is actual. This vector has 1 at its jth place and 0 everywhere else if and only if the proposition $\{w_j\}$ is true in the world w.

Likewise, the function ch over $\mathcal{W}^N$ can be represented by a vector $c = (ch(\{w_1\}), ..., ch(\{w_{2^N}\}))$. The vector c represents the candidate chance distribution over all the singleton propositions $\{w\}$ for $w \in \mathcal{W}^N$. Since we do not know yet whether ch is formally adequate, we cannot say whether $\sum_{j=1}^{2^N} ch(\{w_j\}) = 1$.

Once the possible worlds and candidate chance functions are positioned as vectors in the 2^N-dimensional Euclidean space, we can determine distances between them. Let $\|\delta_w - c\|$ be the Euclidean distance between any two vectors w_i and ch. Such Euclidean distance is always a non-negative quantity defined given by:

$$\|\delta_w - c\| = \sqrt{(v_w(\{w_1\}) - ch(\{w_1\}))^2 + ... + (v_w(\{w_{2^N}\}) - ch(\{w_{2^N}\}))^2}. \quad (2.11)$$

[12]The Brier score—a particular sort of quadratic scoring rule—is named after Glenn Brier (1950), who developed it to measure the accuracy of probabilistic weather forecasts.

The divergence of ch from a world w_i, then, may be represented as a distance thus construed. If the distance of c from δ_w is lesser than the distance of c' from δ_w, we say that ch is more predictively accurate than ch'. More precisely, for any ch, ch' and a world w_i:

$$S(w, ch) < S(w, ch') \iff \|\delta_w - c\| < \|\delta_w - c'\|, \tag{2.12}$$

where $|w_i - ch|$ and $|w_i - ch'|$ stand for Euclidean distances. Within our geometrical representation, the Brier score may be defined as the squared Euclidean distance:

$$S(w, ch) = \|\delta_w - c\|^2. \tag{2.13}$$

It is natural to ask whether there are some special reasons that led me to adopt the Brier score rather than some other measure of predictive accuracy. This is an important issue, for it might be claimed that the whole argument is essentially Brier scoring rule-dependent and thus we need a justification of why this scoring rule is suitable for the task of measuring the predictive accuracy of candidate chance functions. Since my answer to this question involves a few subtle points, I shall postpone it until section 2.7.

2.5.2 Weak Predictive Accuracy and the Expert Role

Suppose that epistemic rationality requires an agent to approximate an "epistemically ideal" credence function over the singleton propositions in $\mathcal{W}^N$. Given a possible world w, the "epistemically ideal" credence function at w is the one that assigns maximal credence, i.e. 1, to propositions that are true at w and minimal credence, i.e. 0, to propositions that are false at w. Thus, for each $w \in \mathcal{W}^N$, such credence function is the characteristic function v_w at w. Many philosophers have argued that proximity to the "epistemically ideal credence" function is *a* fundamental requirement of epistemic rationality for the agent's credence function (see, e.g. Joyce 1998, 2009; Pettigrew 2013b). And even if there are other requirements of epistemic rationality, there are good reasons to believe that they are consistent with this fundamental requirement.[13]

[13] For example, as argued in Pettigrew (2013b), the goal of proximity to the "epistemically ideal credence" and the goal of matching one's credences with one's evidence are not only compatible, but may lead to the same norms for credences.

It seems natural to think that the agent will pursue the goal of approximating such epistemically ideal credence function by adopting the credence function that is best justified in the light of her evidence, including evidence about chances. Let me use the following personification for heuristic purposes. Suppose that the agent has at her disposal a whole "panel" of functions that are the candidate expert functions concerning $\mathcal{W}^N$. We assume that information given by these functions is the only relevant information about the propositions' truth values. Let this "panel" be represented by a set $\mathcal{C}$. Assume also that the agent has no inadmissible evidence. In particular, she does not possess a crystal ball-like evidence that predicts correctly the truth values of all the propositions in $\mathcal{W}^N$. With these assumptions in mind, the question arises: what is required for any function ch in $\mathcal{C}$ to count as an expert concerning $\mathcal{W}^N$ for the agent's credence function?

My answer is that ch must satisfy some condition whereby it deserves, by the agent's lights, the status of expert. That is, our agent would set her credence function equal to the function ch in $\mathcal{C}$ if she knew that ch satisfies that condition. Following Hall (2004, pp. 102-104), we can make this point more precise. Let the agent's admissible evidence E be the proposition that ch meets the required condition, and let C_{ch} be the proposition that the real-valued assignment over $\mathcal{W}^N$ is given by ch. Then, for all singleton propositions in $\mathcal{W}^N$,

$$cr(\{w\}|E \cap C_{ch}) = ch(\{w\}|E), \tag{2.14}$$

which is equivalent to the PP, if ch is "certain" that it satisfies the required condition E, i.e. that $ch(E) = 1$.

But what could this condition be? By the agent's lights, it must be a condition whose satisfaction is necessary for ch to generate a rational credence function as defined by the agent's epistemic rationality. This, in fact, means that by satisfying this condition the chance function (i) generates a credence function which approximates the epistemically ideal credence function, and (ii) makes the agent's credence function *resilient* in the following sense: once the agent's credence function is set equal to ch, she is not epistemically compelled to move to some other ch' in $\mathcal{C}$, which is known to her, on pain of violating epistemic rationality. What is important is that we can reveal such a condition when we focus on *ch*'s predictive accuracy. Obviously, not any kind of *ch*'s predictive accuracy would be suitable for this task. If some proposition $\{w\}$ happens to be true at a world w and ch assigns to it a very low value, then it is hardly informative about the proposition's truth values at w. Here the normative force of the injunction to move to some other chance distribution, say ch', will be quite strong. But what I call the *weak predictive accuracy* is perfectly suitable to be such a necessary condition. One may state this condition, more precisely, as follows:

> **Weak Predictive Accuracy:** Suppose that $S(w, ch)$ is the Brier score. Then, $ch \in \mathcal{C}$ is said to be weakly predictively accurate if and only if there is no other $ch' \in \mathcal{C}$ such that for all $w \in \mathcal{W}^N$,
>
> $$S(w, ch') \leq S(w, ch).$$

In other words, ch is weakly predictively accurate just in case there is no other ch' in $\mathcal{C}$ that could have a better score of predictive accuracy in *every* possible world. Here the operating idea is that no matter which world turns out to be actual, ch must be such that no other ch' could outperform it in reducing the distance to all the possible worlds. The weak predictive accuracy of ch so construed assures that there is no other chance function, known to the agent, that could be a better guide to the truth. That is, there is no other chance among the "panellists" that could be more informative about the truth values of propositions over which the agent's credences are distributed. Hence, the agent who adjusts her credence function to such a chance function would maximize its closeness to the truth, and would not be compelled to move to some other chance function, on pain of epistemic irrationality.

To make this point more vivid, suppose that (i) there are two singleton propositions $\{w_1\}$ and $\{w_2\}$, (ii) there are three candidate chance functions ch_1, ch_2, and ch_3 in $\mathcal{C}$, whose predictions are represented by three different vectors $ch_1 = \left(\frac{1}{4}, \frac{1}{4}\right)$, $ch_2 = \left(\frac{1}{2}, \frac{1}{2}\right)$, and $ch_3 = (1, 0)$, (iii) there are two possible worlds w_1 and w_2 that might be represented by two vectors $\delta_{w_1} = (1, 0)$ and $\delta_{w_2} = (0, 1)$ respectively, and (iv) for each $ch \in \mathcal{C}$ and $w \in \mathcal{W}^N$, the Brier score gives the

Table 2.1: An Example of the Brier Scores for Chances

ch	$\{w_1\}$	$\{w_2\}$	$S(w_1, ch)$	$S(w_2, ch)$
ch_1	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{5}{16}$	$\frac{5}{16}$
ch_2	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$
ch_3	1	0	0	1

score $S(w, ch) = \frac{1}{2}\Big(v_w(\{w_1\}) - ch(\{w_1\})\Big)^2 + \Big(v_w(\{w_2\}) - ch(\{w_2\})\Big)^2$. Table 2.1 summarizes the results of assessing the predictive accuracy of our three candidate chance functions. Could ch_1 be an expert function? Given that it is outperformed by ch_2 in both w_1 and w_2, it would generate a credence function which does not maximize the closeness to the truth, and it would make the agent epistemically compelled to move to some other "panellist". Consequently, it cannot constrain the agent's credence, on pain of epistemic irrationality. In this example, only ch_2 and ch_3 are weakly predictively accurate because there is no other function that could outperform them in w_1 and w_2. Thus, by the agent's light, they both deserve the status of expert.

2.5.3 Other Notions of Predictive Accuracy

Could considerations concerning the predictive accuracy of chance reveal other conditions whose satisfaction might be required to play the expert role? Two such conditions are worth noticing. The first condition may be called the *Laplacean predictive accuracy*. We say that ch is Laplacean predictively accurate just in case ch is the actual truth-value distribution over $\mathcal{W}^N$. I call this notion of predictive accuracy Laplacean, for it bears a close resemblance to Pierre Laplace's idea, famously articulated in "A Philosophical Essay on Probabilities":

> Given for one instant an intelligence which could comprehend all the forces by which nature is animated and the respective situation of the beings who compose it—an intelligence sufficiently vast to submit these data to analysis—it would embrace in the same formula the movements of the greatest bodies of the universe

> and those of the lightest atom; for it, nothing would be uncertain and the future, and the past, would be present to its eyes. (Laplace 1951, p. 4)

To put it somewhat anthropomorphically, *ch* is Laplacean predictively accurate if and only if it can be portrayed as Laplace's demon, i.e. it "knows" which one of the possible worlds is actual. On this account, chance is always an "opinionated" expert with categorical predictions that might be represented numerically by the truth values 0 and 1. Moreover, it is an "opinionated" expert whose categorical predictions, together with the background assumptions, deductively entail the truth or falsehood of a given proposition. Since it knows in advance which of the possible worlds is actual, it makes no sense to speak about the distance between its predictions and the possible truth-value distributions. Though initially attractive, it seems to be too strong a condition for counting *ch* as expert. Many well-confirmed stochastic theories (e.g. population genetics in biology) deserve to be called experts, even though their predictions confer on the propositions some non-extreme values from the unit interval of real numbers. We do not discard such theories as experts just because they cannot tell us what will happen for sure. So a stochastic model of the evolution of some population may be a good predictor of a change in trait frequency, even though at the micro-physical level this process is purely deterministic.

Another important problem facing this condition is that it construes the notion of predictive accuracy in an *absolute* manner. That is, it holds that to be predictively accurate chance must reach some *threshold* in reducing the distance to the truth. But once we agreed that the threshold given by the truth values of a proposition is not applicable to chances, there seems to be no non-arbitrary way to say what the threshold should be. The notion of weak predictive accuracy avoids this problem because it is couched in a comparative way. We ask when a chance is predictively accurate when compared to other chances.

The second condition may be called the *strong predictive accuracy*. We say that *ch* is strongly predictively accurate just in case *ch* matches all possible truth values of the propositions in $\mathcal{W}^N$ better than any other ch' in $\mathcal{C}$. Unlike the condition of Laplecean predictive accuracy, this condition is construed in a *comparative* manner and thus avoids the pitfalls of the absolute reading of predictive accuracy. Moreover, this condition is more liberal than the Laplacean one. For a chance that assigns to propositions in its domain some non-extreme values from the unit interval of real numbers may be regarded as an expert. The only condition is that it must be the best chance in reducing the distance from all possible worlds, i.e. it must yield a prediction that is closest to all possible worlds. Of course, to be philosophically significant, this condition must specify whether "best" refers to "actually best" or maybe to "possibly best". But even if this issue is solved, there is another and more serious wrinkle that besets this condition. It is perfectly possible that for any set containing the purported expert chances, there might be no chance that outperforms all the others in reducing the distance from all possible worlds. This may be due to the infamous problem of underdetermination of theories by evidence. The condition of strong predictive accuracy implies that unless some other criteria determining our choice of these chances are used (e.g. simplicity, *ad hoc*ness, explanatory power), one faces a surprising stalemate: none of these chances could be considered as an expert for you. But it sounds at least odd to say that stochastic theories cannot be considered as experts just because they are predictively equivalent. The oddity evaporates when we recognize that predictively equivalent theories may still be candidates for experts, even though we cannot decide which one of them is the best with respect to predictive accuracy.

The notion of weak predictive accuracy avoids this oddity. The example discussed in subsection 2.5.2 reveals, quite intuitively, that ch_2 that assigns intermediate values might be an expert on a par with the "opinionated" ch_3. Of course, the condition of weak predictive accuracy does not suffice to tell us which of these chances would constrain an agent's credence. After all, it is a necessary condition. But still there is a way in which ch_2 and ch_3 could constrain the agent's credence in that case. The idea is that one should adjust one's credence to a mixture of the known chances, weighted by one's credences assigned to them. In other words, one should adjust credence to one's best estimate of these chances. Of course, this manoeuvre works only if stochastic theories are mutually exclusive. More precisely, for all singleton propositions in $\mathcal{W}^N$,

$$cr(\{w\}) = cr(C_{ch_1})ch_1(\{w\}) + cr(C_{ch_2})ch_2(\{w\}), \tag{2.15}$$

where C_{ch_1} is the proposition that the chance distribution over $\mathcal{W}^N$ is given by ch_1 and C_{ch_2} is the proposition that the chance distribution over $\mathcal{W}^N$ is given by ch_2.

2.6 Proving Formal Adequacy

In this section, I prove that the expert role demands chance to be formally adequate, i.e. to be a finitely additive probability function. It is shown that, relative to the Brier score, any function ch over $\mathcal{W}^N$, which is a finitely additive probability function, is predictively accurate in the weak sense. But before I state and prove the theorem, let me outline a strategy for deducing formal adequacy of ch from the condition of weak predictive accuracy, to wit, a necessary condition for counting ch as expert function.

Let the function $ch : \mathcal{A} \to [0, 1]$ uniquely determine the function $ch : \mathcal{W}^N \to [0, 1]$. Let $\mathcal{V} = \{v_w : w \in \mathcal{W}^N\}$ be the set of consistent truth-value distributions over the propositions in $\mathcal{W}^N$, i.e. a collection of functions $v_w : \mathcal{W}^N \to \{0, 1\}$. Further, we introduce the set of all convex combinations of the v_w's in $\mathcal{V}$, called the convex hull of $\mathcal{V}$. It can be defined as follows:

$$Conv(\mathcal{V}) = \left\{ \sum_{w \in W} \lambda_w v_w : 0 \leq \lambda_w \leq 1, \sum_{w \in W} \lambda_w = 1 \right\}. \tag{2.16}$$

That is, $Conv(\mathcal{V})$ is the smallest set that (i) contains $\mathcal{V}$, and (ii) contains, for any two v_w and $v_{w'}$, every convex combination or mixture of them. i.e. for any $0 \leq \lambda_w \leq 1$, it contains $\lambda_w v_w + (1 - \lambda_w) v_{w'}$.

We say then that $ch : \mathcal{W}^N \to [0, 1]$ is formally adequate if and only if $ch \in Conv(\mathcal{V})$. That is, due to a well-known result of Bruno de Finetti's (De Finetti 1970), finitely additive probability functions are precisely the mixtures of truth-value distributions in $\mathcal{V}$.

How can the formal adequacy thus construed be derived from the condition of weak predictive accuracy? Suppose again that you assess the predictive accuracy of the functions in $\mathcal{C}$. Assume that there are some functions that are not formally adequate. Then, the first step towards a vindication of formal adequacy is to show that the function ch, which is *not* formally adequate, can be outperformed by a formally adequate one in reducing the distance to all possible truth-value distributions over $\mathcal{W}^N$. That is to say, any function ch that is not formally adequate is *not* predictively accurate in the weak sense because there is always another function ch'—a formally adequate one—that could match better every possible world:

> $\neg$ **Formal Adequacy $\Rightarrow$ $\neg$ Weak Predictive Accuracy:** Let ch be a candidate chance function over $\mathcal{W}^N$. Then, if $ch \notin Conv(\mathcal{V})$, then there is another function $ch' \in Conv(\mathcal{V})$ such that
>
> $$\forall w \in \mathcal{W}^N : \quad S(w, ch') \leq S(w, ch).$$

But does ($\neg$ Formal Adequacy $\Rightarrow$ $\neg$ Weak Predictive Accuracy) suffice as an argument for formal adequacy? No, it does not. It only shows that if $ch \notin Conv(\mathcal{V})$, then there is a formally adequate function $ch' \in Conv(\mathcal{V})$ that could outperform ch in reducing the distance to every possible world. This, however, does not preclude the possibility that $ch' \in Conv(\mathcal{V})$ might be outperformed by some other formally adequate function $ch^* \in \mathcal{C}$. To make this point more vivid, suppose that the vectors $\delta_{w_1} = (1, 0)$ and $\delta_{w_2} = (0, 1)$ are two vectors of the truth-value distribution over two mutually exclusive singleton propositions $\{w_1\}, \{w_2\}$. Let c'_1 and c'_2 be two vectors of predicted values given by $ch'_1 \in Conv(\mathcal{V})$ and $ch'_2 \in Conv(\mathcal{V})$ respectively. Although both ch'_1 and ch'_2 are formally adequate, they are formally adequate in a different way; that is to say, $c'_1 = \left(\frac{1}{2}, \frac{1}{2}\right)$ and $c'_2 = (1, 0)$. ($\neg$ Formal Adequacy $\Rightarrow$ $\neg$ Weak Predictive Accuracy) shows that if ch is not formally adequate, then it can be outperformed by some formally adequate function, say either ch'_1 or ch'_2. However, it does not mean that neither of these formally adequate chance functions can be outperformed by the other one. So in order to guarantee that every formally adequate function is weakly predictively accurate, one needs to prove the following:

> **Formal Adequacy $\Rightarrow$ Weak Predictive Accuracy:** Let ch' be a candidate chance function over $\mathcal{W}^N$. Then, if $ch' \in Conv(\mathcal{V})$, then there is no other $ch^* \in \mathcal{C}$ such that
>
> $$\forall w \in \mathcal{W}^N : \quad S(w, ch^*) \leq S(w, ch').$$

With this strategy in mind, we can reach an equivalence between the formal adequacy and the weak predictive accuracy of chances. I shall establish this equivalence by means of the following theorem:

Theorem 2.1. *Suppose that the predictive accuracy of ch over $\mathcal{W}^N$ is scored by the Brier score with the score* $S(w, ch) = \frac{1}{|\mathcal{W}^N|}\sum_{j=1}^{2^N}\Big(v_w(\{w_j\}) - ch(\{w_j\})\Big)^2$. *Then the following two propositions are true:*

(i) If $ch \notin Conv(\mathcal{V})$, then there is another function $ch' \in Conv(\mathcal{V})$ such that

$$\forall w \in \mathcal{W}^N : \quad S(w, ch') \leq S(w, ch).$$

(ii) If $ch' \in Conv(\mathcal{V})$, then there is no other $ch^* \in \mathcal{C}$ such that

$$\forall w \in \mathcal{W}^N : \quad S(w, ch^*) \leq S(w, ch').$$

For the proof, let us introduce the following assumptions:

- Let $ch \in \mathcal{C}$ and $v_w \in \mathcal{V}$ be represented as vectors c and δ_w respectively. We have that $\mathcal{V} \subseteq \mathcal{C}$.
- Let the set $Conv(\mathcal{V})$ be represented by the set

$$\left\{\sum_{w\in\mathcal{W}^N} \lambda_w \delta_w : 0 \leq \lambda_w \leq 1, \sum_{w\in\mathcal{W}^N} \lambda_w = 1\right\},$$

  which is a convex hull of 2^N-dimensional vectors δ_w. Every vector in this set is a vector of finitely additive probability distribution over $\mathcal{W}^N$ denoted by $p = (p(\{w_1\}), ..., p(\{w_{2^N}\}))$, where $p(\{w_j\}) \geq 0$ and $\sum_{j=1}^{2^N} p(\{w_j\}) = 1$. We have that $\mathcal{V} \subseteq Conv(\mathcal{V})$.
- Let $D : \mathcal{C} \times \mathcal{C} \to \mathbb{R}$ be a real-valued function which, for each pair (ch, ch') of $ch \in \mathcal{C}$ and $ch' \in \mathcal{C}$, gives a score $D(ch, ch')$ given by

$$D(ch, ch') = \|c - c'\|^2.$$

The function D measures the Brier score-based distance between any two candidate chance functions ch and ch'. Note that for $ch \in \mathcal{C}$ and $w \in \mathcal{W}^N$,

$$S(w, ch) = D(v_w, ch).$$

Before giving the proof of Theorem 2.1, we state and prove the following technical lemma:

Lemma 2.1. *Let c' be an orthogonal projection $P_\Lambda c$ of c onto $Conv(\mathcal{V})$. Then c' minimizes the distance from c to $Conv(\mathcal{V})$, i.e. for all $c'' \in Conv\,(\mathcal{V})$,*

$$D(ch, ch') \leq D(ch, ch''),$$

where ch, ch', ch' are represented as vectors c, c', c'' respectively.

Proof of Lemma 2.1. Since $c - c'$ is orthogonal to $Conv(\mathcal{V})$, $c - c' \perp Conv(\mathcal{V})$, it also must be orthogonal to $c' - c''$, i.e. $c - c' \perp c' - c''$, for any $c'' \in Conv(\mathcal{V})$. Then, by the Pythagorean theorem, we have that

$$\|c - c''|^2 = \|c - c'\|^2 + \|c' - c''\|^2 \geq \|c - c'\|^2.$$

And thus,

$$D(ch, ch') \leq D(ch, ch''),$$

as required. □

Proof of Theorem 2.1. (i) Suppose that $c \notin Conv(\mathcal{V})$. Then, for the proof of Theorem 2.1 (i), it suffices to show that there is $c' \in Conv(\mathcal{V})$ such that $\|\delta_w - c'\|^2 \leq \|\delta_w - c\|^2$. Hence, $D(v_w, ch') \leq D(v_w, ch)$, and thus $S(w, ch') \leq S(w, ch)$ for all $w \in \mathcal{W}^N$. Suppose that $c' \in \Lambda$. Let $c - c'$ be orthogonal to $Conv(\mathcal{V})$, $c - c' \perp Conv(\mathcal{V})$. Then, c' is an orthogonal projection $P_{Conv(\mathcal{V})}c$ of c onto $Conv(\mathcal{V})$. With Lemma 2.1 in mind, suppose that $c' \neq \delta_w$. Define a line

$$L = \{\lambda c + (1 - \lambda)\, c' : -\infty < \lambda < \infty\},$$

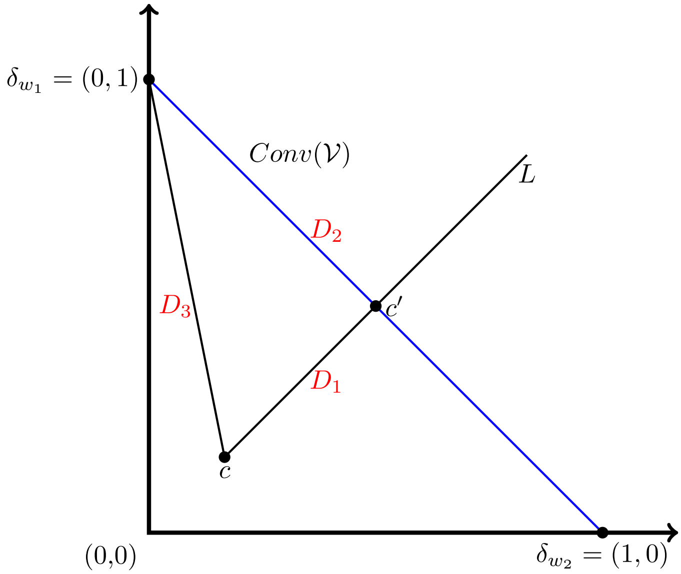

Figure 2.1: Suppose that $c' \in Conv(\mathcal{V})$ is an orthogonal projection $P_L\delta_{w_1}$ of δ_{w_1} onto L, $c \in L$ and $c - c' \perp \delta_{w_1} - c'$. Let D_1 be the distance $\|ch - ch'\|^2$, D_2 be the distance $\|\delta_{w_1} - c'\|^2$, and D_3 be the distance $\|\delta_{w_1} - c\|^2$. Then, by the Lemma 2.1, we have that $D_2 \leq D_3$, as required. The same result can be established for δ_{w_2}.

which is a ray from c through c' to infinity. We know that $\delta_w \notin L$. If so, then we can consider projections of δ_w onto L. Because $c' \in L$ and $\delta_w - c' \perp L$, we have that c' is an orthogonal projection $P_L\delta_w$ of δ_w onto L. Since $c \in L$ and $c - c' \perp \delta_w - c'$, then by the Lemma 2.1 we have

$$D(v_w, ch') \leq D(v_w, ch).$$

Hence,

$$S(w, ch') \leq S(w, ch),$$

as required (see Figure 2.1 for a useful illustration).

(ii)[14] For a reductio, suppose that $c' \in Conv(\mathcal{V})$ and there is c^* such that $\|\delta_w - c^*\|^2 \leq \|\delta_w - c'\|^2$ for all δ_w. Hence, $D(v_w, ch^*) \leq D(v_w, ch')$, and so $S(w, ch^*) \leq S(w, ch')$ for all $w_i \in W^N$. Construe a subset $\mathcal{G} \subseteq Conv(\mathcal{V})$ defined as follows:

$$\mathcal{G} = \{g \in Conv(\mathcal{V}) : D\left(g, ch^*\right) \leq D\left(g, ch'\right)\}.$$

Because the scoring function D is convex, we have that for all $g, g' \in \mathcal{G}$ and for any $0 \leq \lambda \leq 1$,

$$D(\lambda g + (1-\lambda)g', ch^*) \leq D(g, ch^*), D(g', ch^*),$$

and

$$D(\lambda g + (1-\lambda)g', ch') \leq D(g, ch'), D(g', ch'),$$

and thus by the definition of $\mathcal{G}$,

$$D(\lambda g + (1-\lambda)g', ch^*) \leq D(\lambda g + (1-\lambda)g', ch').$$

It follows then that $\mathcal{G}$ must be convex, i.e. for any two g and $g' \in \mathcal{G}$, $\mathcal{G}$ contains every convex linear combination $\lambda g + (1-\lambda)g'$ for any $0 \leq \lambda \leq 1$. Since any convex combination $\lambda g + (1-\lambda)g'$ of two elements in $\mathcal{G}$ is a convex combination $\lambda[\lambda_w v_w] + (1-\lambda)[(1-\lambda_{w'})v'_w]$ of two elements in $Conv(\mathcal{V})$, we have that $\mathcal{G}$ is a convex hull containing $Conv(\mathcal{V})$. Hence, $\mathcal{G} \supseteq Conv(\mathcal{V})$. Since $D(ch', ch') < D(g, ch^*)$, we have that $ch' \notin \mathcal{G}$ and $ch' \in Conv(\mathcal{V})$. But then we have a contradiction because it cannot be true that simultaneously $\mathcal{G} \supseteq Conv(\mathcal{V})$ and $\mathcal{G} \subsetneq Conv(\mathcal{V})$. Therefore, (ii) holds true.

$\square$

To sum up: It has been shown how the expert role demands chance to be a finitely additive probability function. That is, it has been proved that only a finitely additive probability function can satisfy the condition of weak predictive accuracy, and thus can play the expert role.

Note that the result just given has been established for a chance function defined over the most fine-grained partition of $\mathcal{W}^N$, i.e. the set of all singleton

[14] The proof of Theorem 2.1 (ii) hinges on the method used in Pettigrew (2012).

propositions $\{w\}$, for $w \in \mathcal{W}^N$. But, by exploring equation (2.7), this result can hold for a finite algebra $\mathcal{A}$ generated by the cylinder sets A_t^a in $\mathcal{W}^N$. Therefore, the expert role demands ch to be a finitely additive probability function over $\mathcal{A}$, i.e. it ought to be such that: (i) for all $A_t \in \mathcal{A}$, $ch(A_t) \geq 0$, (ii) $ch(\varnothing) = 0$ and $ch\big(\mathcal{W}^N\big) = 1$, and (iii) for any disjoint $A_t, B_t \in \mathcal{A}$, $ch(A_t \cup B_t) = ch(A_t) + ch(B_t)$.

2.7 Motivating the Brier Score

One might claim that the result established in the last section hinges essentially on the application of the Brier score. After all, the proof of the theorem makes use of the convexity of the Brier score. If this is so, then it might be objected that the result is tainted with a disreputable arbitrariness, since no reason has been proposed for why we should prefer the Brier score over some other measure of the predictive accuracy of chance. I this section, I show how the use of the Brier score for measuring the predictive accuracy of chance can be justified.

Given the task of measuring the predictive accuracy of chance, one can isolate certain desirable properties of the measures of predictive accuracy, and try to show that the only measure that satisfies them is the Brier score. Of course, the question arises whether one can isolate such properties in the case of measuring the predictive accuracy of chance. In the remainder of this section, I will try to show how this might be possible.

This proposal can be formulated as follows. The Brier score satisfies two properties, namely *propriety* and *neutrality*, that seem adequate for the task of measuring the predictive accuracy of chances, and no other scoring function can satisfy them simultaneously. The heart of this proposal is Reinhard Selten's result (Selten 1998) showing that only the quadratic scoring function possesses the two properties just mentioned. A quadratic scoring function may be introduced as follows:

Quadratic Scoring Function: Suppose that $S : \mathcal{W}^N \times \mathcal{C} \rightarrow \mathbb{R}_0^+$ is a scoring function for each pair (w, ch). Then, we say that S is a quadratic scoring function if and only if

$$S(w, ch) = \sum_{j=1}^{2^N} \lambda_j \Big(v_w(\{w_j\}) - ch(\{w_j\}) \Big)^2,$$

where $\sum_{j=1}^{2^N} \lambda_j = 1$, and each $\lambda_j > 0$.

The quadratic scoring function allows us to weigh some propositions as more important than other. As it is easy to observe, the Brier score is an instance of the quadratic scoring function: it counts each proposition of the form $\{w_j\}$ equally. Specifically, Selten showed that while propriety is satisfied also by scoring functions other than the quadratic ones, e.g. by the spherical scoring function, only quadratic scoring functions satisfy both propriety and neutrality.

To introduce these two properties, let us define, for all $ch, ch' \in \mathcal{C}$, the *expected* predictive accuracy of the function ch' by the lights of ch and relative to a scoring function S as

$$\mathrm{Exp}_{ch,S}(ch') = \sum_{w \in \mathcal{W}^N} ch(\{w\}) S(w, ch'). \tag{2.17}$$

The expected predictive accuracy of ch' is the sum of scores for ch' at all worlds $w \in \mathcal{W}^N$, weighted by the values assigned by ch to each of these worlds, i.e. to each of the singleton propositions $\{w\}$, for $w \in \mathcal{W}^N$. One may think of $\mathrm{Exp}_{ch,S}(ch')$ as a measure of how good the function ch' is in matching every $w \in \mathcal{W}^N$, according to ch.

Given (2.17), we can characterize the first of the two properties of S, namely propriety:

Propriety. S is said to be *proper* if for all $ch, ch' \in \mathcal{C}$,

$$\mathrm{Exp}_{ch,S}(ch) \leq \mathrm{Exp}_{ch,S}(ch').$$

We say that S is *strictly proper* if $\mathrm{Exp}_{ch,S}(ch) < \mathrm{Exp}_{ch,S}(ch')$ for all $ch, ch' \in \mathcal{C}$ unless $ch = ch'$.

That is, a proper scoring function tells us that setting $ch = ch'$ minimizes the expected score. Thus, if S is proper, then ch does not expect that any other candidate chance function is more predictively accurate. If S is strictly proper, then the expected score is uniquely minimized by setting $ch = ch'$.

Neutrality may be introduced as follows:

Neutrality: S is said to be *neutral* if for all $ch, ch' \in \mathcal{C}$,

$$\mathrm{Exp}_{ch,S}(ch') = \mathrm{Exp}_{ch',S}(ch).$$

Neutrality may be explained in the following way: if one does not know which of the functions ch and ch' is "true" or "correct", then a scoring rule S should not favour any of them in advance.

There is a way to show that a scoring function S, which has both these properties, seems adequate for the task of measuring the predictive accuracy of chances. Here is why I think this is so. Suppose that ch is an expert function, i.e. it satisfies the condition of weak predictive accuracy. If S were not a proper scoring function, then setting $ch = ch'$ would not minimize the expected score. Thus, ch would not expect that no other ch' could be better than it in matching the possible worlds. But if so, ch hardly deserves the status of expert. Neutrality can be utilized in our context as follows. It may be interpreted as saying that if one does not know which of the functions ch and ch' is an expert, then a scoring function S should not favour any of them in advance. This is expressed by saying hypothetically that if ch were an expert and ch' were not, then ch' would be regarded as good at reducing the distance as ch would be if ch' were an expert instead of ch. Neutrality implies that if we are ignorant about which chance function is weakly predictively accurate, then there is no other factor that could affect our evaluation of candidate chance functions besides our ignorance about their predictive accuracy.

I do not deny that there might be other reasons for adopting the Brier score. For example, it might be prudent to defend this scoring function by appealing to its convexity. Let us formulate convexity more precisely as follows:

Convexity: S is said to be *convex* if for all $ch, ch' \in \mathcal{C}$, $w \in \mathcal{W}^N$, and all $\lambda \in [0, 1]$,

$$S(w, \lambda ch + (1 - \lambda)ch') \leq \lambda S(w, ch) + (1 - \lambda)S(w, ch').$$

That is, convexity means that the score of any function intermediate between ch and ch' is less or equal to a linear combination of the score of ch and the score of ch'. *Strict convexity* means that the above inequality is strict unless $ch = ch'$. There is a way to argue that any measure of chance's predictive accuracy should be convex. Suppose that ch and ch' are both expert functions, i.e. both satisfy the condition of weak predictive accuracy. In subsection 2.5.3, I have claimed that, in such a situation, these chance functions can still place constraints on the agent's credence function. The recipe I have alluded to is that the agent should set her credence function equal to a mixture of the expert functions ch and ch'. But why should we believe that this recipe is rational? Importantly, convexity gives an answer to this question. If S is convex, it would render any mixture of the form $\lambda ch + (1 - \lambda)ch'$ an expert function for the agent. This is so because any mixture of this form is at least as predictively accurate as ch and ch'. Hence, we have reason to follow that recipe.

In sum, I have shown that the use of the Brier score to measure the predictive accuracy of chance can be well justified, and thus is not a matter of mere *fiat*. I have isolated certain properties of the Brier score, and have argued that they make the Brier score well suited for the task of measuring the predictive accuracy of chance.

2.8 Measuring Predictive Accuracy: Bregman Divergence

This section shows how our main result established in section 2.6 can hold for a class of predictive accuracy measures that are instances of the so-called Bregman divergence. This in turn shows how this result can be generalized from the Brier score to the Bregman divergence.

For the purposes of this section, we use the following representation:

- Let $ch \in \mathcal{C}$ and $v_w \in \mathcal{V}$ be represented as vectors c and δ_w respectively. We have that $\mathcal{V} \subseteq \mathcal{C}$.

- Let the set $Conv(\mathcal{V})$ be represented by the set

$$\left\{ \sum_{w \in \mathcal{W}^N} \lambda_w \delta_w : 0 \leq \lambda_w \leq 1, \sum_{w \in \mathcal{W}^N} \lambda_w = 1 \right\},$$

  which is a convex hull of 2^N-dimensional vectors δ_w. Every vector in this set is a vector of finitely additive probability distribution over $\mathcal{W}^N$ denoted by $p = (p(\{w_1\}), ..., p(\{w_{2^N}\}))$, where $p(\{w_j\}) \geq 0$ and $\sum_{j=1}^{2^N} p(\{w_j\}) = 1$. We have that $\mathcal{V} \subseteq Conv(\mathcal{V})$.

Now, for our purposes, we can introduce the Bregman divergence between any c and c' in $\mathcal{C}$ as follows (a geometrical interpretation of it is given by Figure 2.2):

> **Bregman Divergence:** Suppose that $\mathcal{C}$ is a convex subset of $\mathbb{R}^N$. Let $\Phi : \mathcal{C} \to \mathbb{R}$ be a strictly convex function whose gradient $\nabla\Phi$ is defined in the interior of $\mathcal{C}$ and extends to a bounded, continuous function on $\mathcal{C}$. Then, for all $c, c' \in \mathcal{C}$, the Bregman divergence $D_\Phi : \mathcal{C} \times \mathcal{C} \to \mathbb{R}$ corresponding to Φ is given by
>
> $$D_\Phi(c', c) = \Phi(c') - \Phi(c) - \nabla\Phi(c) \cdot (c' - c),$$
>
> where $\nabla\Phi(c)$ is the gradient of Φ and $\cdot$ denotes the inner product of two vectors.

Since Φ is strictly convex, it follows that $D_\Phi(c', c) \geq 0$ with equality if and only if $c' = c$. The function D_Φ is the difference between the value of Φ at c and the first-order Taylor expansion of Φ around c evaluated at c'. Now, if the function Φ is defined as $\Phi(c) = \|c\|^2$, then $D_\Phi(c', c) = \|c' - c\|^2$. That is, the squared Euclidean distance is a Bregman divergence. Hence, the Brier score is a Bregman divergence. Other instances of the Bregman divergence include the Kullback-Leibler divergence, the Itakura-Sato distance, or the Hellinger distance (Banerjee et al. 2005).

How could our scoring function $S(w, ch)$ be represented as a Bregman divergence? Following the results due to Predd et al. (2009), we can introduce the scoring function $S(w, ch)$ as a Bregman divergence generated by a proper scoring rule s. Suppose that $s : \{0, 1\} \times [0, 1] \to \mathbb{R}$ is a continuous proper scoring rule.

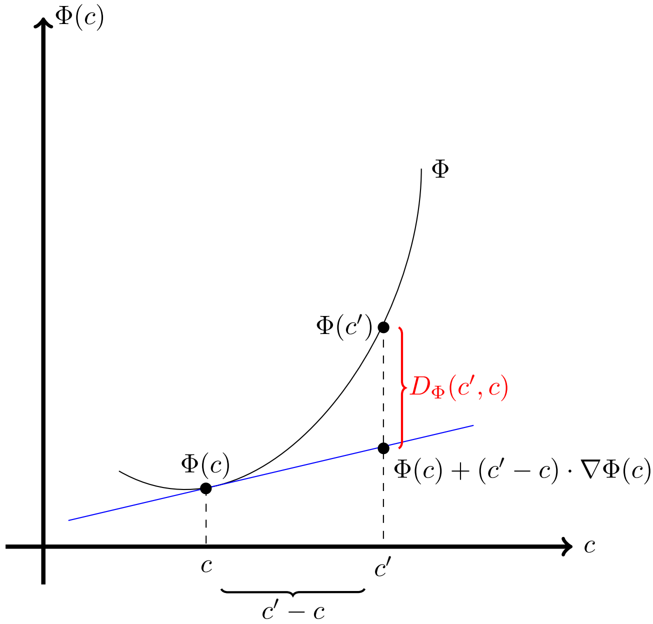

Figure 2.2: Visualizing the Bregman divergence generated by Φ. The Bregman divergence from c to c' is a difference between the value of two functions at c'. The first function is Φ and the second function is the tangent to Φ taken at c.

We say that:

- s is continuous if for all $i \in \{0, 1\}$ and any sequence $x_n \in [0, 1]$ converging to x, $\lim_{n\to\infty} s(i, x_n) = s(i, x)$.
- s is proper if for all $x \in [0, 1]$, $xs(1, y) + (1 - x)s(0, y)$ is uniquely minimized at $y = x$.

Then, the function $\varphi : [0, 1] \to \mathbb{R}$ defined as

$$\varphi(x) = -xs(1, x) - (1 - x)s(0, x) \tag{2.18}$$

is (i) continuous, bounded, strictly convex on $[0,1]$, (ii) continuously differentiable on $(0,1)$, and (iii) such that for all $x \in [0,1]$,

$$s(i,x) = -\varphi(x) - \varphi'(x)(i-x). \tag{2.19}$$

Now, define $\Phi : \mathcal{C} \to \mathbb{R}$ as

$$\Phi(c) = \sum_{j=1}^{2^N} \varphi\big(ch\left(\{w_j\}\right)\big). \tag{2.20}$$

Hence,

$$\Phi(\delta_w) = \sum_{j=1}^{2^N} \varphi\big(v_w(\{w_j\})\big). \tag{2.21}$$

Then,

$$\begin{aligned}
S(w,ch) &= \sum_{j=1}^{2^N} s\Big(v_w(\{w_j\}), ch(\{w_j\})\Big) \\
&= \sum_{j=1}^{2^N} -\varphi\big(ch(\{w_j\})\big) - \varphi'\big(ch(\{w_1\})\big)(v_w(\{w_j\}) - ch(\{w_j\})) \text{ (by (2.18))} \\
&= D_\Phi(\delta_w, c) - \sum_{j=1}^{2^N} \varphi\big(v_w(\{w_j\})\big) \\
&\quad \text{(by the definition of Bregman divergence)} \\
&= D_\Phi(\delta_w, c) + \sum_{j=1}^{2^N} s\Big(v_w(\{w_j\}), v_w(\{w_j\})\Big) \\
&\quad \text{(by the fact that } s\big(v_w(\{w_j\}), v_w(\{w_j\})\big) = -\varphi\big(v_w(\{w_j\})\big)\text{)},
\end{aligned}$$

as required.

The following proposition (for a proof see Predd et al. 2009, p. 4789) captures a feature of any Bregman divergence, feature that will play a crucial role in showing that our main result holds for any Bregman divergence:

Proposition 2.1. *Let $Conv(\mathcal{V}) \subseteq [0,1]^N$ be a closed convex subset of $\mathbb{R}^N$. Then, if $c \in \mathcal{C} - Conv(\mathcal{V})$, there is $c' \in Conv(\mathcal{V})$ such that $D_\Phi(c', c) \leq D_\Phi(c'', c)$ for all $c'' \in Conv(\mathcal{V})$. Moreover, $D_\Phi(c'', c') \leq D_\Phi(c'', c) - D_\Phi(c', c)$ for all $c'' \in Conv(\mathcal{V})$ and all $c \in \mathcal{C} - Conv(\mathcal{V})$.*

The key idea behind this proposition is this. Given a chance function not in $Conv(\mathcal{V})$, there is a chance function in $Conv(\mathcal{V})$ that is at least as close to the chance function outside $Conv(\mathcal{V})$ as any other chance function in $Conv(\mathcal{V})$. The vector c' is called the projection of c onto $Conv(\mathcal{V})$. The second sentence of this proposition expresses the fact that any Bregman divergence satisfies the generalized Pythagorean theorem, i.e. $D_\Phi(c'', c) \geq D_\Phi(c'', c') + D_\Phi(c', c)$.

With these notions and results in hand, we can state and prove the following theorem:

Theorem 2.2. *Let $S(w, ch) = D_\Phi(\delta_w, c) + \sum_{j=1}^{2^N} s\Big(v_w(\{w_j\}), v_w(\{w_j\})\Big)$. Then:*

(i) If $c \notin Conv(\mathcal{V})$, then there is $c' \in Conv(\mathcal{V})$ such that $S(w, ch') \leq S(w, ch)$ for all $w \in \mathcal{W}^N$.

(ii) If $c' \in Conv(\mathcal{V})$, then there is no $c^* \in \mathcal{C}$ such that $c^* \neq c'$ and $S(w, ch^*) \leq S(w, ch')$ for all $w \in \mathcal{W}^N$.

Proof. For the proof of Theorem 2.2 (i), suppose that $c \in \mathcal{C} - Conv(\mathcal{V})$. Then, by Proposition 2.1, there is $c' \in Conv(\mathcal{V})$ such that for all $c'' \in Conv(\mathcal{V})$,

$$D_\Phi(c'', c') \leq D_\Phi(c'', c) - D_\Phi(c', c).$$

And since $\mathcal{V} \subseteq \mathcal{C}$, we have that for all $\delta_w \in \mathcal{V}$,

$$D_\Phi(\delta_w, c') \leq D_\Phi(\delta_w, c) - D_\Phi(c', c).$$

Since $c \notin Conv(\mathcal{V})$ and $c' \in Conv(\mathcal{V})$, we have that $c \neq c'$, and so $D_\Phi(c', c) > 0$. Hence,

$$D_\Phi(\delta_w, c') \leq D_\Phi(\delta_w, c),$$

and by the definition of $S(w, ch)$,

$$S(w, ch') \leq S(w, ch),$$

as required.

For the proof of Theorem 2.2 (ii), suppose that $c' \in Conv(\mathcal{V})$. Now, suppose that $D_\Phi(c'', c^*) \leq D_\Phi(c'', c')$ for all $c'' \in Conv(\mathcal{V})$. Since $c' \in Conv(\mathcal{V})$, it follows that $D_\Phi(c'', c^*) \leq D_\Phi(c', c') = 0$. This implies that $D_\Phi(c'', c^*) = 0$ and $c^* = c'$. But this contradicts the assumption that $c^* \neq c'$. □

I have just shown that relative to the scoring function represented as Bregman divergence, the condition of weak predictive accuracy, to wit, a necessary condition on chance to play the expert role, demands chance to be a finitely additive probability function. Thus, we have shown that the expert role demands chance to be formally adequate, even if the assumption about the Brier score as a measure of predictive accuracy is relaxed.

2.9 Expert Role, Predictive Accuracy, and Calibration

One of the building blocks of our argument is the assumption that the predictive accuracy of chance is the extent to which a chance distribution diverges from a world which is fully characterized by a truth-value distribution. Thus, the predictive accuracy of chance is concerned with "nearness" of the predicted chance distribution to a possible outcome at a particular time represented by a particular truth-value distribution.

But it seems that the predictive accuracy of a probabilistic theory does not only pertain to particular outcomes. Also, we often evaluate probabilistic theories with respect to how well they predict the relative frequency with which a given outcome occurs. In other words, the predictive accuracy of chance is not limited to a single case, but also concerns its predictive success in the long run, i.e. in a long series of trials. In this section, I suggest a way to incorporate the thought about long run predictive accuracy into our basic framework. Specifically, I show that the predictive accuracy measured by the Brier score is able to reflect how well chances are *calibrated* with relative frequencies at a given world.

Informally, the idea of calibration is intuitively simple. Consider a probabilistic theory that repeatedly makes probabilistic predictions. Now, if this theory assigns a 0.7 chance to some coin landing heads over a series of 100 tosses, and it happens that the coin lands heads 70 times over this long run, then there is something good about this probabilistic theory. We say that this theory is *well calibrated*. More generally, a probabilistic theory is well calibrated if, over a long run of trials, for each chance x that it assigns, the proportion of true propositions amongst all propositions to which it assigns x equals x. Let us make this idea more precise. Assume that a probabilistic theory is fully characterized by a chance function $ch : \mathcal{W}^N \to [0,1]$. Then:

> **Calibration:** Let us divide $\mathcal{W}^N$ into mutually exclusive reference classes $W_x^N = \big\{\{w\} \in \mathcal{W}^N : ch(\{w\}) = x\big\}$. That is, two propositions $\{w_1\}, \{w_2\} \in \mathcal{W}^N$ belong to the same reference class if and only if $ch(\{w_1\}) = ch(\{w_2\})$. Define the frequency of true propositions at a world w amongst all propositions to which ch assigns chance x as:
>
> $$freq_w\big(W_x^N\big) = \frac{\big|\big\{\{w\} \in \mathcal{W}^N : ch(\{w\}) = x \text{ and } v_w(\{w\}) = 1\big\}\big|}{|\{\{w\} \in \mathcal{W}^N : ch(\{w\}) = x\}|}$$
>
> Then, ch is *well calibrated* at w if for each value $x \in \mathrm{ran}(ch)$,
>
> $$x = freq_w\big(W_x^N\big).$$

Let me stress some basic features of calibration so introduced. First, calibration has to do with how well a chance distribution matches the frequency of true propositions in a reference class consisting of propositions to which it assigns the same value x. Thus, calibration depends on how the space is carved into reference classes. The above-mentioned way of dividing $\mathcal{W}^N$ into reference classes is due to Bas van Fraassen (1983) and Abner Shimony (1988). The key idea is that the reference class is fixed by the chance function. Second, calibration is a measurable property: we can measure how well calibrated a chance function is, even if it falls short of perfect calibration. But how can calibration be measured?

Let us focus on the calibration score (also called the index of calibration) given in Murphy (1973):

> **Calibration Score:** Let us divide $\mathcal{W}^N$ into mutually exclusive reference classes $W_x^N = \big\{\{w\} \in \mathcal{W}^N : ch(\{w\}) = x\big\}$ for $x \in \mathrm{ran}(ch)$. If n_x is the number of elements in W_x^N, we have that $\sum_{x\in\mathrm{ran}(ch)} n_x = \big|\mathcal{W}^N\big|$. For each $x \in \mathrm{ran}(ch)$, let $freq_w\big(W_x^N\big)$ be the frequency of true propositions in the set labelled by x. Then, the calibration score for ch at world w is given by
>
> $$C(w, ch) = \frac{1}{|\mathcal{W}^N|} \sum_{x\in\mathrm{ran}(ch)} n_x \Big(freq_w\big(W_x^N\big) - x\Big)^2,$$
>
> where $\big|\mathcal{W}^N\big|$ stands for the number of elements in $\mathcal{W}^N$ and n_x is the number of elements in W_x^N.

That is, the calibration score is the sum of weighted squared differences between the frequency of true propositions in W_x^N and the value of chance assigned to all propositions in this set, where the weight is the number of elements in W_x^N. The sum itself is weighted by an equal weight $\frac{1}{|\mathcal{W}^N|}$ assigned to every world $w \in \mathcal{W}^N$.

The idea behind the calibration score is as follows. The chance function ch is well calibrated at w if $C(w, ch) = 0$. This means that half the propositions assigned chance $\frac{1}{2}$ are true, two-thirds of those assigned chance $\frac{2}{3}$ are true, three-fourths of those assigned chance $\frac{3}{4}$ are true, and so forth. It is easy to observe that, for each world w, the truth-value distribution at w is always well calibrated. For v_w assigns 1 to a proposition just in case that proposition is true at w, and so 100% of those propositions to which v_w assigned value 1 are true. And v_w assigns 0 to a proposition just in case that proposition is false at w, and so 0% of those propositions to which v_w assigned value 0 are true.

At first sight, the idea of calibration and predictive accuracy measured by the Brier score are significantly different. Both these measurable quantities capture important aspects concerning the assessment of probabilistic predictions given by chance functions. But whereas the Brier score-based predictive accuracy measures the reliability of a chance function by assessing how informative it is about possible truth-value distributions, the calibration score measures how reliable it is as an indicator of the truth-frequency of propositions. Could these two quantities be reconciled?

One attempted reconciliation, due originally to Murphy (1973) and developed in DeGroot and Fienberg (1983) and in Blattenberger and Lad (1985), says that the Brier score is an aggregate of various measures, one of which is the calibration score. In particular, the Brier score can be separated into the calibration score and the refinement score.[15] Informally, the refinement score measures the variance in truth-value across all the reference classes W_x^N fixed by ch or, equivalently, the extent to which each reference class of propositions assigned by the same chance x is uniform in the truth-value distribution over its members. More precisely, the refinement score can by introduced as follows:

> **Refinement Score:** Let us divide $\mathcal{W}^N$ into mutually exclusive reference classes $W_x^N = \big\{\{w\} \in \mathcal{W}^N : ch(\{w\}) = x\big\}$, for $x \in \mathrm{ran}(ch)$. If n_x is the number of elements in W_x^N, we have that $\sum_{x\in\mathrm{ran}(ch)} n_x = \big|\mathcal{W}^N\big|$. For each $x \in \mathrm{ran}(ch)$, let $freq_w\big(W_x^N\big)$ be the frequency of true propositions in the set labelled by x. Then, the refinement score for ch at world w is given by
>
> $$REF(w, ch) = \frac{1}{|\mathcal{W}^N|} \sum_{x\in\mathrm{ran}(ch)} n_x freq_w\big(W_x^N\big)\Big(1 - freq_w\big(W_x^N\big)\Big),$$
>
> where $\big|\mathcal{W}^N\big|$ stands for the number of elements in $\mathcal{W}^N$ and n_x is the number of elements in W_x^N.

That is, the refinement score is the sum of weighted uniformities of truth values in W_x^N, where the weight is the number of elements in W_x^N. The sum itself is weighted by an equal weight $\frac{1}{|\mathcal{W}^N|}$ assigned to every world $w \in \mathcal{W}^N$.

The refinement score is minimal ($REF(w, ch) = 0$) if ch fixes the reference classes in such a way that the frequency of true propositions amongst the propositions assigned by the same chance x is either 0 or 1. This is a situation in which ch perfectly discriminates truths from falsehoods. It is maximal $\big(REF(w, ch) = \frac{1}{4}\big)$ if the frequency of true propositions in the reference classes is $\frac{1}{2}$.

[15] In fact, Murphy (1973) showed that the refinement score can be separated into two other components: uncertainty and resolution.

The Brier score thus can be separated as follows:

Brier Score Separation: Suppose that $S(w, ch) = \frac{1}{|\mathcal{W}^N|} \sum_{j=1}^{2^N} \Big(v_w(\{w_j\}) - ch(\{w_j\})\Big)^2$. Then,

$$S(w, ch) = C(w, ch) + REF(w, ch).$$

The Brier score so separated might be understood as balancing the two components off and giving an overall measure of predictive accuracy. But why do we need to balance off calibration and refinement? The answer is that it appears that refinement can be improved at a cost of calibration, and vice versa. For example, perfect refinement can be achieved by assigning chance 1 to every false proposition and chance 0 to every true proposition, but in terms of calibration such assignment is maximally miscalibrated.

Given that the Brier score may be understood as a measure of the balance between calibration and refinement, I suggest that the result from section 2.6 can also be read as follows:

- if ch is not a finitely additive probability function, then there is another ch' that is a finitely additive probability function and that could strike a better balance of calibration and refinement in every $w \in \mathcal{W}^N$.
- if ch is a finitely additive probability function, then there is no other ch' such that it could strike a better balance of calibration and refinement in every $w \in \mathcal{W}^N$.

2.10 Concluding Remarks and Further Research

This chapter has shown how a particular use of the Principal Principle can lead to the conclusion that chance is a finitely additive probability function. I have arrived at this conclusion, first, by introducing and defending chance functionalism: the view that chance is whatever plays certain functional roles. Second, by exploring a particular role—the expert role codified in the Principal Principle—I have argued that chance plays the expert role if it is weakly predictively accurate. This condition was taken to be fairly minimal: a chance function is weakly predictively accurate if and only if there is no other chance function, known to the epistemic agent, that could fit better the possible truth-value distributions over some family of propositions. Third, I have proved that chance is weakly predictively accurate if and only if it is a finitely additive probability function. From this, I have concluded that chance plays the expert role if it is a finitely additive probability function. Thus, I have shown that the expert role formulated in the Principal Principle demands chance to be formally adequate.

Obviously, the approach to formal adequacy presented in this chapter leaves some questions untouched. It is worth mentioning that this approach has extensions that point towards future research. Let me briefly indicate some of them. First, by letting $\mathcal{A}$ be a countably infinite σ-algebra, one may try to extend this approach to show that the expert role demands chance to be *countably additive.* Second, one may wish to extend the framework in order to prove that chance satisfies the laws of *quantum probability.* This could be done by replacing $\mathcal{W}^N$ with a Hilbert space $\mathcal{H}_N$, which is a N-dimensional vector space, by generating an algebra of propositions (subspaces in $\mathcal{H}_N$)—an orthomodular lattice $\mathcal{L}$, and finally by showing that ch must be a σ-additive quantum probability (quantum state) $\phi : \mathcal{L} \to [0, 1]$ in order to play the expert role. A third extension worth mentioning concerns an axiomatic treatment of probability in which *conditional probability* is primitive. For example, one may define chance as a mapping $ch : \mathcal{A} \times \mathcal{A} \to [0, 1]$, and show that it satisfies the axioms for Popper functions. These tasks also suggest avenues for further elaboration and clarification of the functionalist approach to formal adequacy.

Chapter 3

Chance and Resiliency

In this chapter, I show how a particular resiliency-centered approach to chance lends support for two conditions that are claimed in the literature to be constitutive of chance. The first condition says that the present chance of some proposition A conditional on the proposition about some later chance of A should be set equal to that later chance of A. The second condition requires the present chance of some proposition A to be equal to the weighted average of possible later chances of A, where the weights are chances assigned by the present chance function to propositions about A's possible later chances. I first introduce, motivate, and make precise a resiliency-centered approach to chance whose basic idea is that any chance distribution should be maximally invariant under variation of experimental factors. Second, I provide resiliency-based arguments for the two conditions: I show that any present chance distribution that violates the two conditions can be replaced by another present chance distribution that satisfies them and is more resilient under variation of experimental factors. Finally, I show that the conditions in the case of Humean accounts of chance lead to hardly acceptable consequences. It is then argued that considerations of the resiliency of chance have substantial repercussions on the question of whether these conditions should be retained in that case.

3.1 Introduction

Several philosophers have claimed that any notion of chance (physical probability) should satisfy certain conditions that appear to be constitutive of chance or might be regarded as platitudes we apparently have about chance (e.g. Loewer 2001; Schaffer 2003, 2007). According to this line of thought, a notion of chance that violates these conditions either refers to something which only approximates genuine chance or does not refer to chance at all. Despite considerable discussion, there is no consensus among philosophers as to how many of such conditions a notion of chance should satisfy. Famously, David Lewis (1986) claimed that his Principal Principle is the sole condition on chance: it captures all we know about chance. But, as shown by Frank Arntzenius and Ned Hall (2003), Lewis's claim cannot be rationally sustained. Similarly, Jonathan Schaffer (2007) has argued that besides the Principal Principle, there is a number of equally plausible conditions that inform our understanding of chance.

Two interesting conditions on chance have been proposed in Bigelow et al. (1993). In addition, they have claimed that the two conditions are constitutive of chance. To introduce these conditions, I assume that Ch denotes a chance function over a finite set of propositions $\mathcal{A}$ generated by a set of possible worlds $\mathcal{W}$.[1] Suppose further that $\mathcal{C}_{\mathcal{F}} = \{ch_w : w \in \mathcal{W}\}$ is a finite set of chance functions over $\mathcal{A}$ indexed by the worlds in $\mathcal{W}$. Each ch_w stands for a chance function in some possible world w. Following Bigelow et al. (1993, p. 458), let Ch be a *present* or *prior* chance function, and let the ch_w's be the possible *later* or *posterior* chance functions. As will be apparent, the former can be understood as an *unrefined* chance function and the latter as the possible *refinements* of Ch. Further, let C_{ch_w} be the proposition that the chance distribution over $\mathcal{A}$ is given by ch_w. Assume that Ch and the ch_w's are probability functions over $\mathcal{A}$. Then, the two conditions on chance may be presented more generally as follows:

> **C1** (*Chance conditional on chance formulation*)**:** For all $A \in \mathcal{A}$ and all possible later chance functions ch_w,
>
> $$Ch(A|C_{ch_w}) = ch_w(A).$$

[1] Here the original formulation of these conditions is slightly rephrased to fit in with the framework presented in this chapter.

C2 (*Chances are equal to the expected values of chances*)**:** For all $A \in \mathcal{A}$,

$$Ch(A) = \sum_{w \in \mathcal{W}} Ch(C_{ch_w}) ch_w(A).$$

In words, C1 tells us that the present chance of some proposition A conditional on the proposition about some later chance of A should be set equal to that later chance of A. C2 requires the present chance of some proposition A to be equal to the weighted average of possible later chances of A, where the weights are chances assigned by the present chance function to propositions about A's possible later chances. In other words, the present chance of some proposition A is the expectation of the possible posterior chances of A. Thus, both C1 and C2 relate any present chance distribution to the possible later chance distributions in a certain way.

To get a better grasp of these conditions, one might frame the present chance function Ch, the possible later chance functions, the ch_w's, and the relations between them covered by C1 and C2 within the theory of expert functions developed by Haim Gaifman (1988) and Bas van Fraassen (1989, chapter 8). With this theory in mind, each possible later chance function might be interpreted as a first-order chance function over $\mathcal{A}$, and assuming that every proposition of the form C_{ch_w} belongs to $\mathcal{A}$, the present chance function Ch might be regarded as a second-order chance function over $\mathcal{A}$ so enriched. Following Gaifman's terminology, we may interpret the possible first-order chance functions that figure in conditions C1 and C2 as expert functions for any present chance function.

The question arises: why should we believe that the two conditions are constitutive of chance? It seems that these conditions are neither trivially true, nor self-evident, nor do they follow from platitudes we apparently have about chance. This question appears to be even more interesting once we acknowledge that in the case of Humean accounts of chance these conditions lead to hardly acceptable consequences. This observation in turn prompts the following problem: should we reject Humean accounts of chance or the two conditions? Given that Humean accounts of chance have been invoked successfully in explanations of certain physical phenomena as well as in explanations of some philosophical puzzles concerning chances, their rejection seems to be too hasty.[2] After all, in such cases, one might start to question the very constitutive role of the two conditions, rather than Humean accounts of chance themselves. From this point of view, then, bringing up a rationale for these conditions seems to be important. Also, the task of providing such rationale appears to be challenging. For on the face of things, there seems to be no straightforward answer to the question of what considerations concerning chances could lend credence to the two conditions. We might require that whatever those considerations might be, they should point towards some fundamental feature of chance, feature that itself does not require any deeper justification. But alternatively, we might tackle this question in a more modest way by showing that some well-motivated considerations concerning chance provide support for the two conditions. As will be apparent, this chapter employs the latter strategy.

My primary aim in this chapter is to provide support for conditions C1 and C2 by appealing to the idea of chance's resiliency. I show that chances that violate these two conditions do not maximize resiliency suitably understood. Resiliency is taken to be a kind of stability property of chance: it reflects the approximate invariance of a chance distribution under variation of experimental factors that bear on a given chance set-up. I introduce and motivate a norm, according to which chances that figure in statistical laws should maximize resiliency over a given set of experimental factors. This idea draws on Brian Skyrms's (1977; 1978; 1980a) resiliency-based account of chance. Skyrms (1980a) has used the notion of resiliency to explain the status of chances or physical probabilities posited in statistical laws. He has argued that such chances should be highly resilient probabilities. In this chapter, this idea is employed in the following way: (i) the experimental factors over which the resiliency of chance is evaluated form a partition, (ii) each cell of the partition singles out a possible world at which the cell holds true, (iii) we associate with every such world a possible later chance function which is the present chance function *refined* by accommodating

[2] Let me mention two such attempts. Loewer (2001) uses Lewis's conception of Humean chance to resolve what he calls the paradox of deterministic probabilities, to wit, the problem of reconciling the fact that some theories posit non-trivial probabilities for events not to occur with the fact that those events are determined to occur. In a similar vein, though more specifically, Frigg and Hoefer (2015) use their Humean account of chance to explain the nature of probabilities posited by classical statistical mechanics in deterministic settings.

information about an experimental factor (a given cell of the partition) that obtains at that world, (iv) in each possible world, we measure the resiliency of the present chance function by showing how "close" that chance function is to the possible later chance function at that world. Given these assumptions, it is then shown, for each of the conditions C1 and C2 separately, that (i) if the present chance function does not obey that condition, then there is another present chance function that obeys it and is "closer" to every possible later chance function, and (ii) if a present chance function obeys that condition, then there is no other present chance function that is closer to every possible later chance function. This result then shows that any present chance function which violates that condition is inadmissible with respect to the resiliency measure. That is, it is ruled out as inadmissible by a seemingly plausible standard of resiliency.

Secondarily, I show that considerations concerning the resiliency of chance have substantial repercussions on whether the two conditions should be retained in the case of so-called Humean accounts of chance. Since these conditions lead to hardly acceptable consequences in that case, one might question their plausibility, let alone their constitutive character. I suggest, however, that our resiliency-centered arguments for these conditions provide reason for treating Humean accounts of chance with caution. For if chances are not constrained by these conditions, they do not necessarily maximize resiliency which appears to be one of the fundamental virtues of a good probabilistic theory.

Before I continue some remarks are in order. First, as it stands in Bigelow et al. (1993, p. 458), the words "present" and "later" that characterize the chance functions in C1 and C2 carry no specific meaning. This in turn leaves much room for interpretation. In section 3.2, I will introduce some of their possible understandings, and I will propose to understand the present chance function as an unrefined chance function and the possible later ones as its possible refinements. I will argue that this understanding is sufficiently broad to make sense of various cases involving the relation between the chance functions that figure in the two conditions. Second, though the chapter employs Skyrms's idea of resiliency, it does not adopt his theory of chance built on this concept. Skyrms (1980a; 1984) developed a pragmatic theory of chance, according to which chance is a resilient subjective or personal probability. He thus thought of resiliency as a feature of one's degrees of belief or credences over some family of propositions. In this chapter, resiliency is taken to be a measurable property of chance distributions, without claiming that these chance distributions can be analysed in terms of one's resilient degrees of belief.

The chapter is structured as follows. Section 3.2 expounds the two conditions. Section 3.3 introduces, motivates, and makes precise the idea of chance's resiliency. Section 3.4 provides two resiliency-based arguments for the two conditions. Section 3.5 discusses some philosophical consequences of the resiliency-based approach to chance for the debate concerning the plausibility of Humean accounts of chance. Section 3.6 concludes.

3.2 A Closer Look at the Conditions

In this section, I explain the two conditions more precisely. First, I give some examples that (i) show how these conditions could work and (ii) provide some insight into how the present and the possible later chance functions could be understood. Second, in order to give a more general account of the present and the possible later chance functions, I introduce a rule called a *refinement rule* for chances. Finally, I explore some relations that hold between these conditions and state some of their consequences by combining them with some additional, fairly intuitive assumptions.

Let me start with some examples illustrating how conditions C1 and C2 could work. Also, the purpose of these examples is to highlight some possible understandings of the words "present" and "later" that characterize the chance functions that figure in C1 and C2, and to motivate a fairly general way of thinking about them.

> **Example 1:** Suppose that the coin is loaded with iron and that its bias depends on whether the electromagnet is *off* or *on*. Assume that if the electromagnet is on, the coin is biased towards heads, and if it is off, the coin is biased towards tails. We have two possible worlds, one in which the electromagnet is on, the other in which it is off. Also, we have two corresponding chance functions: the one assigning greater chance to heads at world "on", and the other assigning greater

chance to tails at world "off". In this scenario, C1 tells us that the present chance of heads, given that the chance of heads is determined by the chance function at world "on", is equal to the chance of heads at world "on". C2 states that the chance of heads is a mixture of the chance of heads at world "on" and the chance of heads at world "off".

Example 2:[3] Suppose that you are in Atlantic city playing a "backward bandit" game. It is like playing a one-armed bandit game except the following: while in the ordinary one-armed bandit game the chance of winning the jackpot is determined (of course, not the outcome itself) when you pull the handle, in the "backward bandit" game the chance of winning the jackpot is not. In fact, peculiarly, your present chance of winning the jackpot is *counterfactually dependent* on whether you actually win! We have two peculiar counterfactuals: (i) if world "win" were actual, the chance of winning the jackpot would be low, and (ii) if world "not win" were actual, the chance of winning the jackpot would be high. In this case, C1 tells us that the present chance of winning the jackpot, given that the chance of winning the jackpot is determined by the chance function at world "win", is equal to the chance of winning the jackpot at world "win". And C2 requires the chance of winning the jackpot to be a mixture of the chance of winning the jackpot at world "win" and the chance of winning the jackpot at world "not win".

Example 3:[4] Merlin is about to cast a risky spell on a bottle at time t. There is a 50% chance that his spell will turn out to be a dark one, and there is a 50% chance that it will turn out to be a light one. We consider two possible worlds: the one at which Merlin's spell is light, and the other with a dark spell. If Merlin's spell turns out to be a light one, then by time $t+1$ the bottle turns a milky colour and at time $t+1$ there is a 60% chance that a genie will emerge from the bottle. If Merlin's spell turns out to be a dark one, then by time

[3] This is a slight reformulation of the example given by Bigelow et al. (1993).
[4] This example is based on a case discussed by Fisher (2006).

$t + 1$ the bottle turns a murky colour and at time $t + 1$ there is a 40% chance that a genie will emerge from the bottle. By C1, the present (at t) chance of the genie emerging from the bottle, given the chance at $t + 1$ of this event at world with a light spell, is 60%. By C2, the present (at t) chance of the genie emerging from the bottle is a mixture of the chance at $t + 1$ of this event at world with a light spell and the chance at $t + 1$ of this event at world with a dark spell.

Importantly, the examples given above show that the words "present" and "later" have three different meanings. In cases like example 1, the possible later chance functions are two possible "fine-grained" chance functions. Here no temporal shift in chance distributions is necessarily involved: we consider two ways in which a "coarse-grained" chance function could be made finer.[5] Example 2 points towards subjunctive supposing or hypothesizing. We imagine what the present chance function would be if some possible world turned out to be actual. Here the possible later chance functions are counterfactual chance functions. In example 3, we consider two possible shifts of the present chance distribution across time. The possible later chance functions result from two ways by which a present chance function might develop at some later time. Thus, here we relate the current and the possible future chance functions.

A question that naturally arises is: could we provide a more general account of the present and the possible later chance functions that figure in C1 and C2 to capture the different meanings mentioned above? To answer this question, I introduce, in an abstract way, a rule that relates a chance function Ch, a certain partition of propositions, and the chance functions ch_w. Consider a chance function Ch and a rule RE which might be called a refinement rule for chances. We assume that $\mathcal{F} = \{F_w : w \in \mathcal{W}\}$ is a finite partition of $\mathcal{W}$ into propositions describing experimental factors that obtain at some possible world w, and that could affect a given chance set-up, e.g. a coin tossing, a slot machine like the aforementioned "backward bandit", or even a spell casting on a bottle. For this idea to work, we might think of an element in $\mathcal{W}$ as specifying not only the outcome of a chancy experiment, but also the experimental arrangement. More

[5] More precisely, what is made finer is an experimental set-up to which a given chance function is ascribed.

generally, I take any F_w in $\mathcal{F}$ to stand for a supposition upon which the chance function Ch can be refined. We can think of it as a possible world at which a given experimental factor obtains. Furthermore, such a supposition might be a matter-of-fact supposition as well as a subjunctive supposition. In particular, F_w may be a more fine-grained description of a given chance set-up as in example 1, or it may describe a subjunctive supposition on which the chance set-up is counterfactually dependent as in example 2, or it may characterize some history of chancy outcomes in the interval from time t to $t+1$, as in example 3.

Now, we can think of RE as a function that takes element F_w of the partition $\mathcal{F}$ and returns a chance function $RE(Ch, \mathcal{F}, F_w)$. That is, the refinement rule is a way to transform Ch in order to accommodate some supposition. Given the refinement rule for chances so understood, I propose to think of a present chance function and the possible later chance functions as a chance function before refinement (an unrefined chance function) and the possible refinements respectively. The latter are the chance functions obtained from the present one by accommodating the suppositions from partition $\mathcal{F}$. As shown by the three examples given above, there are various ways by which Ch might be refined. That is, Ch may be refined by considering a partition $\mathcal{F}$ containing propositions about possible ways in which a chance set-up could be made finer, or a partition $\mathcal{F}$ containing propositions about possible counterfactual situations describing factors that could affect the set-up, or a partition $\mathcal{F}$ of propositions about some possible intervening histories between t and $t+1$.

Notice that if we take Ch to stand for a chance function at t and the ch_ws as the possible chance functions at $t+1$ (characterized as the possible chance functions obtained from Ch by accommodating information about intervening histories between t and $t+1$), then conditions C1 and C2 parallel the well-known *Reflection principles* (van Fraassen 1984, van Fraassen 1995) that govern subjective probabilities. Under this interpretation, C1 says that the current chance of A conditional on the proposition about some future chance of A should be equal to that future chance of A. And C2 says that the current chance of A should be the expectation of the possible future chances of A. But given that the refinement rule for chances is broadly understood, other interpretations of these conditions are possible.

Since the refinement rule for chances is an abstract rule, it is tempting to ask: how could the later chance function $RE(Ch, \mathcal{F}, F_w)$ be defined? One candidate is the rule of conditionalization. More precisely:

Refinement by Conditionalization (REC): For all $A \in \mathcal{A}$,

$$RE(Ch, \mathcal{F}, F_w)(A) = Ch(A|F_w),$$

provided that $Ch(F_w) > 0$.

That is, the later chance function is equal to the present chance function conditional on some experimental factor F_w in $\mathcal{F}$ that obtains at w.

Importantly, Lewis (1986) famously claimed that REC captures fully the kinematics of chance, that is, the way in which chances develop through time. He considered a specific partition of possible intervening histories of chancy outcomes in the interval between t and $t+1$, with each intervening history obtaining at some possible world. To illustrate this point consider again example 3. There are two possible histories between t and $t+1$. If the world will develop so that the bottle has a milky colour, then the chance of a genie emerging from it would change from the chance at t to the chance at $t+1$ by conditionalizing on the intervening history between t and $t+1$. Similarly, the chance of a genie emerging from the bottle changes upon the fact that the bottle has a murky colour. Although Lewis's kinematics of chance is seemingly attractive, it captures only one kind of chance refinement: the one alluded to in example 3. But our goal is to deal also with other cases of chance refinement. As it is easy to observe, examples 1 and 2 cannot be adequately governed by Lewis's kinematics of chance, for they do not describe the evolution of chances through time. For this reason, I reject the possibility of defining $RE(Ch, \mathcal{F}, F_w)$ as the present chance function conditional on intervening histories. Even though, we can still consider REC as a viable option, for it is not limited to the intervening histories. Should we assume that any kind of chance refinement on a partition proceeds by the rule of conditionalization? The following considerations show that we should not.

First, it seems that REC requires justification. Similarly, Bayesians do not take for granted the parallel claim that rational credence is updated on evidence by the rule of conditionalization. Rather, they argue for this claim by appealing to various arguments: Dutch book arguments or arguments from minimizing expected inaccuracy. By analogy, we should ask what considerations concerning chances justify the claim that REC is a correct refinement rule for chances. Lacking an answer to this question, it seems that REC is only a conjecture. Moreover, as recently argued in Lange (2006), REC faces serious problems in connection with higher-order chances and so is not as evident as we might think.

Second, and more importantly for the purpose of this chapter, if we adopt REC as a refinement rule for chances, we thereby assume that any present chance function satisfies C2. This in turn undermines the whole project of providing an independent support for this condition. To see this, observe that if Ch is refined by conditionalizing on supposition F_w, then

$$Ch(A) = \sum_{w \in \mathcal{W}} Ch(A|F_w)Ch(F_w), \tag{3.1}$$

where the sum extends over all w such that $Ch(F_w) > 0$ and $\sum_{w \in \mathcal{W}} Ch(F_w) = 1$. That is, the present chance of A is the expectation of the possible conditional present chances of A. Given that $ch_w = Ch(A|F_w)$, we have that the present chance of A is a convex combination of possible later chances of A. And this is exactly what C2 requires.

For these reasons, I proceed in the next sections without a definition of $RE(Ch, \mathcal{F}, F_w)$. I take it that Ch is an unrefined chance function and the ch_w's are its possible refinements obtained from Ch by accommodating in a certain way a partition of suppositions describing experimental factors. Thus, I treat the possible later chance functions as fixed by *some* refinement rule. Also, as will be shown, this account of Ch and the ch_w's underpins a resiliency measure which is one of the building blocks of our resiliency-based arguments for conditions C1 and C2.

Despite this abstract characterization of the refinement rule, we can require it to satisfy the following conditions:

(R1) For all $F_w \in \mathcal{F}$ and any refinement $RE(Ch, \mathcal{F}, F_w)$, $RE(Ch, \mathcal{F}, F_w)(F_w) = 1$.

(R2) Suppose that $\mathcal{RE}$ is a set of possible refinements of some Ch. Then, if

$Ch \in \mathcal{RE}$, $Ch(\cdot) = RE(Ch, \mathcal{F}, F_w)(\cdot)$ for some $F_w \in \mathcal{F}$.

(R3) For any $A \in \mathcal{A}$ that implies F_w, $RE(Ch, \mathcal{F}, F_w)(A) \geq Ch(A)$.

Condition R1 should be straightforward: the chance function produced by the refinement rule accommodates fully the experimental factor that prompts revision, and so it assigns chance 1 to the proposition about that factor. Condition R2 says if Ch has already accommodated an experimental factor, then it cannot be further refined by that factor. Finally, condition R3 expresses the thought that the refinement rule never decreases the chances of propositions that imply the proposition about a given experimental factor.

I now turn to some relations between C1 and C2. The first important relation between the two conditions is straightforward. It is captured by the following proposition:

Proposition 3.1. *If Ch satisfies C1, then it follows that Ch satisfies C2.*

Proof. Suppose that Ch satisfies C1 and that $C_{ch_w} \in \mathcal{A}$ for all w in $\mathcal{W}$. Then,

$$\begin{aligned} Ch(A) &= Ch\left(\bigvee_{w \in \mathcal{W}} (A \wedge C_{ch_w})\right) \\ &= \sum_{w \in \mathcal{W}} Ch(A \wedge C_{ch_w}) \\ &= \sum_{w \in \mathcal{W}} Ch(A|C_{ch_w})Ch(C_{ch_w}) \\ &= \sum_{w \in \mathcal{W}} Ch(C_{ch_w})ch_w(A) \text{ (by C1).} \end{aligned}$$

□

What other relations could hold between conditions C1 and C2? Interestingly, under a certain assumption, C1 and C2 come as equivalent conditions. The assumption says that any possible later chance function ch_w is "certain" in giving the correct chance distribution, i.e. $ch_w(C_{ch_w}) = 1$. Call such a later chance function an *immodest* expert function.[6] Given this assumption, we can prove the following proposition:

[6] For a more thorough analysis of epistemic expert functions, see Pettigrew and Titelbaum (2014).

Proposition 3.2. *Suppose that every ch_w is an immodest expert function. Then, Ch satisfies C1 if and only if it satisfies C2.*

Proof. Suppose that $ch_w(C_{ch_w}) = 1$. We prove that Ch satisfies C1 iff it satisfies C2 in the following two steps.

(a) C1 $\Rightarrow$ C2. Suppose that Ch satisfies C1, $ch_w(C_{ch_w}) = 1$, and $C_{ch_w} \in \mathcal{A}$ for all w in $\mathcal{W}$. Then,

$$\begin{aligned} Ch(A) &= Ch\left(\bigvee_{w\in\mathcal{W}} (A \wedge C_{ch_w})\right) \\ &= \sum_{w\in\mathcal{W}} Ch(A \wedge C_{ch_w}) \\ &= \sum_{w\in\mathcal{W}} Ch(A|C_{ch_w})Ch(C_{ch_w}) \\ &= \sum_{w\in\mathcal{W}} Ch(C_{ch_w})ch_w(A) \text{ (by C1)}, \end{aligned}$$

as required.

(b) C2 $\Rightarrow$ C1. Suppose that Ch satisfies C2, $ch_w(C_{ch_w}) = 1$, and $C_{ch_w} \in \mathcal{A}$ for all w in $\mathcal{W}$. Then,

$$\begin{aligned} Ch(A|C_{ch_w}) &= \frac{Ch(A \wedge C_{ch_w})}{Ch(C_{ch_w})} \\ &= \frac{\sum_{w'\in\mathcal{W}} Ch(C_{ch_{w'}})ch_{w'}(A \wedge C_{ch_w})}{\sum_{w'\in\mathcal{W}} Ch(C_{ch_{w'}})ch_{w'}(C_{ch_w})} \text{ (by C2)} \\ &= \frac{ch_w(A \wedge C_{ch_w})}{ch_w(C_{ch_w})} \\ &\text{(since } ch_{w'}(C_{ch_{w'}}) = 1) \\ &= ch_w(A|C_{ch_w}) \\ &= ch_w(A), \end{aligned}$$

as required.

□

Assuming that every later chance function is immodest, this result in fact shows that we deal not with two conditions, but with one condition. The condition relates two chance or physical probability functions in a way that parallels the way in which various chance-credence principles relate one's subjective probability and physical probability functions. That is, it requires a present chance function to line up with some possible later chance function. Putting some qualifications aside, like the admissibility clause, this condition might be regarded as the chance-counterpart of Lewis's famous Principal Principle.

But why should we assume that every later chance function is immodest? Interestingly, it turns out that this assumption plays an important role when it is conjoined with condition C1 and with the demand that any chance function should be a probability function. If we allow any possible later chance function to be *modest*, i.e. $ch_w(C_{ch_w}) < 1$, then we get a contradiction when we require any present chance function Ch to obey C1. To see this, suppose that a given later chance function ch_w is modest and that the present chance function Ch satisfies C1. Assume further that $Ch(C_{ch_w}) > 0$. Then, we get the following two contradictory conclusions:

- By probability theory, $Ch(C_{ch_w}|C_{ch_w}) = 1$.
- By C1, we have $Ch(C_{ch_w}|C_{ch_w}) = ch_w(C_{ch_w}) < 1$, since ch_w is modest.

Thus, it appears that a consistent application of condition C1 requires any possible later chance function to be immodest.

Before closing this section, let me state one consequence of condition C2. It can be formulated as a condition relating present and possible later expectations, calculated relative to present and later chance functions. To introduce this condition, let us define the expectation of a random variable $X : \mathcal{W} \to \mathbb{R}$, calculated relative to a chance function:

$$E_{ch}(X) = \sum_{w \in \mathcal{W}} ch(w)X(w). \tag{3.2}$$

Then, the following condition is a consequence of C2:

C3 (*Present expectations are equal to the expected values of later expectations*)**:** For all Ch and ch_w,

$$E_{Ch}(X) = \sum_{w \in \mathcal{W}} Ch(C_{ch_w}) E_{ch_w}(X).$$

That is, C3 requires an expectation of a random variable, calculated relative to the present chance function, to be equal to an expectation of later expectations, calculated relative to the possible later chance functions. To show that C2 entails C3, let me prove the following proposition:

Proposition 3.3. *If Ch satisfies C2, then it follows that Ch satisfies C3.*

Proof. Suppose that Ch satisfies C2 and that $C_{ch_w} \in \mathcal{A}$ for all w in $\mathcal{W}$. Then,

$$\begin{aligned}
E_{Ch}(X) &= \sum_{w \in \mathcal{W}} Ch(w)X(w) \\
&= \sum_{w \in \mathcal{W}} \left(\sum_{w \in \mathcal{W}} Ch(C_{ch_w}) ch_w(w) X(w) \right) \quad \text{(by C2)} \\
&= \sum_{w \in \mathcal{W}} \sum_{w \in \mathcal{W}} Ch(C_{ch_w}) ch_w(w) X(w) \\
&= \sum_{w \in \mathcal{W}} Ch(C_{ch_w}) \sum_{w \in \mathcal{W}} ch_w(w) X(w) \\
&= \sum_{w \in \mathcal{W}} Ch(C_{ch_w}) E_{ch_w}(X),
\end{aligned}$$

which yields C3. $\square$

3.3 Chance and Resiliency

In this section, I introduce, motivate, and make precise a norm for chances called *Maximizing Resiliency.* This norm captures a fundamental intuition concerning chance, to wit, the intuition that chances should be as stable as possible under variation of experimental factors. This norm, suitably understood, is the linchpin of our resiliency-centered arguments for conditions C1 and C2. That is, I will show in the next sections how these two conditions follow from the norm of

Maximizing Resiliency.

3.3.1 Maximizing Resiliency and Resiliency Measure

Our fundamental norm governing chances runs as follows:

> **Maximizing Resiliency:** Chance should maximize its resiliency in the presence of experimental factors. That is, a chance distribution ought to be maximally invariant upon variation of experimental factors.

Before I elaborate on this norm, let me briefly provide a pre-theoretic motivation for the resiliency-centered approach to chance given in this chapter. I begin with a presentation of Skyrms's (1980a) reasons for why resiliency considerations are so crucial to chance.

First, Skyrms argues that scientific methodology aims at achieving resilient chances that figure in statistical laws. This goal stems from an important canon of scientific methodology known as Bacon's rule of varying the circumstances. It says that to test a law, we have to vary as much as possible the conditions that are not described by the law. Second, he claims that the resiliency of chance is closely tied to the notion of *statistical necessity*. One way to clarify this claim is to say that a high resiliency of chance posited in a statistical law means that this law obtains almost unconditionally. Thus, resiliency is a way to capture the necessity of statistical laws. Since resiliency comes in degrees, this observation in turn implies that the necessity of statistical laws might be gradual. In recent years a similar idea has been developed in connection with attempts to explain lawful generalizations. Various argumentative strategies have been put forward to defend the view that any lawful generalization, including a statistical one, must be sufficiently "resilient", "stable", "robust", or "invariant" under certain sorts of changes, and that the degree of its stability indicates how immutable or necessary it is.

Let me mention two attempts of this kind. Mitchell (2000) defends the view that generalizations admit of various degrees of stability, and consequently there is no simple divide between accidental and lawful generalizations. On Mitchell's view, a generalization's stability is the extent to which it is contingent. As an example, generalizations covering biological phenomena, like Mendel's laws, seem to be more contingent (less robust) than fundamental physical laws, but are less contingent (more robust) than generalizations of the kind "All the coins in my pocket are 1 cent euro coins". Using a different conceptual framework, Woodward (2003, chapter 6) argues that invariance across certain changes is the key to understand successful explanatory generalizations. One of his major goals is to account for the explanatory role of generalizations of the special sciences. Although those generalizations have exceptions, they should be regarded as lawful because of their stability under a wide range of changes or interventions, he claims. Like Mitchell, he argues that, given that invariance comes with gradations, the standard law versus accident framework fails to give us a full understanding of possible explanatory and causal relationships that various generalizations are supposed to cover.

Third, and perhaps most importantly, Skyrms claims that the resiliency of a chance (propensity) is a mark of the *confirmatory value* of a statistical law that posits this propensity, and argues that well-confirmed statistical laws posit highly resilient chances. Resiliency thus is evidentially significant. More precisely, he argued that "A propensity statement is well confirmed if the epistemic probability is high that the correct objective distribution gives the probability attribution at issue high resiliency. In other words, confirmation of propensity statements goes by epistemic expectation of objective resiliency" (Skyrms 1980a, p. 71). Well-confirmed statistical laws thus should have a high epistemic expectation of the resiliency of chances or propensities they assign to experimental outcomes. In sum, Skyrms shows that the resiliency of chance is closely related to various concepts concerning statistical laws, their modal status and their testing.

Interestingly, there are other important reasons for adopting a resiliency-based approach to chance. First, and more generally, resiliency viewed as a property of chance distributions is strictly connected with the strategy of scientific model building that aims at finding "robust" models. In his classical paper on model building in population biology, Richard Levins (1966) developed the idea that model building in population biology involves a trade-off among realism, precision, and generality, and that population biology should aim at finding robust models, including robust statistical models. According to Levins, such models supply an access to the truth about biological reality. On this view thus, robustness is truth conducive. Here I do not want to adjudicate whether robustness is a mark of truth. Perhaps, robust models are worthwhile, regardless of whether or not they are truth conducive. What is important in the context of this chapter is that the idea of chance resiliency seems to fit well with the idea of robust statistical models. One natural thought seems to be that a robust statistical model posits a chance distribution that is invariant across variation in data. The robustness of a model, or a degree thereof, might in turn be revealed by the extent to which that chance distribution is resilient across variation of data.

Second, resiliency appears to be a key feature of a certain sort of objective probabilities that figure in probabilistic explanations. Aidan Lyon (2010) calls such objective probabilities *counterfactual probabilities*. These probabilities are distinguished by playing a specific conceptual role in probabilistic explanations: they convey modally comparative information about a system. On this view, counterfactual probability is a measure of how robust a proposition is under variation of experimental factors that could impinge on the system. This variation proceeds by considering counterfactual situations. As an example of counterfactual probabilities, consider probabilities that figure in explanations given by classical statistical mechanics. For concreteness, consider an explanation of why a cup of coffee cools down. Classical statistical mechanics tells us that this is because, given a probability distribution over initial conditions, it is overwhelmingly probable that its micro-state is one that lies on a trajectory that deterministically takes it into the macro-state "cooled down". What makes these explanations satisfactory is that, among other things, these probabilities are highly resilient: we are told that the overwhelming majority of micro-states compatible with a given macro-state would evolve to a higher-entropy macro-state. So even if we knew these micro-states, this would have a negligible impact on the probabilistic explanation given by classical statistical mechanics.

Resiliency thus appears to be an important virtue of chances. Naturally, one may ask: how important is this virtue? Clearly, resiliency is not the sole virtue of chances. There are other virtues of chances that scientists seem to value. For example, they value chance distributions that match relative frequencies with which events actually occur or chance distributions that are predictively accurate.

However, to run our resiliency-based approach we do not need to decide how important resiliency is when compared to these other virtues. For our purposes, it suffices to assume that resiliency is one of many virtues that a good probabilistic theory should have.

Having said that, I now turn to clarify the norm of Maximizing Resiliency. In so doing, let me first provide a more precise way of understanding our key concept, to wit, the resiliency of a chance function. Like Skyrms (1980a), I take it that the resiliency of a chance distribution is a measurable property. My proposal is that a resiliency measure should tell us how stable a present chance distribution over some family of propositions is when compared to a chance distribution that is obtained from the present one by accommodating information about an experimental factor, understood as a supposition, that obtains at some possible world. Given that this accommodation proceeds via some refinement rule, we compare in fact an unrefined chance function with its possible refinements. To provide a resiliency measure of this sort, we use the following background assumptions:

- Let $\mathcal{F}$ be a set of experimental factors. $\mathcal{F}$ is taken to be a finite partition $\{F_w : w \in \mathcal{W}\}$ of $\mathcal{W}$. Since $\mathcal{F}$ is a partition, no two elements from $\mathcal{F}$ are true at the same w.

- For each possible world w and an experimental factor F_w that holds true at w, we define the possible later chance function as:

$$ch_w(\cdot) = RE(Ch, \mathcal{F}, F_w).$$

  That is, a possible later chance function ch_w is obtained from the present chance function Ch by accommodating via a refinement rule the experimental factor F_w that holds true at w.

- We assume that $\mathcal{C}$ is a set of present chance functions and $\mathcal{C}_{\mathcal{F}} = \{ch_w : w \in \mathcal{W}\}$ is the set of possible later chance functions.

With this background in mind, we can introduce a resiliency measure for the present chance function Ch at a world w. Mathematically, it is a function $\mathbf{R}$ that takes the present chance function Ch and a world w, and returns a score

$\mathbf{R}(Ch, w)$, a number in $\mathbb{R}$, that measures the resiliency of Ch at w. Given this notion, it is natural to ask: how should the resiliency of Ch at w be computed?

My proposal is that it is computed by measuring a divergence between the present chance function Ch and a possible later chance function ch_w that comes from Ch by accommodating via a refinement rule the F_w, which is true at w. Thus, the resiliency at w expresses how "close" the present chance function is to the possible later chance function, ch_w, which is a possible refinement of Ch at w. More precisely, if $[0,1]^n \subset \mathbb{R}^n$ is the unit cube of chance functions over a finite set of propositions $\mathcal{A} = \{A_1, ..., A_n\}$, then $D : [0,1]^n \times [0,1]^n \to \mathbb{R}$ is a divergence that takes $Ch \in [0,1]^n$ and $ch_w \in [0,1]^n$, and gives a real number in $\mathbb{R}$. We assume that for all $Ch, ch_w \in [0,1]^n$,

- $D(ch_w, Ch)$ is non-negative, i.e. $D(ch_w, Ch) \geq 0$ with equality if and only if $Ch = ch_w$.

However, we do not need to assume that $D(ch_w, Ch)$ is a *metric* in the mathematical sense. That is, we do not require it to satisfy the triangle inequality nor to be symmetric. For concreteness, let us specify $D(ch_w, Ch)$ as

$$D(ch_w, Ch) = \sum_{A \in \mathcal{A}} \big(ch_w(A) - Ch(A)\big)^2. \tag{3.3}$$

That is, $D(ch_w, Ch)$ is given by the sum over the propositions $A \in \mathcal{A}$ of the squares of differences between the present chance assigned to A by Ch and the later chance assigned to A by ch_w. If $\mathcal{A} = \{A_1, ..., A_n\}$, then we can think of Ch and ch_w as real-valued vectors, $(Ch(A_1), ..., Ch(A_n))$ and $(ch_w(A_1), ..., ch_w(A_n))$ respectively, in n-dimensional space equipped with the Euclidean inner product. Then, $D(ch_w, Ch)$ is the squared Euclidean distance between these vectors. Now, a resiliency measure that corresponds to this divergence function can be given by

$$\mathbf{R}(Ch, w) = \sum_{A \in \mathcal{A}} \big(ch_w(A) - Ch(A)\big)^2. \tag{3.4}$$

Having specified the resiliency measure, we can now clarify the norm of Maximizing Resiliency. Let $\mathcal{C}$ be a set of present chance functions. Suppose that Ch, Ch' are in $\mathcal{C}$. We say that:

- Ch' *strongly resiliency-dominates* Ch if $\mathbf{R}(Ch', w) < \mathbf{R}(Ch, w)$ for all worlds w in $\mathcal{W}$,
- Ch' *weakly resiliency-dominates* Ch if $\mathbf{R}(Ch', w) \leq \mathbf{R}(Ch, w)$ for all worlds w in $\mathcal{W}$, and $\mathbf{R}(Ch', w) < \mathbf{R}(Ch, w)$ for at least one world w in $\mathcal{W}$.

That is, Ch' *strongly resiliency-dominates* Ch if it is less divergent from the later chance function in every possible world. And Ch' *weakly resiliency-dominates* Ch if it is at least as divergent as Ch is from the later chance function in every possible world, and it is less divergent from the later chance function in at least one possible world.

Now, I propose to think of the norm of Maximizing Resiliency more precisely as follows:

> **Maximizing Resiliency*:** Suppose that Ch and Ch' are in $\mathcal{C}$. If
>
> (i) Ch' *strongly resiliency-dominates* Ch, and
>
> (ii) there is no other Ch'' in $\mathcal{C}$ that *weakly resiliency-dominates* Ch',
>
> then Ch is inadmissible with respect to the resiliency measure.

Thus, Maximizing Resiliency* tells us which present chance functions in $\mathcal{C}$ are inadmissible with respect to the resiliency measure. According to this norm, an inadmissible present chance function is one that is strongly resiliency-dominated by some other present chance function that itself is not weakly resiliency-dominated by any other present chance function in $\mathcal{C}$. Our norm might be regarded, *mutatis mutandis*, as a version of the Dominance Principle exploited in statistical decision theory. This norm goes as follows:

> **Dominance:** Let o and o' be two options in a set of options $\mathcal{O}$. Then, if
>
> (i) o' *strongly dominates* o, and
>
> (ii) there is no other o'' in $\mathcal{O}$ that *weakly dominates* o', then
>
> o is an inadmissible choice from $\mathcal{O}$.

Like Dominance, Maximizing Resiliency* signals which element of some set should be ruled out as inadmissible. At this point, one might wonder whether clause (ii) in Maximizing Resiliency* could be dropped. That is, one might argue for the following alternative norm:

> **Maximizing Resiliency**:** Suppose that Ch and Ch' are in $\mathcal{C}$. If Ch' *strongly resiliency-dominates* Ch, then Ch is inadmissible with respect to the resiliency measure.

Maximizing Resiliency** does not require that at least one present chance function in $\mathcal{C}$ is undominated. It tells us that no chance function that is strongly resiliency-dominated is admissible. However, Maximizing Resiliency* does not seem satisfactory. For it allows for the possibility that all present chance functions in $\mathcal{C}$ are ruled out as inadmissible with respect to resiliency. This is so because it does not guarantee that there is a present chance function that is not dominated by some other chance function in $\mathcal{C}$. That is, it does not guarantee that there is a stable dominance-stopping chance function in $\mathcal{C}$. Maximizing Resiliency* makes it explicit that strong dominance signals inadmissibility in case there is a dominance-stopping chance function in $\mathcal{C}$.

In the next section, I show how Maximizing Resiliency* establishes conditions C1 and C2. That is, I show that any present chance function that violates each of the conditions C1 and C2 is strongly resiliency-dominated by a present chance function in $\mathcal{C}$ that satisfies these conditions, whereas any chance function that satisfies each of the conditions is not even weakly resiliency-dominated by some other chance function in $\mathcal{C}$. Thus, we show that any chance function that violates each of the conditions C1 and C2 must be regarded as inadmissible with respect to the resiliency measure.

3.3.2 Comparison with Skyrms's Resiliency Measure

The resiliency measure presented above differs importantly from Skyrms's resiliency measure. Slightly reformulated for the purpose of comparison, Skyrms's resiliency measure might be understood as a mathematical function $\mathbf{R}_{\mathrm{S}}$ that takes proposition $A \subseteq \mathcal{W}$, a present chance of A, $Ch(A)$, and a finite set $\mathcal{B}$ of propositions describing truth-functional combinations of experimental factors that are logically consistent with A and $\neg A$, and gives a measure of the resiliency of the chance of A under $\mathcal{B}$. Resiliency, so understood, is then measured by

$$\mathbf{R}_{\mathrm{S}}(A, Ch(A), \mathcal{B}) = 1 - \max_{B \in \mathcal{B}} \big(Ch(A) - ch_B(A)\big), \tag{3.5}$$

where B ranges over propositions describing truth-functional combinations of experimental factors, and $ch_B(A)$ is the conditional chance of A obtained from $Ch(A)$ by conditioning on B. That is, the resiliency of $Ch(A)$ under $\mathcal{B}$ is the maximal change of $Ch(A)$ over all elements in $\mathcal{B}$. The divergence that underpins Skyrms's resiliency measure is the variation distance between $Ch(A)$ and $ch_B(A)$.

Let us point out some crucial differences between Skyrms's resiliency measure and the resiliency measure given in (3.4). First, Skyrms's resiliency measure measures the resiliency of a particular present chance assignment for A, $Ch(A)$, while our resiliency measure measures the resiliency of the whole present chance distribution, Ch, over $\mathcal{A}$. Second, Skyrms deals with the resiliency under the algebra $\mathcal{B}$ that contains non-empty and consistent Boolean combinations of propositions describing experimental factors, whereas we deal with the resiliency at a particular world w or, equivalently, under a particular experimental factor from $\mathcal{F}$ that obtains at w. We thus assume that $\mathcal{F}$ is a partition: no two elements F and F' in $\mathcal{F}$ are both true at w. This is so because if the elements of $\mathcal{F}$ are not disjoint, then F_w is not well defined and hence ch_w cannot be determined. Third, Skyrms's resiliency measure compares the present chance of A with the chance of A that is "obtainable" from the present one by *conditioning* on some truth-functional combination of experimental factors. For reasons given in section 3.2, we do not assume REC as the refinement rule for chances. Simply, to measure the resiliency of a present chance function at w, we take a divergence between that function and its refinement at w obtained by accommodating via some refinement rule an experimental factor that obtains at w. Thus, the refinement of a present chance in the light of some experimental factor proceeds in a "black-box": we are only given information about the input (a present chance function) and the output(a refined present chance function at w), without specifying precisely the rule by which the present chance distribution is refined. Fourth, on Skyrms's view, the divergence between $Ch(A)$ and $ch_B(A)$ is given by the largest amount of disagreement between $Ch(A)$ and $ch_B(A)$, where B ranges over truth-functional combinations of propositions describing experimental factors, whereas, on our view, the divergence between Ch and ch_w is given by the sum over the propositions $A \in \mathcal{A}$ of the squares of differences between Ch and ch_w.

3.3.3 Generalizing Resiliency Measure: Bregman Divergence

The resiliency-based arguments for C1 and C2 to be presented do not hinge essentially on the resiliency measure given in (3.4). As will be apparent, these arguments hold for a class of resiliency measures that are instances of the so-called Bregman divergence.

To introduce Bregman divergence more generally in this chapter, let $\mathcal{A} = \{A_1, ..., A_n\}$, and let $\mathbf{x}$ and $\mathbf{y}$ be real-valued vectors in n-dimensional space equipped with the Euclidean inner product. Suppose that $\mathcal{X}$ is a convex subset of $\mathbb{R}^n$. Let $\Phi : \mathcal{X} \to \mathbb{R}$ be a strictly convex function whose gradient $\nabla\Phi$ is defined in the interior of $\mathcal{X}$ and extends to a bounded, continuous function on $\mathcal{X}$. Then for all $\mathbf{x}, \mathbf{y} \in \mathcal{X}$, the Bregman divergence $D_\Phi : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ corresponding to Φ is given by

$$D_\Phi(\mathbf{y}, \mathbf{x}) = \Phi(\mathbf{y}) - \Phi(\mathbf{x}) - \nabla\Phi(\mathbf{x}) \cdot (\mathbf{y} - \mathbf{x}), \tag{3.6}$$

where $\nabla\Phi(\mathbf{x})$ is the gradient vector of Φ evaluated at $\mathbf{x}$ and $\cdot$ denotes the inner product of two vectors (Figure 3.1 visualizes the Bregman divergence). Since Φ is strictly convex, it follows that $D_\Phi(\mathbf{y}, \mathbf{x}) \geq 0$ with equality if and only if $\mathbf{y} = \mathbf{x}$. The function D_Φ is the difference between the value of Φ at $\mathbf{x}$ and the first-order Taylor expansion of Φ around $\mathbf{x}$ evaluated at $\mathbf{y}$. Now, if the function Φ is defined as $\Phi(\mathbf{x}) = \|\mathbf{x}\|^2$, then $D_\Phi(\mathbf{y}, \mathbf{x}) = \|\mathbf{y} - \mathbf{x}\|^2$. That is, squared Euclidean distance is a Bregman divergence (Banerjee et al. 2005).

More specifically, for our purposes let us assume that $\mathcal{X}$ is the unit cube, $[0, 1]^n$, in $\mathbb{R}^n$ and let Ch and ch_w be represented as real-valued vectors of chances, respectively $Ch = (Ch(A_1), ..., Ch(A_n))$ and $ch_w = (ch_w(A_1), ..., ch_w(A_n))$, in $\mathbb{R}^n$ equipped with the Euclidean inner product. Then for all $Ch, ch_w \in [0, 1]^n$,

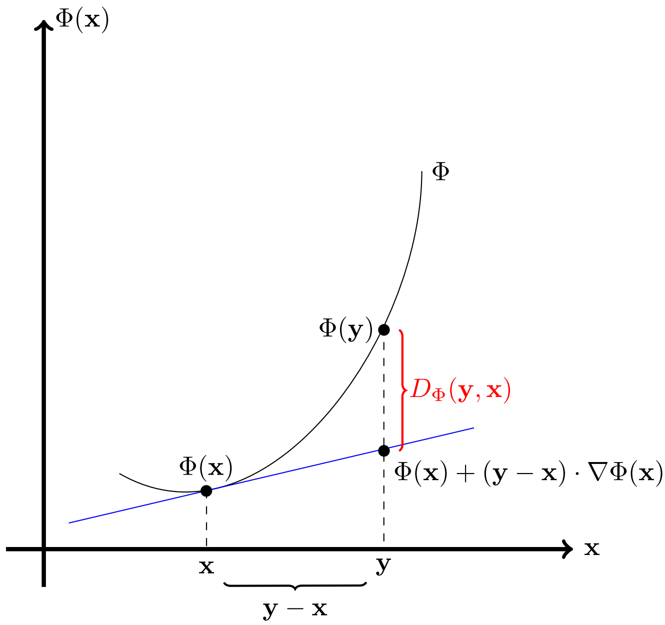

Figure 3.1: Visualizing the Bregman divergence generated by Φ. The Bregman divergence from $\mathbf{y}$ to $\mathbf{x}$ is a difference between the value of two functions at $\mathbf{y}$. The first function is Φ and the second function is the tangent to Φ taken at $\mathbf{x}$.

the Bregman divergence $D_\Phi : [0,1]^n \times [0,1]^n \to \mathbb{R}$ corresponding to Φ is given by

$$D_\Phi(ch_w, Ch) = \Phi(ch_w) - \Phi(Ch) - \nabla\Phi(Ch) \cdot (ch_w - Ch). \tag{3.7}$$

The following proposition (for a proof see Predd et al. 2009, p. 4789) captures a feature of any Bregman divergence, feature that will play a crucial role in our resiliency-centered arguments for C1 and C2:

Proposition 3.4. *Let $\mathcal{Z} \subseteq [0,1]^n$ be a closed convex subset of $\mathbb{R}^n$. Then, if $Ch \in [0,1]^n - \mathcal{Z}$, there is $Ch' \in \mathcal{Z}$ such that $D_\Phi(Ch', Ch) \leq D_\Phi(ch_w, Ch)$ for all $ch_w \in \mathcal{Z}$. Moreover, $D_\Phi(ch_w, Ch') \leq D_\Phi(ch_w, Ch) - D_\Phi(Ch', Ch)$ for all*

$ch_w \in \mathcal{Z}$ and all $Ch \in [0,1]^n - \mathcal{Z}$.

The key idea behind this proposition is this. Given an unrefined chance function that is not in $\mathcal{Z}$, there is an unrefined chance function in $\mathcal{Z}$ that is at least as close to the chance function outside $\mathcal{Z}$ as any other chance function in $\mathcal{Z}$. The vector of chances Ch' is called the projection of Ch onto $\mathcal{Z}$. The second sentence of this proposition expresses the fact that any Bregman divergence satisfies the generalized Pythagorean theorem, i.e. $D_\Phi(ch_w, Ch) \geq D_\Phi(ch_w, Ch') + D_\Phi(Ch', Ch)$.

3.4 Resiliency-based Arguments for C1 and C2

We first establish a theorem which is crucial for the task of showing how conditions C1 and C2 follow from the stricture of maximizing resiliency. Let $\mathcal{F} = \{F_w : w \in \mathcal{W}\}$ be a finite partition of experimental factors, and let $\mathcal{C}_\mathcal{F} = \{ch_w : w \in \mathcal{W}\}$ be the corresponding set of possible later chance functions. These chance functions are possible refinements of some chance function in $\mathcal{C}$. We introduce the set of all convex combinations of the possible later chance functions in $\mathcal{C}_\mathcal{F}$, called the convex hull of $\mathcal{C}_\mathcal{F}$. It can be defined as follows:

$$Conv(\mathcal{C}_\mathcal{F}) = \left\{ \sum_{w \in \mathcal{W}} \lambda_w ch_w : 0 \leq \lambda_w \leq 1, \sum_{w \in \mathcal{W}} \lambda_w = 1 \right\}. \tag{3.8}$$

That is, $Conv(\mathcal{C}_\mathcal{F})$ is the smallest set that (i) contains $\mathcal{C}_\mathcal{F}$, and (ii) contains, for any two later chance functions ch_w and $ch_{w'}$, every convex combination or mixture of them. i.e. for any $0 \leq \lambda_w \leq 1$, it contains $\lambda_w ch_w + (1 - \lambda_w) ch_{w'}$. Alternatively, $Conv(\mathcal{C}_\mathcal{F})$ may be defined as the intersection of all convex sets containing $\mathcal{C}_\mathcal{F}$.

Now, endowed with these notions, we can establish the following theorem:

Theorem 3.1. *Let $\mathbf{R}(Ch, w) = D_\Phi(ch_w, Ch)$. Suppose $\mathcal{C}_\mathcal{F} \subseteq [0,1]^n$. Then:*

(i) If $Ch \notin Conv(\mathcal{C}_\mathcal{F})$, then there is $Ch' \in Conv(\mathcal{C}_\mathcal{F})$ such that $D_\Phi(ch_w, Ch') < D_\Phi(ch_w, Ch)$ for all $ch_w \in \mathcal{C}_\mathcal{F}$.

(ii) If $Ch' \in Conv(\mathcal{C}_\mathcal{F})$, then there is no $Ch'' \in [0,1]^n$ such that $Ch'' \neq Ch'$ and $D_\Phi(ch_w, Ch'') \leq D_\Phi(ch_w, Ch')$ for all $ch_w \in \mathcal{C}_\mathcal{F}$, and $D_\Phi(ch_w, Ch'') <$

$D_\Phi(ch_w, Ch')$ for some $ch_w \in \mathcal{C}_\mathcal{F}$.

Proof. For the proof, we assume that:

- $\mathcal{A} = \{A_1, ..., A_n\}$ is a finite set of propositions.
- Ch and ch_w are represented as real-valued vectors of chances in the n-dimensional Euclidean space $\mathbb{R}^n$.
- $Conv(\mathcal{C}_\mathcal{F})$ is the convex hull of $\mathcal{C}_\mathcal{F}$. Thus, $\mathcal{C}_\mathcal{F} \subseteq Conv(\mathcal{C}_\mathcal{F})$.

For the proof of Theorem 3.1 (i), suppose that $Ch \in [0,1]^n - Conv(\mathcal{C}_\mathcal{F})$. Then, by Proposition 3.4, for all $ch_w \in Conv(\mathcal{C}_\mathcal{F})$ and hence for all $ch_w \in \mathcal{C}_\mathcal{F}$,

$$D_\Phi(ch_w, Ch') \leq D_\Phi(ch_w, Ch) - D_\Phi(Ch', Ch).$$

And so,

$$D_\Phi(ch_w, Ch) \geq D_\Phi(ch_w, Ch') + D_\Phi(Ch', Ch).$$

Since $Ch \notin Conv(\mathcal{C}_\mathcal{F})$ and $Ch' \in Conv(\mathcal{C}_\mathcal{F})$, we have that $Ch \neq Ch'$, and so $D_\Phi(Ch', Ch) > 0$. So

$$D_\Phi(ch_w, Ch') < D_\Phi(ch_w, Ch),$$

as required.

For the proof of Theorem 3.1 (ii), suppose that $Ch' \in Conv(\mathcal{C}_\mathcal{F})$. Now, suppose that $D_\Phi(ch_w, Ch'') \leq D_\Phi(ch_w, Ch')$ for all $ch_w \in Conv(\mathcal{C}_\mathcal{F})$, and hence for all $ch_w \in \mathcal{C}_\mathcal{F}$. Since $Ch' \in Conv(\mathcal{C}_\mathcal{F})$, it follows that $D_\Phi(ch_w, Ch'') \leq D_\Phi(Ch', Ch') = 0$. This implies that $D_\Phi(ch_w, Ch'') = 0$ and $Ch'' = Ch'$. But this contradicts the assumption that $Ch'' \neq Ch'$. □

Our theorem has two parts. In the first part, it says that if a present chance function Ch is not in $Conv(\mathcal{C}_\mathcal{F})$, then there is another present chance function Ch' that is in $Conv(\mathcal{C}_\mathcal{F})$ and is more resilient at all possible worlds or, equivalently, is closer to every possible later chance function. The basic idea of this part of our theorem is that any unrefined chance function that lies outside the convex hull of $\mathcal{C}_\mathcal{F}$ can be replaced by some unrefined chance function in the convex hull of

$\mathcal{C}_{\mathcal{F}}$ that is more resilient at all possible worlds. The second part of the theorem ensures that once the unrefined chance function Ch' is in $Conv(\mathcal{C}_{\mathcal{F}})$, there is no other unrefined chance function $Ch'' \in \mathcal{C}$ that could be more resilient at all possible worlds. In other words, the theorem says that if the chance function Ch is in the convex hull of $\mathcal{C}_{\mathcal{F}}$, then its resiliency would not be improved by replacing it with any other chance function in $\mathcal{C}$.

It has to be emphasized that this theorem hinges on the assumption that the set of possible refinements $\mathcal{C}_{\mathcal{F}}$ is hold fixed: it is fixed by some refinement rule that maps Ch and the elements of $\mathcal{F}$ to the ch_w's. We then consider whether Ch should be replaced with some other unrefined chance function Ch' from $\mathcal{C}$ by looking at how the two unrefined chance functions diverge from every element of the fixed set $\mathcal{C}_{\mathcal{F}}$. Thus, our result relies heavily on the assumption that the chance functions in $\mathcal{C}$ give rise to the same refined chances in $\mathcal{C}_{\mathcal{F}}$. At this juncture, however, it might be objected that since every possible refined chance depends both on the element of $\mathcal{F}$ and on the unrefined chance function Ch, the assumption is an unacceptable oversimplification. We should expect that for any two different unrefined chance functions in $\mathcal{C}$, the refinement rule would produce different sets of possible refined chances. Consequently, we should consider different convex hulls of these sets.

My response to this objection is that our framework does not require the refinement rule to track or conserve possible unrefined chance functions in $\mathcal{C}$. Hence, it is entirely possible for this rule to produce refined chance functions that track only the experimental factors in $\mathcal{F}$ and do not conserve unrefined chance functions. That is, any refined chance function could satisfy only the constraints imposed by the experimental factors and be far removed from the unrefined chance function. An instance of a refinement rule that tracks both the experimental factors in $\mathcal{F}$ and the unrefined chance function is REC: a refined chance is a given unrefined chance function conditional on a true member of $\mathcal{F}$. But we have given reasons in section 3.2 for why REC is not a good choice in our framework.

3.4.1 A Resiliency-based Argument for C2

I start by providing a resiliency-based argument for C2. The reason for approaching our task in this order is that a resiliency-based argument for C1 hinges on the assumption that every $ch_w \in \mathcal{C}_{\mathcal{F}}$ is an immodest expert function. And given Proposition 3.2, it follows that such an argument for C1 is automatically an argument for C2. But I would like to examine whether resiliency considerations provide support for C2 independently of how C1 and C2 are related. I thus proceed in this section without assuming that the possible later chance functions are immodest. However, I make another additional assumption about the elements in $\mathcal{C}_{\mathcal{F}}$: it is required that each later chance of A is the expectation of the later chances of A calculated relative to the later chance function. That is, for all $v \in \mathcal{W}$ and all $A \in \mathcal{A}$,

$$ch_v(A) = \sum_{w \in \mathcal{W}} ch_v(C_{ch_w}) ch_w(A). \tag{3.9}$$

Thus, each ch_v "expects" itself to give the later chance of A. It is easy to make this assumption seem plausible. To display more effectively the role played by this assumption, recall the theory of expert functions adduced at the beginning of this chapter. If $Ch(A)$ is the expectation of the expert values for A, then it seems that each expert value for A should itself be an expectation of the expert values for A. And this intuition is captured by (3.9): each expert value for A expects itself to give the expert value for A.

Our resiliency-based argument for C2 comprises the norm of Maximizing Resiliency*, Theorem 3.1, and the following theorem:

Theorem 3.2. *Suppose that $\mathcal{C}_{\mathcal{F}}$ is the set of possible later chance functions. And suppose that for all $v \in \mathcal{W}$ and all $A \in \mathcal{A}$, $ch_v(A) = \sum_{v \in \mathcal{W}} ch_v(C_{ch_w}) ch_v(A)$. Then, Ch satisfies C2 if and only if $Ch \in Conv(\mathcal{C}_{\mathcal{F}})$.*

Proof. The first step is to show that if $Ch \in Conv(\mathcal{C}_{\mathcal{F}})$, then Ch satisfies C2. Since $Conv(\mathcal{C}_{\mathcal{F}})$ contains $\mathcal{C}_{\mathcal{F}}$ and every convex combination of any two elements in $\mathcal{C}_{\mathcal{F}}$, we proceed as follows. First, we show that if $Ch \in \mathcal{C}_{\mathcal{F}}$, then Ch satisfies C2. And this is straightforward, since for all $v \in \mathcal{W}$ and all $A \in \mathcal{A}$, $ch_v(A) = \sum_{w \in \mathcal{W}} ch_v(C_{ch_w}) ch_w(A)$.

Second, we show that if Ch and Ch' satisfy C2, then so does any convex combination of Ch and Ch'. Suppose that Ch and Ch' satisfy C2. Then,

$$\begin{aligned}\lambda Ch + (1-\lambda)Ch'(A) &= \sum_{w\in\mathcal{W}} \lambda Ch(C_{ch_w})ch_w(A) + \sum_{w\in\mathcal{W}} (1-\lambda)Ch'(C_{ch_w})ch_w(A)\\ &= \sum_{w\in\mathcal{W}} \lambda Ch(C_{ch_w}) + (1-\lambda)Ch'(C_{ch_w})ch_w(A),\end{aligned}$$

as required.

The second step is to show that if Ch satisfies C2, then $Ch \in Conv(\mathcal{C}_{\mathcal{F}})$. Suppose that Ch satisfies C2. It follows then that Ch is a convex combination of the posterior chance functions, the ch_ws, in $\mathcal{C}_{\mathcal{F}}$. That is, there are λ_w such that $\sum_{w\in\mathcal{W}} \lambda_w = 1$ and for $\lambda_w = Ch(C_{ch_w})$,

$$Ch = \sum_{w\in\mathcal{W}} \lambda_w ch_w.$$

□

Theorem 3.2 thus identifies the set $Conv(\mathcal{C}_{\mathcal{F}})$ with the set of all possible later chance functions that satisfy C2. To establish this result, we have used the following strategy. Since $Conv(\mathcal{C}_{\mathcal{F}})$ is the intersection of all convex sets containing $\mathcal{C}_{\mathcal{F}}$ and a convex set is one that contains each convex combination of the elements in $\mathcal{C}_{\mathcal{F}}$, we have shown that for all convex sets containing $\mathcal{C}_{\mathcal{F}}$, (i) every element in $\mathcal{C}_{\mathcal{F}}$ satisfies C2 and (ii) every convex combination of the elements in $\mathcal{C}_{\mathcal{F}}$ satisfies C2.

Next, we infer the following proposition from Theorem 3.1 and Theorem 3.2:

Proposition 3.5. *Let* $\mathbf{R}(Ch, w) = D_\Phi(ch_w, Ch)$*. Suppose* $\mathcal{C}_{\mathcal{F}} \subseteq [0,1]^n$*. Then:*

(i) If Ch does not satisfy C2, then there is $Ch' \in \mathcal{C}$ that satisfies C2 and $D_\Phi(ch_w, Ch') < D_\Phi(ch_w, Ch)$ for all $ch_w \in \mathcal{C}_{\mathcal{F}}$.

(ii) If Ch' satisfies C2, then there is no other $Ch'' \in \mathcal{C}$ such that $Ch'' \neq Ch'$ and $D_\Phi(ch_w, Ch'') \leq D_\Phi(ch_w, Ch')$ for all $ch_w \in \mathcal{C}_{\mathcal{F}}$, and $D_\Phi(ch_w, Ch'') < D_\Phi(ch_w, Ch')$ for some $ch_w \in \mathcal{C}_{\mathcal{F}}$.

From Proposition 3.5 and the norm of Maximizing Resiliency*, we conclude that Ch is inadmissible with respect to the resiliency measure. That is, any Ch that violates C2 is strongly resiliency-dominated by some other Ch'' that itself is not even weakly resiliency-dominated by any other element in $\mathcal{C}$.

The argument just given thus may be formalized as follows:

(1) Maximizing Resiliency*

(2) Theorem 3.1

(3) Theorem 3.2

(4) Proposition 3.5

(5) Therefore, Ch is inadmissible with respect to the resiliency measure.

3.4.2 A Resiliency-based Argument for C1

Our resiliency-based argument for C1 combines Maximizing Resiliency*, Theorem 3.1, and the following theorem:

Theorem 3.3. *Suppose that $\mathcal{C}_{\mathcal{F}}$ is the set of all possible later chance functions. And suppose that each ch_w is an immodest expert function, i.e. $ch_w(C_{ch_w}) = 1$. Then, Ch satisfies C1 if and only if $Ch \in Conv(\mathcal{C}_{\mathcal{F}})$.*

Proof. The first step is to show that if $Ch \in Conv(\mathcal{C}_{\mathcal{F}})$, then Ch satisfies C1. Since $Conv(\mathcal{C}_{\mathcal{F}})$ contains $\mathcal{C}_{\mathcal{F}}$ and every convex combination of any two elements in $\mathcal{C}_{\mathcal{F}}$, we proceed as follows. First, we show that if $Ch \in \mathcal{C}_{\mathcal{F}}$, then Ch satisfies C1. Suppose that $v, w \in \mathcal{W}$. Then, if $ch_v(C_{ch_w}) > 0$, we have that $ch_v = ch_w$, and so

$$ch_v(A|C_{ch_w}) = ch_w(A|C_{ch_w}) = ch_w(A),$$

as required.

Second, we show that if Ch and Ch' satisfy C1, then any convex combination of Ch and Ch' satisfies C1. Suppose that Ch and Ch' are in $Conv(\mathcal{C}_{\mathcal{F}})$ and satisfy C1. Then,

$$\begin{aligned}(\lambda Ch + (1-\lambda)Ch')(A|C_{ch_w}) &= \frac{\lambda Ch(A \wedge C_{ch_w}) + (1-\lambda)Ch'(A \wedge C_{ch_w})}{\lambda Ch(C_{ch_w}) + (1-\lambda)Ch'(C_{ch_w})} \\ &= \frac{\lambda Ch(A|C_{ch_w})Ch(C_{ch_w})}{\lambda Ch(C_{ch_w})}\end{aligned}$$

$$\frac{+\ (1-\lambda)Ch'(A|C_{ch_w})Ch'(C_{ch_w})}{+\ (1-\lambda)Ch'(C_{ch_w})}$$
$$= \frac{\lambda ch_w(A)Ch(C_{ch_w}) + (1-\lambda)ch_w(A)Ch'(C_{ch_w})}{\lambda Ch(C_{ch_w}) + (1-\lambda)Ch'(C_{ch_w})}$$
(by the fact that Ch and Ch' satisfy C1
and by the axioms of probability)
$$= ch_w(A),$$

as required.

The second step of our proof is to show that if Ch satisfies C1, then $Ch \in Conv(\mathcal{C_F})$. Suppose that Ch satisfies C1 and that $C_{ch_w} \in \mathcal{A}$ for each $w \in \mathcal{W}$. Then, for any $A \in \mathcal{A}$,

$$\begin{aligned} Ch(A) &= Ch\left(\bigvee_{w\in\mathcal{W}} (A \wedge C_{ch_w})\right) \\ &= \sum_{w\in\mathcal{W}} Ch(A \wedge C_{ch_w}) \\ &= \sum_{w\in\mathcal{W}} Ch(A|C_{ch_w})Ch(C_{ch_w}) \\ &= \sum_{w\in\mathcal{W}} ch_w(A)Ch(C_{ch_w}) \text{ (by C1).} \end{aligned}$$

And for $\lambda_w = Ch(C_{ch_w})$, we have that $Ch \in Conv(\mathcal{C_F})$, as required. □

Much like Theorem 3.2, the theorem just established identifies the set $Conv(\mathcal{C_F})$ with the set of all possible later chance functions that satisfy C1. To establish this result, we have shown that for all convex sets containing $\mathcal{C_F}$, (i) every element in $\mathcal{C_F}$ satisfies C1 and (ii) every convex combination of the elements in $\mathcal{C_F}$ satisfies C1.

Now, from Theorem 3.1 and Theorem 3.3, we infer that:

Proposition 3.6. *Let* $\mathbf{R}(Ch, w) = D_\Phi(ch_w, Ch)$. *Suppose* $\mathcal{C_F} \subseteq [0,1]^n$. *Then:*

(i) If Ch *does not satisfy C1, then there is* $Ch' \in \mathcal{C}$ *that satisfies C1 and* $D_\Phi(ch_w, Ch') < D_\Phi(ch_w, Ch)$ *for all* $ch_w \in \mathcal{C_F}$.

(ii) *If* Ch' *satisfies C1, then there is no other* $Ch'' \in \mathcal{C}$ *such that* $Ch'' \neq Ch'$ *and* $D_\Phi(ch_w, Ch'') \leq D_\Phi(ch_w, Ch')$ *for all* $ch_w \in \mathcal{C}_\mathcal{W}$*, and* $D_\Phi(ch_w, Ch'') < D_\Phi(ch_w, Ch')$ *for some* $ch_w \in \mathcal{C}_\mathcal{F}$*.*

From Proposition 3.6 and from the norm of Maximizing Resiliency*, we conclude that Ch is inadmissible with respect to the resiliency measure. It means that any Ch that violates C1 is strongly resiliency-dominated by some other Ch'' that itself is not even weakly resiliency-dominated by any other element in $\mathcal{C}$.

In sum, the resiliency-based argument for C1 may be formalized as follows:

(1) Maximizing Resiliency*

(2) Theorem 3.1

(3) Theorem 3.3

(4) Proposition 3.6

(5) Therefore, Ch is inadmissible with respect to the resiliency measure.

3.4.3 Summary

We have just shown how considerations of resiliency can be adduced in order to provide support for conditions C1 and C2. We have shown that any present chance function that violates each of our conditions should be deemed as inadmissible by the norm of Maximizing Resiliency*. Informally, the gist of our resiliency-based arguments is the thought that an unrefined chance function that satisfies each of these conditions belongs to a convex hull of its possible refinements. And once this unrefined chance function is in that convex hull, it cannot be even weakly resiliency-dominated by some other unrefined chance function. But if an unrefined chance function violates these conditions, then it lies outside this convex hull and so it always can be strongly resiliency-dominated by some other unrefined chance function that satisfies each of these conditions and thus belongs to that convex hull. Therefore, any unrefined chance function that violates each of these conditions is ruled out as inadmissible by the norm of Maximizing Resiliency*.

It has to be emphasized that though considerations concerning the resiliency of chance shed new light on conditions C1 and C2, they should not be taken as providing a “bedrock” or ultimate justification of these conditions. Rather, they should be understood as providing a support for the claim that these conditions should be requirements for chance.

3.5 Some Consequences: Humean Supervenience

In this section, I discuss some philosophical consequences of our resiliency-based arguments for the debate concerning the plausibility of so-called Humean accounts of chance. First, I show that conditions C1 and C2 imposed in the case of Humean accounts of chance might lead to unacceptable consequences. Second, I focus on how our resiliency-centered approach to chance contributes to the widely discussed controversy over Humean accounts of chance, known as the Big Bad Bug. Third, I show that there are Humean accounts of chance that accommodate perfectly the two conditions.

Let me start by providing a characterization of Humean accounts of chance. They are grounded on the metaphysical doctrine of Humean Supervenience (HS). In general, HS states that modal concepts such as laws of nature, dispositions, counterfactuals, causation, or chances supervene on the global histories of non-modal facts. For the purpose of this chapter, HS applied to chance can be formulated slightly more precisely as follows:

> $\mathbf{HS_{chance}}$**:** For any worlds w and v and any time t, if w and v have the same global (past, present, and future) histories of chancy outcomes, then w and v have the same chance function at t.

In view of HS_{chance}, Humean accounts of chance might be called reductionist, for they treat chances are being fixed or determined by non-modal features of a world. But not only does the Humean accept HS_{chance}, also she denies the following doctrine:

> $\mathbf{HS^*_{chance}}$**:** For any worlds w and v and any time t, if w and v have the same past and present histories of chancy outcomes (up to and including t), then w and v have the same chance function at t.

That is, the Humean accepts that chances supervene on the whole of history of chancy outcomes and denies that they supervene on the present and past history alone. Now, the conjunction of HS_{chance} with the denial of HS^{*}_{chance} leads to the following peculiar conclusion:

> **Underminability:** For any world w and any time t, it is possible that w will have a future history of chancy outcomes that (i) is compatible with w's history up to and including t, (ii) has positive chance at t to come out true at w, and (iii) is incompatible with the chance function at t.

That is, the Humean metaphysical credo leads to the possibility of so-called *undermining futures*, i.e. certain future histories of chancy outcomes such that if they were to come out true, they would determine, together with the past and present history, a chance distribution that differs from the present one. We say then that such a future undermines the present chance distribution. Lewis gave the following example of an undermining future:

> For instance, there is some minute present chance that far more tritium atoms will exist in the future than have existed hitherto, and each one of them will decay in only a few minutes. If this unlikely future came to pass, presumably it would complete a chancemaking pattern on which the half-life of tritium would be very much less than the actual 12.26 years. (Lewis 1994, p. 482)

To show how the conjunction of HS_{chance} with the denial of HS^{*}_{chance} leads to Underminability, suppose for reductio that Underminability is false. Then, the present chance function at w is compatible with any future history of chancy outcomes at w. This means in turn that the present chance function at w is fixed by the past and present history alone. But this is just to say that the present chance function at w supervenes on the past and present history alone. Hence, HS^{*}_{chance} is true, and this contradicts the Humean assumptions.

The foregoing is a brief summary of the metaphysical doctrine underpinning Humean accounts of chance. The question arises: how does this metaphysical doctrine relate to conditions C1 and C2? A simple way of framing a version of Underminability within the talk about an unrefined chance function and its possible refinements shows that both C1 and C2 cannot be satisfied. But in view of Proposition 3.1, this is a hardly acceptable conclusion.

To show this, let $\mathcal{G} = \{G_w : w \in \mathcal{W}\}$ be a finite set of mutually exclusive propositions describing possible global (past, present, and future) histories of chancy outcomes. According to HS_{chance}, each G_w is equivalent to the proposition C_{ch_w}, which says that the chances are given by ch_w. That is, for each $w \in \mathcal{W}$, $C_{ch_w} \equiv G_w$. Suppose that each ch_w is a possible refinement of some Ch. Now, consider such refinement at world v at the beginning of v's history, and suppose that it assigns positive chances to undermining global histories. That is, it assigns positive chances to global histories such that if they came about at v, they would fix a refinement different from ch_v. Thus, we have the chance function at v that assigns positive chances to undermining global histories and is fixed by one global history which will come about at v. If this is so, however, we have that, for every $A \in \mathcal{A}$, the expected chance of A at v, computed relative to the chance function ch_v, may differ from the chance of A at v. That is, for all $A \in \mathcal{A}$, it might be the case that

$$ch_v(A) \neq \sum_{w \in \mathcal{W}} ch_v(C_{ch_w}) ch_w(A). \tag{3.10}$$

This shows that it might be the case that ch_v does not expect to give the chances at v. A simple example brings home this point. A coin is tossed three times. There are 2^3 possible global histories, and hence, by HS_{chance}, there are 2^3 propositions about possible chance functions. Further, there are four possible chance functions that assign a chance to the proposition that the coin lands heads: (i) $ch_{w_0}(H) = 0$, (ii) $ch_{w_1}(H) = \frac{1}{3}$, (iii) $ch_{w_2}(H) = \frac{2}{3}$, and (iv) $ch_{w_3}(H) = 1$. Now, according to HS_{chance}:

- $C_{ch_{w_0}} \equiv TTT$
- $C_{ch_{w_1}} \equiv TTH \vee THT \vee HTT$
- $C_{ch_{w_2}} \equiv THH \vee HHT \vee HTH$
- $C_{ch_{w_3}} \equiv HHH$

Consider, for example, the chance of heads given by ch_{w_1}. Does this chance function expect itself to give the chance of heads? No, it does not since:

$$\begin{aligned}
\frac{1}{3} = ch_{w_1}(H) &= \sum_{i=0}^{3} ch_{w_1}(C_{ch_{w_i}})ch_{w_i}(H) \\
&= ch_{w_1}(TTT)ch_{w_0}(H) + ch_{w_1}(TTH \vee THT \vee HTT)ch_{w_1}(H) + \\
&\quad ch_{w_1}(THH \vee HHT \vee HTH)ch_{w_2}(H) + ch_{w_1}(HHH)ch_{w_3}(H) \\
&= \frac{31}{81} \\
&\neq \frac{1}{3}.
\end{aligned}$$

Now, (3.10) when conjoined with conditions C1 and C2 leads to the following problem. Let Ch be some unrefined chance function. Suppose that Ch satisfies C1. In particular, suppose that for all $A \in \mathcal{A}$,

$$Ch(A|C_{ch_v}) = ch_v(A). \tag{3.11}$$

Then, by (3.10):

$$ch_v \neq \sum_{w \in \mathcal{W}} ch_v(C_{ch_w})ch_w(A). \tag{3.12}$$

But (3.12) is a violation of C2, for C2 requires that

$$ch_v = \sum_{w \in \mathcal{W}} ch_v(C_{ch_w})ch_w(A). \tag{3.13}$$

So if Ch satisfies C1, then it does not satisfy C2. But this result is hardly acceptable given Proposition 3.1, which says that any unrefined chance function that satisfies C1 also satisfies C2. This suggests that there is a tension between Humean accounts of chance and conditions C1 and C2.

Where does this observation leave us? It prompts the following question: should we still believe that the conditions are plausible requirements for chance, let alone constitutive to it? By appealing to resiliency considerations, I suggest that it is Humean accounts of chance that should be treated with suspicion, not the two conditions. A defender of Humean accounts of chance might react by claiming that C1 and C2 are not plausible conditions on chance. However, our resiliency considerations provide reason to doubt whether this reaction is correct. For what we have shown in the preceding sections is that there is an important sense in which any account of chance that collides with the two conditions is wrong. That is, if it allows an unrefined chance function to violate these conditions, then this chance function is inadmissible with respect to resiliency.

Let me now turn to another worry surrounding Humean accounts of chance, known as the Big Bad Bug (hereafter, the Bug). As presented originally in Lewis (1994), the Bug is a reductio which aims to show that Humean accounts of chance contradict the Principal Principle. The Bug is essentially an *epistemological* argument: it appeals to the Principal Principle which tells us how an agent's credences in propositions concerning chances should be related to her credences in other propositions. But, as shown in Bigelow et al. (1993) and in Briggs (2009b), a version of the Bug can be presented without making this epistemological detour. More specifically, the Bug can be formulated as a genuine *metaphysical* argument (hereafter, the metaphysical Bug). The key is to replace the Principal Principle with a non-epistemic condition C1. The metaphysical argument shows that the application of C1 in the context of Humean accounts of chance leads to inconsistency. Again, let $\mathcal{G} = \{G_w : w \in \mathcal{W}\}$ be a finite set of mutually exclusive propositions describing possible global (past, present, and future) histories of chancy outcomes. Consider a chance function ch_v at v that assigns a positive chance to every member of $\mathcal{G}$. In particular, it assigns a positive chance to some undermining global history G_w at v, i.e. $ch_v(G_w) > 0$. Hence, by HS_{chance}, $ch_v(C_{ch_w}) > 0$. Now, by C1:

$$Ch(C_{ch_w}|C_{ch_v}) = ch_v(C_{ch_w}) > 0. \tag{3.14}$$

But since C_{ch_w} and C_{ch_v} are mutually exclusive, by probability theory:

$$Ch(C_{ch_w} \wedge C_{ch_v}) = 0. \tag{3.15}$$

Thus,

$$Ch(C_{ch_w}|C_{ch_v}) = 0. \tag{3.16}$$

So by (3.14) and (3.16) we arrive at a contradiction. This argument thus shows that C1 is incompatible with Ch's possible refinements that allow for undermining global histories.

But it might be objected that the metaphysical Bug is not a problem for the Humean. For it rests on C1 and, as we have just shown, this condition appears to be in a serious tension with Humean accounts of chance. So unless there is a good reason to accept C1, the Humean might not be convinced by the metaphysical Bug. But the core idea of this chapter is that such reason in fact could be given. If one cares about the resiliency of chance, then C1 should be a requirement for chance. Therefore, both the Humean and non-Humean should accept C1. If this is so, then the metaphysical Bug remains a serious challenge for the Humean.

My final point in this section concerns the question of whether there might be Humean accounts of chance that avoid the two problems presented above. It appears that not all Humean accounts of chance lead to Underminability. Schaffer (2003) has offered an account of chance that satisfies HS_{chance} and avoids Underminability. According to Schaffer, for any world w and $A \in \mathcal{A}$, the correct chance of A at w, denoted by $ch^*_w(A)$, is the chance of A conditional on the proposition that the chances at w are given by ch_w, i.e. $ch^*_w(A) = ch_w(A|C_{ch_w})$. Since each chance function ch_w satisfies HS_{chance}, and ch^*_w is defined from ch_w, ch^*_w satisfies HS_{chance} as well. To see how this theory works, recall our example given above. According to Schaffer's view, the possible four chances of the coin landing heads are:

- $ch^*_{w_0}(H) = ch_{w_0}\big(H|C_{ch_{w_0}}\big) = ch_{w_0}(H|TTT)$
- $ch^*_{w_1}(H) = ch_{w_1}\big(H|C_{ch_{w_1}}\big) = ch_{w_0}(H|TTH \vee THT \vee HTT)$
- $ch^*_{w_2}(H) = ch_{w_2}\big(H|C_{ch_{w_2}}\big) = ch_{w_2}(H|THH \vee HHT \vee HTH)$
- $ch^*_{w_3}(H) = ch_{w_3}\big(H|C_{ch_{w_3}}\big) = ch_{w_3}(H|HHH)$

As it is easy to observe, if ch_{w_i} is a chance function at a world in which the coin lands heads i times, then so is $ch^*_{w_i}$. Thus, ch^*_w satisfies HS_{chance}. To show that Schaffer's Humean account of chance avoids Underminability, suppose that $ch_v(C_{ch_w}) > 0$, where C_{ch_w} is equivalent to some undermining global history at v. Since C_{ch_v} and C_{ch_w} are incompatible, it follows that $ch_v(C_{ch_w}|C_{ch_v}) = 0$.

But this is just to say that $ch_v^*(C_{ch_w}) = 0$, and so ch^* does not assign a positive chance to the undermining global history at v.

Because Schaffer's Humean account of chance avoids Underminability, it accommodates perfectly C1, and hence is immune to the metaphysical Bug. To show the latter, notice that if $ch_v^*(C_{ch_w}) = 0$, then it is also true that $ch_v^*\big(C_{ch_w^*}\big) = 0$. So by C1:

$$Ch\big(C_{ch_w^*}|C_{ch_v^*}\big) = ch_v^*\big(C_{ch_w^*}\big) = 0, \tag{3.17}$$

which just says that $C_{ch_w^*}$ and $C_{ch_v^*}$ are incompatible relative to Ch.

3.6 Conclusions

I have shown that conditions C1 and C2 are an essential feature of resilient chance functions. They follow from the injunction to maximize the resiliency of chance. Based on the norm of Maximizing Resiliency suitably understood, it has been shown that an unrefined chance function that violates each of these conditions can be resiliency-dominated by an unrefined chance function that satisfies these conditions. Thus, as I hope to have shown, conditions C1 and C2 can be motivated on grounds of resiliency. Also, I have shown how the resiliency-based motivation for these conditions could contribute to the debate over some worries surrounding Humean accounts of chance. I have thus identified some interesting consequences of the resiliency-based approach to chance.

Our result may well be read as follows. Even if one is not convinced that conditions C1 and C2 are constitutive to chance, but cares about the resiliency of chance, one should regard the two conditions as a basic requirement for chances. It has to be emphasized that it was not my goal to show that conditions C1 and C2 are requirements for chances under all circumstances. Rather, the strategy offered in this chapter was to provide a framework and within it the assumptions under which these conditions can be supported.

Chapter 4

Kinematics of Chance: Conditionalization and Resiliency

The question of how an agent's credences or subjective probabilities change through time is one of the central topics discussed in Bayesian epistemology. It is widely believed that the rule of conditionalization or Bayes's rule captures the salient features of the way epistemic agents should update their credences through time upon receipt of new evidence. But the question of how chances or physical probabilities change through time appears to be a less explored issue in the philosophy of chance. An interesting approach to this problem was proposed by David Lewis (1986). He located the problem, and an answer to it, within the context of chance-credence coordination. More specifically, he argued that a particular kinematical model for chances follows from his Principal Principle—a principle prescribing a particular way of coordinating credences with chances. The model that follows from the Principal Principle is a form of Bayesian conditionalization. According to this model, any later chance function is equal to an earlier chance function conditional on the intervening history of categorical-property instantiations in between.

This chapter discusses in a systematic way Lewis's kinematical model for chances. First, it investigates the conditions that any kinematical model needs to satisfy to count as Lewis's kinematics of chance. Second, it presents and discusses Lewis's argument for his kinematics of chance: it shows how this kinematical model for chance follows from the Principal Principle. Third, the chapter presents an alternative argument for Lewis's kinematics of chance that does not appeal to the Principal Principle. Instead, the argument appeals to a principle that relates chance functions at different times. This principle in turn is motivated by resiliency considerations similar to those presented in chapter 3.

4.1 Introduction

If there are chances attributable to the world, how can we account for the way they develop over time? Consider David Lewis's (1986, p. 91) case of a labyrinth that you enter at 11:00 a.m. and you choose your turn at any branch point you reach by tossing a coin. At the starting point you have a 42% chance of reaching the center by noon. But suppose that at 11:30 you have turned into a region of the labyrinth from which it is hard to reach the center, and so your chance of reaching it has changed to 26%. We may think that the change in your chance of reaching the center was prompted by a change in the chance set-up (that is, you in the labyrinth) between 11:00 and 11:30. After all, between these two times you moved to a different region from which the center is hardly reachable, and so your chance of reaching it decreased.

Lewis famously claimed that the kinematics of chance, or the way in which chances develop over time, is fully captured by the rule of conditionalization, also known as Bayes's rule. Call Lewis's claim *Bayesian Kinematics of Chance.* More precisely, Lewis (1986, p. 101) claimed that "a later chance distribution comes from an earlier one by conditionalizing on the complete history of the interval in between". Bayesian Kinematics of Chance thus tells us how chances at one time relate to chances at a different time. Lewis argued for this claim by appealing to his Principal Principle—a particular chance-credence principle that relates an agent's credences in propositions concerning chances to her credences in other propositions.

Lewis's argument for Bayesian Kinematics of Chance is based on two assumptions. The first assumption is a formulation of the Principal Principle, according to which for any time t, a chance distribution at t is one's credence (subjective probability) distribution conditional on the complete theory of chance and the complete history up to and including t. The second assumption is that the complete history up to t', $t' > t$, can be decomposed into the conjunction of the complete histories of subsequent intervals, i.e. into the complete history up to t and the complete history of the interval between t and t'. As Lewis showed, from these two assumptions it follows that the chance distribution at t' equals the preceding chance distribution at t conditional on the complete history of the interval between t and t'.

Interestingly, the idea that the kinematics of chance is governed by conditionalization parallels the well-known view in Bayesian epistemology, which says that the kinematics of an agent's credence is governed by conditionalization. Lewis wrote:

> The evolution of chance is parallel to the evolution of credence for an agent who learns from experience, as he reasonably might, by conditionalizing. In that case a later credence function comes from an earlier one by conditionalizing on the total increment of evidence gained in the interval in between. For the evolution of chance we simply put the world's chance distribution in place of the agent's credence function, and the totality of particular fact about a time in place of the totality of evidence gained at that time. (Lewis 1986, p. 101)

Typically, however, the idea that updating on evidence is governed by the rule of conditionalization is not taken for granted in Bayesian epistemology. Various arguments have been employed to show that, as a way of updating the agent's credences, the rule of conditionalization can be justified from more fundamental rationality requirements for credences. Some authors (Teller 1973; Skyrms 1987a; Lewis 1999) have argued that the rule of conditionalization as a way of updating one's degrees of belief follows from the injunction to avoid sure loss. Here so-called diachronic Dutch book arguments have been devised to show that any agent who violates conditionalization is vulnerable to a set of bets which ensure that she suffers a net loss. Other authors (Brown 1976; Maher 1992) have shown that any agent should update by conditionalization in order to maximize expected utility of her acts with respect to her prior credence function. Similarly, some authors (Greaves and Wallace 2006; Leitgeb and Pettigrew 2010b; Easwaran 2013) have shown that any agent should update her credences by conditionalization in order to maximize expected epistemic utility of her credences with respect to her prior credence function. Also, it has been argued that the rule of conditionalization follows from certain symmetry requirements for degrees-of-belief updating (van Fraassen 1989, pp. 331-337) and from Reflection principles that relate an agent's present and later credences (van Fraassen 1999).

It thus seems natural to ask: could Bayesian Kinematics of Chance be justified from more fundamental requirements for chance? If we follow Lewis in taking the Principal Principle as such requirement, the answer is: Yes, it could. But, as many authors pointed out, the Principal Principle cannot be the sole requirement for chances: there are other equally fundamental requirements for chances that do not follow from the Principal Principle (Arntzenius and Hall 2003; Schaffer 2003, 2007). If so, could we show that Bayesian Kinematics of Chance follows from these other requirements for chances? Moreover, the Principal Principle provides only an *epistemic* justification for Bayesian Kinematics of Chance in the sense that we derive this kinematics of chance from a principle that says how one's epistemic state described by a credence function should be constrained by evidence of chances. But if chances are to be objective features in the world, it might be claimed that any justification of the way by which they evolve through time had better be *objective*, i.e. appealing to objective, agent-independent features of chances. After all, one might reasonably wonder what the evolution of chances over time has to do with one's credences. It is thus tempting to ask: could we justify Bayesian Kinematics of Chance by appealing to the requirements for chances that do not make an epistemological detour to the Principal Principle?

My aim in this chapter is to show how Bayesian Kinematics of Chance can be justified without appealing to the Principal Principle. I shall argue that, under certain fairly plausible conditions, Bayesian Kinematics of Chance is equivalent to a particular principle connecting prior and possible posterior chances. I call this principle *Generalized Chance Expectation*, for it requires chances at a particular time to be equal to a certain expectation of possible chances at some later time.

As will be apparent, Generalized Chance Expectation has nothing to do with the way chances regulate one's credences, and the support it lends for Bayesian Kinematics of Chance is grounded only on a relation between chances at two different times.

A closely related principle to Generalized Chance Expectation, together with other plausible requirements for chances, has been introduced by John Bigelow, John Collins, and Robert Pargetter (1993) as constitutive to our notion of chance. They have argued that chances violating this principle are not only peculiar, but in fact they are not chances at all. In chapter 3, I have shown that a formulation of this principle can be supported by appealing to the idea of chance's resiliency. Similarly, in this chapter I will also show how Generalized Chance Expectation can be supported by resiliency considerations. That is, I will show that there is a certain sense in which this principle is an essential feature of resilient or stable chance distributions, under suitable conditions.

The structure of this chapter is as follows. In section 4.2, I give a precise statement of Bayesian Kinematics of Chance, and examine the conditions under which a transition from one chance function to another counts as Bayesian Kinematics of Chance. In section 4.3, I present a detailed reconstruction of Lewis's argument for Bayesian Kinematics of Chance, and show how it can be extended to justify separately two conditions whose conjunction is equivalent to Bayesian Kinematics of Chance. In section 4.4, I give some motivation for why we might want to give a justification of Bayesian Kinematics of Chance without appealing to the Principal Principle. In section 4.5, I present an argument for Bayesian Kinematics of Chances that rests solely on Generalized Chance Expectation. In section 4.6, I will bolster the argument given in section 4.5 by showing how resiliency considerations support Generalized Chance Expectation. Finally in section 4.7, I discuss the scope of Bayesian Kinematics of Chance.

4.2 Bayesian Kinematics of Chance: A Precise Statement

In this section, I present Bayesian Kinematics of Chance more precisely. In particular, I show how it can be located in a broader context concerning a transition from one chance function to another, and what conditions need to be satisfied for Bayesian Kinematics of Chance to work. Before doing so, however, let me introduce some terminology and notation.

Suppose that $\mathcal{W}$ is a set of possible outcomes of a trial conducted on some chance set-up and that $\mathcal{F}$ is a full and finite algebra of propositions generated by $\mathcal{W}$. We might think of the elements of $\mathcal{W}$ as possible worlds and the elements of $\mathcal{F}$ as propositions describing those worlds. Let ch_t stand for a chance function over $\mathcal{F}$ at time t. It is important to note that Lewis thought of chance functions as relative not only to time, but also to possible worlds. For simplicity, I assume that ch_t holds at some particular world, and so I will not use the world parameter explicitly. Further, for any times t' and t, $t' > t$, let $\mathcal{I}^t_{t'}$ be a finite partition of $\mathcal{W}$. Each proposition $I^t_{t'}$ in $\mathcal{I}^t_{t'}$ describes a complete history of changes that a chance set-up may undergo in the interval between t and t'. According to Lewis, this complete intervening history is given by the complete intervening history of categorical-property instantiations. These categorical properties do not involve chances, for chances according to Lewis supervene on the global distribution of these properties' instantiations. More precisely, Lewis (1983; 1994) claimed that categorical properties are qualitative, perfectly natural (i.e. non-gerrymandered, e.g. unlike the property of being an emerald), involve nothing modal like propensities, chances or counterfactuals, and are capable of being possessed by spacetime points or occupants thereof. For the purpose of this section, I assume Lewis's characterization of a complete history of the interval. This, however, is not to say that Bayesian Kinematics of Chance works only if this characterization is assumed. In section 4.7, I will argue that a complete history of the interval can be understood broadly as the history of changes in a chance set-up.

Now, given the set $\mathcal{I}^t_{t'}$ and the set $\mathcal{C}$ of chance functions over $\mathcal{F}$, we can introduce a *kinematics rule* for chances. This rule manages to model a transition from one chance function to another chance function. Formally, the kinematics rule for chances is a function $K : \mathcal{C} \times \mathcal{I}_{t'}^{t} \to \mathcal{C}$ that, given the chance function $ch_t(\cdot)$ and a complete intervening history $I_{t'}^{t} \in \mathcal{I}_{t'}^{t}$, produces another chance function $K\big(ch_t, I_{t'}^{t}\big)\big(\,\cdot\,\big)$. That is, if $I_{t'}^{t}$ is a complete intervening history in the interval between t and t', then the kinematics rule induces a shift or transition from the chance function $ch_t(\cdot)$ at t to the revised chance function $K\big(ch_t, I_{t'}^{t}\big)\big(\,\cdot\,\big)$ at t'. Thus, the kinematics rule tells us how to revise $ch_t(\cdot)$ in order to accommodate the truth of $I_{t'}^{t}$.

With these assumptions in place, Bayesian Kinematics of Chance may be presented more precisely as follows:

> **Bayesian Kinematics of Chance:** If $I_{t'}^{t} \in \mathcal{I}_{t}^{t}$ is the complete history of the interval between t and t', then for all $A \in \mathcal{F}$,
>
> $$K\big(ch_t, I_{t'}^{t}\big)\big(A\big) = ch_t\big(A|I_{t'}^{t}\big) := \frac{ch_t\big(A \wedge I_{t'}^{t}\big)}{ch_t\big(I_{t'}^{t}\big)},$$
>
> providing $ch_t\big(I_{t'}^{t}\big) > 0$.

In words, the chance function at t' is equal to the chance function at t conditional on the complete history of the interval in between.

To illustrate how this kinematical model for chances works, consider the following example. Suppose that there are two urns, each one containing 100 marbles. One urn contains 40 red and 60 white marbles, and the other urn contains 60 red and 40 white marbles. There is also a device that first selects with 0.5 chance one of these urns, and then selects a marble at random from it.

As Figure 4.1 shows, there are two possible changes in the marble-selecting device's categorical properties between t and t', each one determining a different chance at t' that a red (blue) marble will be selected. The device has different categorical properties at t' depending on which urn it has selected at t'. Each possible change in the device's state has a 0.5 chance at t to come out true at t'. If Urn 1 is selected, the chance of Urn 1 being selected raises from 0.5 to 1, while the chance of Urn 2 being selected lowers to 0. Suppose that the top path in Figure 4.1 represents the actual history from t to t''. Then, if there is a determinate chance at t that a red (blue) marble will be selected, Bayesian Kinematics of Chance takes the chance at t' of a red (blue) marble being selected

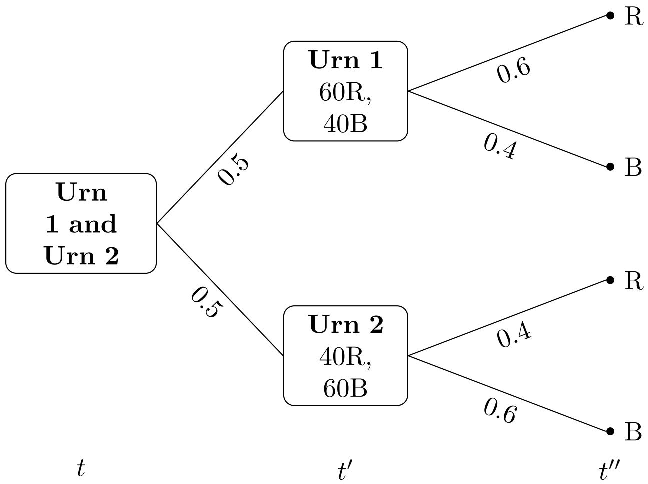

Figure 4.1: The chances at t' of a red marble (R) and a blue marble (B) being selected at t'' are determined by the chances at t and the change of the device's categorical properties between t and t'.

to be the conditional chance at t given the selection of Urn 1 at t'.

It can be shown that Bayesian Kinematics of Chance is equivalent to two conditions that I call *Chance Certainty* and *Chance Rigidity*. These two conditions can be thought as chance-counterparts of two analogous conditions that are typically imposed on an agent's credence function in the case of belief updating (e.g. Jeffrey 1988). Suppose that a shift from $ch_t(\cdot)$ to $K\big(ch_t, I^t_{t'}\big)\big(\cdot\big)$ satisfies the following two conditions:

Chance Certainty: For all $I^t_{t'} \in \mathcal{F}$,

$$K\big(ch_t, I^t_{t'}\big)\big(I^t_{t'}\big) = 1.$$

Chance Rigidity: For all A and $I^t_{t'} \in \mathcal{F}$,

$$K\big(ch_t, I^t_{t'}\big)\big(A|I^t_{t'}\big) = ch_t\big(A|I^t_{t'}\big).$$

Then, the following proposition holds true:

Proposition 4.1. *Chance Certainty and Chance Rigidity ⇔ Bayesian Kinematics of Chance.*

Proof. (⇒) Suppose that Chance Certainty and Chance Rigidity hold true. Then,

$$\begin{aligned} K\big(ch_t, I^t_{t'}\big)\big(A\big) &= K\big(ch_t, I^t_{t'}\big)\big(A|I^t_{t'}\big) \text{ (by Chance Certainty)} \\ &= ch_t\big(A|I^t_{t'}\big) \text{ (by Chance Rigidity),} \end{aligned}$$

as required.

(⇐) Suppose that Bayesian Kinematics of Chance holds true. Then,

$$\begin{aligned} K\big(ch_t, I^t_{t'}\big)\big(I^t_{t'}\big) &= ch_t\big(I^t_{t'}|I^t_{t'}\big) \\ &= 1, \end{aligned}$$

which yields Chance Certainty, and

$$\begin{aligned} K\big(ch_t, I^t_{t'}\big)\big(A|I^t_{t'}\big) &= \frac{K\big(ch_t, I^t_{t'}\big)\big(A \wedge I^t_{t'}\big)}{K\big(ch_t, I^t_{t'}\big)\big(I^t_{t'}\big)} \\ &= \frac{ch_t\big(A \wedge I^t_{t'}|I^t_{t'}\big)}{ch_t\big(I^t_{t'}|I^t_{t'}\big)} \\ &= ch_t\big(A \wedge I^t_{t'}|I^t_{t'}\big) \\ &= \frac{ch_t\big(A \wedge I^t_{t'} \wedge I^t_{t'}\big)}{ch_t\big(I^t_{t'}\big)} \\ &= ch_t\big(A|I^t_{t'}\big), \end{aligned}$$

which gives Chance Rigidity. □

Chance Certainty appears to be a highly intuitive condition. If a particular $I^t_{t'}$ is true, then the chance at t' of $I^t_{t'}$'s truth should be 1. After all, at t' the proposition $I^t_{t'}$ describes the past intervening history, and as Lewis claims: "what's past is no longer chancy" (Lewis 1986, p. 93). At any time up to t' the chance of $I^t_{t'}$'s truth fluctuates, but is fixed at 1 once $I^t_{t'}$ is true. The intuitive force of Chance Certainty notwithstanding, in section 4.3 I will show how this condition can be supported additionally by the Principal Principle.

But is Chance Rigidity an equally intuitive condition for the evolution of chances? There seems to be no immediate rationale for imposing Chance Rigidity on the kinematical rule for chances. Why should prior and posterior chances conditional on $I^t_{t'}$ stay unchanged? Typically, an analogous condition in the case of an agent's credences is defended by means of the following intuitive argument:

> Suppose that A is all and everything that is learned, that all changes to the agent's partial belief are rational effects of her learning that A, but that her new degrees of belief are not her old degrees of belief conditional on A. Then her conditional degrees of belief given A must have changed. But the truth of A is not itself a reason to change one's conditional beliefs given A, so something more than A must have been learned. But that is contrary to the supposition that A is all that is learned. (Bradley 2005, p. 345)

Adapting this line of thought in the case of chances, we could argue that if the chances at t and t' conditional on $I^t_{t'}$ are different, then $I^t_{t'}$ does not describe the whole change that a chance set-up undergoes between t and t', and thus something more must have occurred in between. But also it is well known that the Rigidity condition in the case of an agent's credences might fail to be satisfied: though A is all and everything that is learned, it could, together with the agent's other credences, prompt a change in her conditional degrees of belief given A. Likewise, consider the example visualized in Figure 4.1: it might well be that the change of the marble-selecting device's categorical properties at t' prompts a change in the device's causal mechanism which in turn fixes different chances conditional on $I^t_{t'}$. Hence, in such circumstances, the proposition $I^t_{t'}$ would not describe all the changes that the device undergoes. These observations notwithstanding, in section 4.3 I will show how Chance Rigidity could follow from the Principal Principle, and in section 4.5 how it could be supported by Generalized Chance Expectation.

We can illustrate the idea behind these two conditions by means of van Fraassen's (1989, pp. 161-162) muddy Venn diagrams. Suppose that propositions from $\mathcal{W}$ are regions inside a box representing $\mathcal{W}$, as presented in Figure 4.2. Imagine that we heap one unit of mud on the box, and the amount of mud

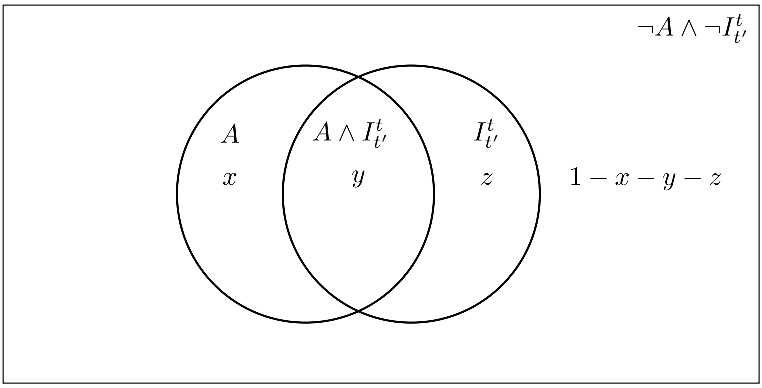

Figure 4.2: A distribution of the mud over regions representing propositions A, $A \wedge I^t_{t'}$, $I^t_{t'}$, and $\neg A \wedge \neg I^t_{t'}$ from $\mathcal{W}$.

in each region is the chance of the proposition represented by that region. Now, as presented in Figure 4.3, when a particular $I^t_{t'}$ is true, we wipe away all the mud that lies outside the region $I^t_{t'}$ and leave the rest of the mud where it is in. This results, first, in removing all of the mud that was not already in $I^t_{t'}$'s region into it. This in turn increases the amount of mud in that region to unity, and so $K\big(ch_t, I^t_{t'}\big)\big(I^t_{t'}\big) = 1$. Hence, Chance Certainty holds true. Second, the proportion of $I^t_{t'}$'s mud that is also in A is kept unchanged: the swept-away mud is redistributed in $I^t_{t'}$ in such a way that the proportions of the mud in regions within $I^t_{t'}$ are retained.

As it is easy to observe from Figure 4.2 and 4.3, before redistribution the ratio of the mud in $A \wedge I^t_{t'}$ to the mud in $I^t_{t'}$ is $\frac{y}{y+z}$, for the amount of mud in $I^t_{t'}$ is $y + z$ and the amount of mud in $A \wedge I^t_{t'}$ is y. After redistribution the total amount of mud is in $I^t_{t'}$ and the amount of mud in $A \wedge I^t_{t'}$ is $\frac{y}{y+z}$. Hence, the ratio of the mud in $A \wedge I^t_{t'}$ to the mud in $I^t_{t'}$ is $\frac{y}{y+z}$, and so it is left unchanged after redistribution. And if we write $ch_t\big(A|I^t_{t'}\big)$ for the proportion of $I^t_{t'}$'s mud that is also in A before redistribution and $K\big(ch_t, I^t_{t'}\big)\big(A|I^t_{t'}\big)$ for the proportion of $I^t_{t'}$'s mud that is also in A after redistribution, we get Chance Rigidity, i.e. $K\big(ch, I^t_{t'}\big)\big(A|I^t_{t'}\big) = ch_t\big(A|I^t_{t'}\big)$.

It is worth noticing that, in view of Proposition 4.1, any argument supporting Bayesian Kinematics of Chance also supports Chance Certainty and Chance

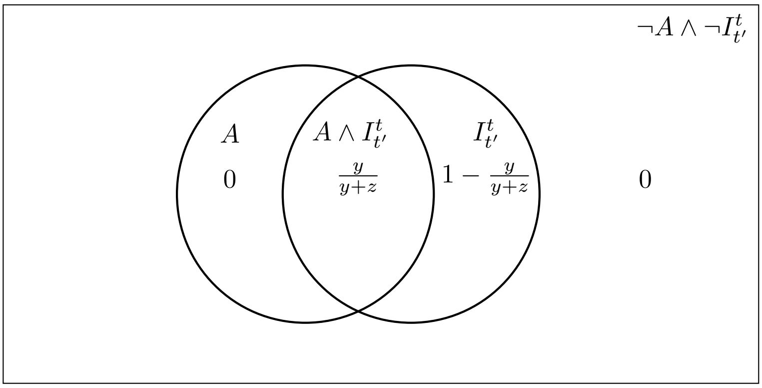

Figure 4.3: A redistribution of the mud over regions representing propositions A, $A \wedge I^t_{t'}$, $I^t_{t'}$, and $\neg A \wedge \neg I^t_{t'}$ from $\mathcal{W}$, given the truth of $I^t_{t'}$.

Rigidity, and vice versa. In section 4.3, I introduce Lewis's argument for Bayesian Kinematics of Chance. Although it justifies the two conditions automatically, I will also use the main idea of his argument to provide an independent Lewis-style justification for each of the two conditions.

4.3 Bayesian Kinematics of Chance: Lewis's Argument

Lewis derives Bayesian Kinematics of Chance from his Principal Principle. In fact, for this derivation he uses a particular reformulation of that principle. In this section, I first show that the Principal Principle and this particular reformulation are equivalent principles. Second, I state precisely Lewis's argument for Bayesian Kinematics of Chance by using this reformulation of the Principal Principle. Finally, I use the basic idea behind Lewis's argument to justify Chance Certainty and Chance Rigidity.

Lewis's Principal Principle can be introduced as follows:

> **Principal Principle:** Suppose that cr is an agent's reasonable initial credence function. Let ch_t be the chance function over $\mathcal{F}$ at time t, and let $\ulcorner ch_t(A) = x \urcorner$ be the proposition that the chance of A at t is

x, where x is a real number in the closed unit interval. If E is any proposition that is admissible at t, then

$$cr(A|\ulcorner ch_t(A) = x\urcorner \wedge E) = x,$$

providing $cr(\ulcorner ch_t(A) = x\urcorner \wedge E) > 0$.

In words, if the agent started with a reasonable initial credence in A, received at time t the information that the chance of A at t equals x, and the rest of her evidence E were admissible at t, then she should have credence in A equal to x.

To provide Lewis's reformulation of the Principal Principle, we need to introduce his characterization of admissibility:

> Admissible propositions are the sort of information whose impact on credence about outcomes comes entirely by way of credence about the chances of those outcomes. (Lewis 1986, p. 92)

Lewis gave two examples of propositions that satisfy his account of admissible information: information about the complete history of categorical-property instantiations up to and including time t, and information about the laws of nature, including probabilistic laws. Moreover, Lewis assumed that any Boolean combination of the two is also admissible at time t.

With these notions in mind, we are in a position to introduce what Lewis calls the Principal Principle Reformulated. It can be presented as follows:

The Principal Principle Reformulated: Suppose that cr is an agent's reasonable initial credence function, and let t be any time. Let H_t be a proposition about the complete history of categorical-property instantiations up to and including time t and let L be a proposition about the laws of nature. Then, for all $A \in \mathcal{F}$,

$$ch_t(A) = cr(A|H_t \wedge L),$$

providing $cr(H_t \wedge L) > 0$.

That is, this principle tells us that the chance of A at time t is equal to the conditional reasonable initial credence in A given the complete history up to and including t and the laws of nature. To show precisely that the Principal

Principle and the Reformulated Principal Principle are equivalent, let us prove the following proposition:

Proposition 4.2. *The Principal Principle* $\Leftrightarrow$ *The Principal Principle Reformulated.*

Proof. ($\Leftarrow$) Following Lewis (1986, p. 99), assume that $\ulcorner ch_t(A) = x \urcorner \leftrightarrow H_t^1 \wedge L^1 \vee ... \vee H_t^n \wedge L^n$ for all $H_t^i \wedge L^i$ such that $ch_t(A) = x$. Assume further that E is admissible and compatible with $\ulcorner ch_t(A) = x \urcorner$. Now, since every $H_t^i \wedge L^i$ is admissible and compatible with $\ulcorner ch_t(A) = x \urcorner$, we can replace E with $H_t^1 \wedge L^1 \vee ... \vee H_t^n \wedge L^n$. Then,

$$
\begin{aligned}
cr(A|E \wedge \ulcorner ch_t(A) = x \urcorner) &= cr(A|H_t^1 \wedge L^1 \vee ... \vee H_t^n \wedge L^n \wedge \ulcorner ch_t(A) = x \urcorner) \\
&= cr(A|H_t^1 \wedge L^1 \vee ... \vee H_t^n \wedge L^n) \\
&= \frac{cr(A \wedge (H_t^1 \wedge L^1 \vee ... \vee H_t^n \wedge L^n))}{cr(H_t^1 \wedge L^1 \vee ... \vee H_t^n \wedge L^n)} \\
&= \frac{\sum_i cr(A \wedge H_t^i \wedge L^i)}{\sum_i cr(H_t^i \wedge L^i)} \\
&= \frac{\sum_i cr(H_t^i \wedge L^i) cr(A|H_t^i \wedge L^i)}{\sum_i cr(H_t^i \wedge L^i)} \\
&= \frac{\sum_i cr(H_t^i \wedge L^i) ch_t(A)}{\sum_i cr(H_t^i \wedge L^i)} \\
&\text{(by the Reformulated Principal Principle)} \\
&= \frac{\sum_i cr(H_t^i \wedge L^i)}{\sum_i cr(H_t^i \wedge L^i)} x \text{ (by assumption)} \\
&= x,
\end{aligned}
$$

as required.

($\Rightarrow$) Suppose that $ch_t(A) = x$ and E is admissible and compatible with $\ulcorner ch_t(A) = x \urcorner$. Then, $H_t \wedge L \rightarrow \ulcorner ch_t(A) = x \urcorner$, and we have that

$$
\begin{aligned}
cr(A|H_t \wedge L) &= cr(A|\ulcorner ch_t(A) = x \urcorner \wedge H_t \wedge L) \\
&= x \\
&\text{(by the Principal Principle and the fact that } H_t \wedge L \text{ is admissible)}
\end{aligned}
$$

$$= ch_t(A) \text{ (by assumption)},$$

as required. □

The crucial step in the proof of Proposition 4.2 involves the fact that the proposition that the chance of A at t is x is equivalent to the disjunction of the $H_t^i \wedge L^i$'s, for each of which the chance of A at t is x. The proposition that the chance of A at t is x and the disjunction of the $H_t^i \wedge L^i$'s are equivalent in the sense that the former is true just in case some $H_t^i \wedge L^i$ picks out the value $ch_t(A) = x$.

With the Principal Principle Reformulated in hand, I will now present Lewis's argument for Bayesian Kinematics of Chance as a series of equations which I will then chain together.

First, since $K\big(ch_t, I_{t'}^t\big)\big(A\big)$ is the chance of A at t', it follows from the Principal Principle Reformulated that

$$K\big(ch_t, I_{t'}^t\big)\big(A\big) = cr(A|H_{t'} \wedge L). \tag{4.1}$$

Second, following Lewis (1986, p. 101), assume that any complete history of categorical-property instantiations up to and including time t' can be decomposed into the conjunction of complete histories of subsequent intervals. Formally, $H_{t'} = H_t \wedge I_{t'}^t$, where H_t is the complete history of categorical-property instantiations in the interval up to and including time t, and $I_{t'}^t$ is the complete history of categorical-property instantiations in the interval from t up to and including t'. Then,

$$cr(A|H_{t'} \wedge L) = cr(A|H_t \wedge I_{t'}^t \wedge L). \tag{4.2}$$

By the definition of conditional probability:

$$cr\big(A|H_t \wedge I_{t'}^t \wedge L\big) = \frac{cr\big(A \wedge H_t \wedge I_{t'}^t \wedge L\big)}{cr\big(H_t \wedge I_{t'}^t \wedge L\big)}. \tag{4.3}$$

By the product rule of probability theory:

$$\frac{cr\big(A \wedge H_t \wedge I_{t'}^t \wedge L\big)}{cr\big(H_t \wedge I_{t'}^t \wedge L\big)} = \frac{cr(H_t \wedge L)cr\big(A \wedge I_{t'}^t|H_t \wedge L\big)}{cr(H_t \wedge L)cr\big(I_{t'}^t|H_t \wedge L\big)}. \tag{4.4}$$

By the Reformulated Principal Principle again:

$$\begin{aligned}\frac{cr(H_t \wedge L)cr\big(A \wedge I^t_{t'}|H_t \wedge L\big)}{cr(H_t \wedge L)cr\big(I^t_{t'}|H_t \wedge L\big)} &= \frac{cr(H_t \wedge L)ch_t\big(A \wedge I^t_{t'}\big)}{cr(H_t \wedge L)ch_t\big(I^t_{t'}\big)} \\ &= \frac{ch_t\big(A \wedge I^t_{t'}\big)}{ch_t\big(I^t_{t'}\big)}.\end{aligned} \tag{4.5}$$

And by the definition of conditional probability again:

$$\frac{ch_t\big(A \wedge I^t_{t'}\big)}{ch_t\big(I^t_{t'}\big)} = ch_t\big(A|I^t_{t'}\big). \tag{4.6}$$

Chaining together equations (4.1)—(4.6), we get

$$K\big(ch_t, I^t_{t'}\big)\big(A\big) = ch_t\big(A|I^t_{t'}\big), \tag{4.7}$$

which yields Bayesian Kinematics of Chance. This argument thus shows that if chance satisfies the Reformulated Principal Principle, then it also satisfies Bayesian Kinematics of Chance.

The idea behind Lewis's argument is simple. If we define the chance function at t' as the agent's reasonable initial credence function conditional on the complete history up to t' and the laws of nature, and we extend that history by conjoining complete histories of subsequent intervals, then we get, by using the Principal Principle Reformulated, the conclusion that the chance function at t' is equal to the chance function at t conditional on complete history of the interval between t and t'. Thus, Bayesian Kinematics of Chance follows from a principle that tells us how chance should guide one's credence.

So far in this section, I have reconstructed Lewis's argument for Bayesian Kinematics of Chance. But the basic idea behind Lewis's argument can also be used to show how Chance Certainty and Chance Rigidity follow from the Reformulated Principal Principle. As shown in section 4.2, the two conditions taken together are equivalent to Bayesian Kinematics of Chance. Thus, given an argument for Bayesian Kinematics of Chance, we also get an argument for the two conditions taken together. However, one might want to provide support for each of these conditions separately. Interestingly, the Reformulated Principal

Principle provides such support.

Let me first provide an argument for Chance Certainty, by giving a series of equations and then chaining them together. Assume again that $H_{t'} = H_t \wedge I^t_{t'}$, and the Principal Principle Reformulated is the rule that connects credences with chances. Then, by the Principal Principle Reformulated:

$$K\big(ch_t, I^t_{t'}\big)\big(I^t_{t'}\big) = cr\big(I^t_{t'}|H_{t'} \wedge L\big). \tag{4.8}$$

By the fact that $H_{t'} = H_t \wedge I^t_{t'}$ and by the probability theory:

$$\begin{aligned} cr\big(I^t_{t'}|H_{t'} \wedge L\big) &= cr\big(I^t_{t'}|H_t \wedge I^t_{t'} \wedge L\big) \\ &= 1. \end{aligned} \tag{4.9}$$

Now, chaining together equations (4.8)—(4.9), we get

$$K\big(ch_t, I^t_{t'}\big)\big(I^t_{t'}\big) = 1, \tag{4.10}$$

as required. That is, it follows from the Principal Principle Reformulated that the chance at t' of the interval-history between t and t' must be 1. In other words, an application of this principle leads to the conclusion that any past intervening history of categorical-property instantiations is fixed, and so the chance of its occurring must be 1.

Using the same assumptions as above, we can give the following argument for Chance Rigidity. By the definition of conditional probability:

$$K\big(ch_t, I^t_{t'}\big)\big(A|I^t_{t'}\big) = \frac{K\big(ch_t, I^t_{t'}\big)\big(A \wedge I^t_{t'}\big)}{K\big(ch_t, I^t_{t'}\big)\big(I^t_{t'}\big)}. \tag{4.11}$$

By the Principal Principle Reformulated:

$$\frac{K\big(ch_t, I^t_{t'}\big)\big(A \wedge I^t_{t'}\big)}{K\big(ch_t, I^t_{t'}\big)\big(I^t_{t'}\big)} = \frac{cr\big(A \wedge I^t_{t'}|H_{t'} \wedge L\big)}{cr\big(I^t_{t'}|H_{t'} \wedge L\big)}. \tag{4.12}$$

By the fact that $H_{t'} = H_t \wedge I^t_{t'}$ and by the probability theory:

$$\frac{cr\big(A \wedge I^t_{t'} | H_{t'} \wedge L\big)}{cr\big(I^t_{t'} | H_{t'} \wedge L\big)} = \frac{cr\big(A \wedge I^t_{t'} | H_t \wedge I^t_{t'} \wedge L\big)}{cr\big(I^t_{t'} | H_t \wedge I^t_{t'} \wedge L\big)} = cr\big(A \wedge I^t_{t'} | H_t \wedge I^t_{t'} \wedge L\big). \tag{4.13}$$

By the definition of conditional probability:

$$cr\big(A \wedge I^t_{t'} | H_t \wedge I^t_{t'} \wedge L\big) = \frac{cr\big(A \wedge I^t_{t'} \wedge H_t \wedge I^t_{t'} \wedge L\big)}{cr\big(H_t \wedge I^t_{t'} \wedge L\big)}. \tag{4.14}$$

By the product rule of probability theory:

$$\frac{cr\big(A \wedge I^t_{t'} \wedge H_t \wedge L\big)}{cr\big(H_t \wedge I^t_{t'} \wedge L\big)} = \frac{cr\big(H_t \wedge L\big) cr\big(A \wedge I^t_{t'} | H_t \wedge L\big)}{cr\big(H_t \wedge L\big) cr\big(I^t_{t'} | H_t \wedge L\big)}. \tag{4.15}$$

And, by the Principal Principle Reformulated and probability theory:

$$\begin{aligned} \frac{cr\big(H_t \wedge L\big) cr\big(A \wedge I^t_{t'} | H_t \wedge L\big)}{cr\big(H_t \wedge L\big) cr\big(I^t_{t'} | H_t \wedge L\big)} &= \frac{cr\big(H_t \wedge L\big) ch_t\big(A \wedge I^t_{t'}\big)}{cr\big(H_t \wedge L\big) ch_t\big(I^t_{t'}\big)} \\ &= \frac{ch_t\big(A \wedge I^t_{t'}\big)}{ch_t\big(I^t_{t'}\big)} \\ &= ch_t\big(A | I^t_{t'}\big). \end{aligned} \tag{4.16}$$

Now, chaining together equations (4.11)—(4.16):

$$K\big(ch_t, I^t_{t'}\big)\big(A | I^t_{t'}\big) = ch_t\big(A | I^t_{t'}\big), \tag{4.17}$$

as required. That is, by using the Principal Principle Reformulated, we get the conclusion that the chance functions at t and t' conditional on $I^t_{t'}$ stay unchanged. Thus, Chance Rigidity receives an independent motivation by appealing to the chance-credence relation.

4.4 Do We Need a New Argument?

If we were to follow Lewis in regarding the Principal Principle, and hence the Principal Principle Reformulated, as the key requirement for chance, then it seems that there would be no way to question his argument for Bayesian Kinematics of Chance. After all, the argument derives this kinematical model from the Principal Principle. But it seems that Lewis's idea about the role of the Principal Principle in establishing Bayesian Kinematic of Chance is not so innocuous as one might think. In this section, I give two reasons for thinking this is so. By doing so, I also give motivation for seeking a new argument in favour of Bayesian Kinematics of Chance.

The first reason for doubting Lewis's argument is that it seems contentious to believe that the Principal Principle—a principle that relates chance to credence—could be used to establish the way by which chances should evolve through time. While it is uncontroversial to think that the Principal Principle captures an expert role of chance, to wit, its role in guiding one's belief, it is problematic to think that this principle dictates the way in which chances should evolve through time. For if chances are attributable to the mind-independent world, the way they evolve through time might be quite independent of the way they constrain one's credences.

The worry stated above stems from a more general observation, due to Bigelow, Collins, and Pargetter (1993). They claim, *pace* Lewis, that it is not so evident that the Principal Principle gives us the ultimate justification for attributing chances to the world in order to capture various physical phenomena. For if this were so, adhering to the Principal Principle would mean that

> [...] when physicists assign half-lives, they can only be justified in doing so if they have a justification for assigning specific rational degrees of belief. And yet that cannot be right! Physicists are not investigating rational degrees of belief. They are investigating physical phenomena. A physicist justifies attribution of a specific half-life by appealing to *physical grounds* alone. *No* appeal is made to suppositions about which degrees of belief are rational. The physical justification for attributing a half-life might be by a demonstration that the attribution of the half-life explains various phenomena. In particular, the attribution of this half-life might explain various observed frequencies. Justifications of this sort seem entirely adequate for the attribution of half-lives. They also seem to be justifications of a

> sort which it would be appropriate for physicists to provide, entirely within their role as physicists. Such a justification is adequate for the attribution of half-lives, or chances, properly so called. And yet justifications of the sort outlined make no appeal at all to rational degrees of belief. (Bigelow et al. 1993, p. 447)

The view advocated by Bigelow, Collins, and Pargetter is that in physics facts about chances are not necessarily derivable from facts about credences. This in turn makes dubious the claim that we could regulate the way chances evolve through time by appealing to facts about credences. Of course, this is not to say that the Principal Principle says nothing about chances. Rather, it is to say that it could not provide the defining role for chance. As Bigelow and his co-authors claim, we should think that once the chances are derived from our best scientific theories, our credences ought to be constrained by them. That is:

> The physicist attributes chances—properly so called—without paying any heed to rational degrees of belief. Then the nature of *rationality* is such that certain rational degrees of belief will be consequences of certain chances. And, hence, certain credences must match certain chances–and so the Principal Principle holds. (Bigelow et al. 1993, p. 448)

The second, and perhaps more important, reason for questioning Lewis's argument is that even if we regard the chance-credence relation as a tool for deriving certain facts about chances, we do not know the exact formulation that this relation should take. And this seems to be problematic because if we replace the Principal Principle with some other formulation of that relation, we may derive a kinematical model for chances that differs from Bayesian Kinematics of Chance. To illustrate this point, consider Ned Hall's (1994) and Michael Thau's (1994) New Principle:

> **The New Principle:** Suppose that cr is an agent's reasonable initial credence function, and let t be any time. Let H_t be a proposition about the complete history of categorical-property instantiations up to and including time t, and let L be a proposition about the laws of nature. Then, for all $A \in \mathcal{F}$,
>
> $$ch_t(A|L) = cr(A|H_t \wedge L),$$
>
> providing $cr(H_t \wedge L) > 0$.

That is, the New Principle tells us that the chance of A at time t conditional on the laws of nature is equal to the conditional reasonable initial credence in A given the complete history up to and including t and the laws of nature. Assuming the New Principle, one can ask: does this principle lead to Bayesian Kinematics of Chance? But the answer to this question is: No, it does not.

To see this, observe first that it follows from the New Principle that

$$K\big(ch_t, I^t_{t'}\big)\big(A|L\big) = cr(A|H_{t'} \wedge L), \tag{4.18}$$

where $K\big(ch_t, I^t_{t'}\big)(A|L)$ is the chance of A at t' conditional on L. Then, assuming that $H_{t'} = H_t \wedge I^t_{t'}$, we get

$$cr(A|H_{t'} \wedge L) = cr(A|H_t \wedge I^t_{t'} \wedge L). \tag{4.19}$$

By the definition of conditional probability:

$$cr(A|H_t \wedge I^t_{t'} \wedge L) = \frac{cr(A \wedge H_t \wedge I^t_{t'} \wedge L)}{cr(H_t \wedge I^t_{t'} \wedge L)}. \tag{4.20}$$

By the product rule of probability theory:

$$\frac{cr(A \wedge H_t \wedge I^t_{t'} \wedge L)}{cr(H_t \wedge I^t_{t'} \wedge L)} = \frac{cr(H_t \wedge L)cr\left(A \wedge I^t_{t'}|H_t \wedge L\right)}{cr(H_t \wedge L)cr\left(I^t_{t'}|H_t \wedge L\right)}. \tag{4.21}$$

By the New Principle again:

$$\begin{aligned}\frac{cr(H_t \wedge L)cr\left(A \wedge I^t_{t'}|H_t \wedge L\right)}{cr(H_t \wedge L)cr\left(I^t_{t'}|H_t \wedge L\right)} &= \frac{cr(H_t \wedge L)ch_t\left(A \wedge I^t_{t'}|L\right)}{cr(H_t \wedge L)ch_t\left(I^t_{t'}|L\right)} \\ &= \frac{ch_t(A \wedge I^t_{t'}|L)}{ch_t(I^t_{t'}|L)}.\end{aligned} \tag{4.22}$$

And by the definition of conditional probability again:

$$\frac{ch_t(A \wedge I^t_{t'}|L)}{ch_t(I^t_{t'}|L)} = ch_t(A|I^t_{t'} \wedge L). \tag{4.23}$$

Chaining together equations (4.18)—(4.23), we get

$$K\big(ch_t, I^t_{t'}\big)\big(A|L\big) = ch_t(A|I^t_{t'} \wedge L), \tag{4.24}$$

which yields a kinematical model for chances conditional on L. And this model agrees with Bayesian Kinematics of Chance only if, for any t, $ch_t(L) = 1$. Thus, by varying the formulation of the chance-credence relation, we may get different kinematical models for chances.

The foregoing suggests that there are two sorts of motivation for providing an argument in favour Bayesian Kinematics of Chance that would make no detour to the Principal Principle. First, if facts about chances cannot be derived from facts about credences, then presumably they could be derived from some facts about physicists' scientific practice, or from some features of well-confirmed scientific laws, as suggested by Bigelow, Collins, and Pargetter. Hence, it is natural to ask whether Bayesian Kinematics of Chance could be justified by appealing to these other facts. Second, in providing a justification of Bayesian Kinematics of Chance we would like to avoid the unpleasant consequence of Lewis's argument: that a change in the formulation of the chance-credence relation can lead to a different kinematical model for chances.

There is also a third kind of motivation that should be considered. Although we might think that the Principal Principle tells us a great deal about chances, and could be used to derive Bayesian Kinematics of Chance, there are other equally important principles that inform our understanding of chances. A view of this sort is advocated by Schaffer (2007), who argues extensively that the Principal Principle could not be all we know about chance. Recall that Schaffer identifies a set of platitudes about chance, one of which is the Principal Principle. And, though the Principal Principle is perhaps the most plausible and acceptable of these conditions, Schaffer claims that chance is what best satisfies all of these conditions. Of course, Schaffer does not intend to provide an exhaustive list of platitudes about chance: his point is that there is more to our understanding of chance than the platitude given by the Principal Principle.

Similarly, Arntzenius and Hall (2003) have argued that the Principal Principle cannot be the sole constraint on chance, since it fails to explain why chances must explain physical symmetries of time and space: for example, that if two coin tosses in different regions of spacetime are exactly alike, their outcomes have the same chances. They have shown that what appears to play the role defined by the Principal Principle cannot explain these physical symmetries. From this observation they have concluded that the Principal Principle cannot be the sole constraint for chance.

Now, if there are other equally important principles for chances, it is tempting to ask: could we provide an argument in favour of Bayesian Kinematics of Chance by appealing to such principles? Note that this sort of motivation does not stem from a critique of Lewis's argument. Someone who is motivated by this sort of considerations might well endorse Lewis's argument. Still, she might want to seek a different justification for Bayesian Kinematics of Chance.

Whether or not the problems discussed above make Lewis's argument unacceptable, I will present below an argument for Bayesian Kinematics of Chance that makes no detour to the Principal Principle, but instead appeals to a principle relating prior and posterior chances.

4.5 A New Argument for Bayesian Kinematics of Chance

In this section, I formulate a principle for chances called Generalized Chance Expectation. Under a slightly different name, a similar principle, together with a set of other fairly plausible conditions for chances, has been proposed by Bigelow, Collins, and Pargetter (1993). I then show that, under suitable conditions, Generalized Chance Expectation is equivalent to Bayesian Kinematics of Chance.

Let us first introduce a condition that relates chance functions at two different times. I call this condition Chance-Chance Principle, for it might be read as a way of coordinating different chance functions through time. A similar condition has been defended by Bigelow, Collins, and Pargetter (1993) under the name "Chance Conditional on Chance Formulation", and has been put forward to rule out conceptions of chance that allow chances to be counterfactually dependent on future courses of history. Our condition might be formulated as follows:

Chance-Chance Principle: Suppose that ch_t is the chance function over $\mathcal{F}$ at time t and $K\big(ch_t, I^t_{t'}\big)$ is the chance function over $\mathcal{F}$ at time t', $t' > t$. Let $\ulcorner K\big(ch_t, I^t_{t'}\big) = ch_{t'}\urcorner$ be the proposition that the chance function over $\mathcal{F}$ at t' is $ch_{t'}$.[1] Then, for all $A \in \mathcal{F}$,

$$ch_t\big(A\big|\ulcorner K\big(ch_t, I^t_{t'}\big) = ch_{t'}\urcorner\big) = ch_{t'}(A),$$

if $ch_t\big(\ulcorner K\big(ch_t, I^t_{t'}\big) = ch_{t'}\urcorner\big) > 0$.

That is, the condition requires the chance function at t conditional on the proposition about some possible later chance function to be equal to that later chance function. It is easy to observe that our condition allows chances to be ascribed to propositions about other chances. Thus, we might think of the condition as relating first- and second-order chances.

Although the talk about higher-order chances might strike us as odd, it is not hard to find cases that in fact involve such chances. Recall the example of the marble-selecting device discussed in section 4.2. We might well think that at t there is a 0.5 chance of there being a 0.6 chance at t' that the device will select a red marble, and there is also a 0.5 chance of there being a 0.4 chance at t' that the device will select a blue marble. Thus, in this case, it seems reasonable to speak about chances assigned to other chances.

Importantly, in cases like the marble-selecting device, Chance-Chance Principle plays a crucial role in determining the initial first-order chance of an outcome (selecting a red marble). Note that, in the case discussed above, there is no determinate value for the chance at time t of the device's selecting a red marble. But assuming that chances are probabilities, we can use the law of total probability and Chance-Chance Principle to determine the value of the chance at time t. For concreteness, let us assume that $K\big(ch_t, I^t_{t'}\big)\big(A\big)$ is the chance at t' of the device's selecting the red marble and that $K\big(ch_t, J^t_{t'}\big)\big(A\big)$ is the other chance at t' of the device's selecting the red marble, where $I^t_{t'}, J^t_{t'} \in \mathcal{I}^t_t$ and $I^t_{t'} \neq J^t_{t'}$. Then,

$$ch_t(A) = ch_t\big(A\big|\ulcorner K\big(ch_t, I^t_{t'}\big) = ch_{t'}\urcorner\big)ch_t\big(\ulcorner K\big(ch_t, I^t_{t'}\big) = ch_{t'}\urcorner\big)$$

[1] Here I assume that $\ulcorner K(ch_t, I^t_{t'}) = ch_{t'}\urcorner \in \mathcal{F}$ for all $K\big(ch_t, I^t_{t'}\big)$ in the set of chance functions over $\mathcal{F}$.

$$
\begin{aligned}
&+ ch_t\big(A\big|\ulcorner K\big(ch_t, J^t_{t'}\big) = ch^*_{t'}\urcorner\big)ch_t\big(\ulcorner K\big(ch_t, J^t_{t'}\big) = ch^*_{t'}\urcorner\big)\\
&= ch_{t'}(A)ch_t\big(\ulcorner K\big(ch_t, I^t_{t'}\big) = ch_{t'}\urcorner\big)\\
&+ ch^*_{t'}(A)ch_t\big(\ulcorner K\big(ch_t, J^t_{t'}\big) = ch^*_{t'}\urcorner\big)\\
&= (0.4)(0.5) + (0.6)(0.5)\\
&= 0.5.
\end{aligned}
$$

Now, let us focus on certain consequences stemming from Chance-Chance Principle, consequences that are the linchpin of a new justification of Bayesian Kinematics of Chance. If we define the set of possible later chance functions at t' as:

$$\mathcal{C}_{(t,t')} = \left\{ch_{t'} : ch_t\big(\ulcorner K\big(ch_t, I^t_{t'}\big) = ch_{t'}\urcorner\big) > 0\right\}, \tag{4.25}$$

then Chance-Chance Principle entails the following principle:

Chance Expectation: For all $A \in \mathcal{F}$,

$$ch_t(A) = \sum_{ch_{t'} \in \mathcal{C}_{(t,t')}} ch_t\big(\ulcorner K\big(ch_t, I^t_{t'}\big) = ch_{t'}\urcorner\big)ch_{t'}(A).$$

This principle tells us that the chance function at t ought to be a weighted average of possible posterior chance functions, where the weights are chances at t assigned to propositions about those possible posterior chance functions.

In order to show how Chance-Chance Principle entails Chance Expectation, let us prove the following proposition:

Proposition 4.3. *Chance-Chance Principle ⇒ Chance Expectation.*

Proof. Suppose that ch_t satisfies Chance-Chance Principle, and assume that $\ulcorner K\big(ch_t, I^t_{t'}\big) = ch_{t'}\urcorner \in \mathcal{F}$ for all $ch_{t'}$ in $\mathcal{C}_{(t,t')}$. Then,

$$
\begin{aligned}
ch_t(A) &= ch_t\left(\bigvee_{ch_{t'} \in \mathcal{C}_{(t,t')}} \big(A \wedge \ulcorner K\big(ch_t, I^t_{t'}\big) = ch_{t'}\urcorner\big)\right)\\
&= \sum_{ch_{t'} \in \mathcal{C}_{(t,t')}} ch_t\big(A \wedge \ulcorner K\big(ch_t, I^t_{t'}\big) = ch_{t'}\urcorner\big)\\
&= \sum_{ch_{t'} \in \mathcal{C}_{(t,t')}} ch_t\big(A\big|\ulcorner K\left(ch_t, I^t_{t'}\right) = ch_{t'}\urcorner\big)ch_t\big(\ulcorner K\big(ch_t, I^t_{t'}\big) = ch_{t'}\urcorner\big)
\end{aligned}
$$

$$= \sum_{ch_{t'} \in \mathcal{C}_{(t,t')}} ch_t\big(\ulcorner K\,(ch_t, I^t_{t'}) = ch_{t'}\urcorner\big) ch_{t'}(A)$$

(by Chance-Chance Principle),

as required. □

We can also introduce a generalization of Chance Expectation. To this end, we define a set of all convex combinations of the possible later chance functions in $\mathcal{C}_{(t,t')}$, called the convex hull of $\mathcal{C}_{(t,t')}$:

$$Conv\big(\mathcal{C}_{(t,t')}\big) = \left\{ \sum_{ch_{t'} \in \mathcal{C}_{(t,t')}} \lambda_{ch_{t'}} ch_{t'} \,:\, 0 \leq \lambda_{ch_{t'}} \leq 1, \sum_{ch_{t'} \in \mathcal{C}_{(t,t')}} \lambda_{ch_{t'}} = 1 \right\}. \tag{4.26}$$

That is, $Conv\big(\mathcal{C}_{(t,t')}\big)$ is the smallest set that (i) contains $\mathcal{C}_{(t,t')}$, and (ii) contains, for any two later chance functions $ch_{t'}$ and $ch^*_{t'}$, every convex combination or mixture of them, i.e. for any $0 \leq \lambda_{ch_{t'}} \leq 1$, it contains $\lambda_{ch_{t'}} ch_{t'} + \big(1 - \lambda_{ch_{t'}}\big)\, ch^*_{t'}$. Alternatively, $Conv\big(\mathcal{C}_{(t,t')}\big)$ may be defined as the intersection of all convex sets containing $\mathcal{C}_{(t,t')}$. With this notion in mind, we can state the following generalization of Chance Expectation:

> **Generalized Chance Expectation:** There are $\lambda_{ch_{t'}}$'s, with $0 \leq \lambda_{ch_{t'}} \leq 1$ and $\sum_{ch_{t'} \in \mathcal{C}_{(t,t')}} \lambda_{ch_{t'}} = 1$, such that for all $A \in \mathcal{F}$,
>
> $$ch_t(A) = \sum_{ch_{t'} \in \mathcal{C}_{(t,t')}} \lambda_{ch_{t'}} ch_{t'}(A).$$

That is, Generalized Chance Expectation requires the chance function at t to be in the convex hull of possible later chance functions. This means that the chance function at t should lie within the range spanned by possible later chance functions. It is straightforward to observe that Chance Expectation entails Generalized Chance Expectation, for $0 \leq ch_t\big(\ulcorner K\big(ch_t, I^t_{t'}\big) = ch_{t'}\urcorner\big) \leq 1$ and $\sum_{ch_{t'} \in \mathcal{C}_{(t,t')}} ch_t\big(\ulcorner K\big(ch_t, I^t_{t'}\big) = ch_{t'}\urcorner\big) = 1$. That is, it follows from Chance Expectation that ch_t is a convex combination of the elements in $\mathcal{C}_{(t,t')}$. But while Chance Expectation makes it clear that the coefficients $\lambda_{ch_{t'}}$ are the chances at time t assigned to propositions about possible later chance functions, Generalized

Chance Expectation says only that these coefficients are non-negative numbers that sum to one. The latter thus allows us to stay neutral as to whether chances can be assigned to propositions about some other chances. But although I focus only on Generalized Chance Expectation, the coming results can also be applied to Chance Expectation.

Interestingly, it can be shown that Generalized Chance Expectation—a principle that relates chance functions at two different times—is intimately connected with Bayesian Kinematics of Chance. In fact, under certain assumptions, these two requirements for chances are equivalent. This relation can be established by proving the following theorem:

Theorem 4.1. *Suppose that* $ch_{t'}\big(I^t_{t'}\big) = 1$ *for all* $I^t_{t'} \in \mathcal{I}^t_{t'}$*. Then, Generalized Chance Expectation* $\Leftrightarrow$ *Bayesian Kinematics of Chance.*

Proof. ($\Rightarrow$) Suppose that Generalized Chance Expectation holds true. For notational convention, assume that $K\big(ch_t, I^t_{t'}\big) = ch_{t'}$. Since $\mathcal{I}^t_{t'}$ is a partition and $ch_{t'}$ is a probability function over $\mathcal{F}$, it follows that if $ch_{t'}\big(I^t_{t'}\big) = 1$, then for all $J^t_{t'} \in \mathcal{I}^t_{t'}$ such that $J^t_{t'} \neq I^t_{t'}$, $ch_{t'}\big(J^t_{t'}\big) = 0$. Then,

$$\begin{aligned} ch_t\big(I^t_{t'}\big) &= \sum_{ch_{t'} \in \mathcal{C}_{(t,t')}} \lambda_{ch_{t'}} ch_{t'}\big(I^t_{t'}\big) \\ &= \lambda_{ch_{t'}} ch_{t'}\big(I^t_{t'}\big) \\ &= \lambda_{ch_{t'}}. \end{aligned}$$

So we have that

$$\begin{aligned} ch_t(A) &= \sum_{ch_{t'} \in \mathcal{C}_{(t,t')}} \lambda_{ch_{t'}} ch_{t'}(A) \\ &= \sum_{ch_{t'} \in \mathcal{C}_{(t,t')}} ch_t\big(I^t_{t'}\big) ch_{t'}(A) \\ &= \sum_{ch_{t'} \in \mathcal{C}_{(t,t')}} ch_t\big(I^t_{t'}\big) K\big(ch_t, I^t_{t'}\big)\big(A\big). \end{aligned}$$

Since $\mathcal{I}^t_{t'}$ is a partition, we have that $ch_t(A) = \sum_{I^t_{t'} \in \mathcal{I}^t_{t'}} ch_t\big(A \wedge I^t_{t'}\big)$, and hence

$$\sum_{I^t_{t'} \in \mathcal{I}^t_{t'}} ch_t\big(A \wedge I^t_{t'}\big) = \sum_{ch_{t'} \in \mathcal{C}_{(t,t')}} ch_t\big(I^t_{t'}\big) K\big(ch_t, I^t_{t'}\big)\big(A\big).$$

And, by the assumption that $ch_{t'}\big(I^t_{t'}\big) = 1$, we have that

$$ch_t\big(A \wedge I^t_{t'}\big) = ch_t\big(I^t_{t'}\big) K\big(ch_t, I^t_{t'}\big)\big(A\big).$$

Dividing the last equation by $ch_t\big(I^t_{t'}\big)$, we get

$$\begin{aligned} K\big(ch_t, I^t_{t'}\big)\big(A\big) &= \frac{ch_t\big(A \wedge I^t_{t'}\big)}{ch_t\big(I^t_{t'}\big)} \\ &= ch_t\big(A | I^t_{t'}\big), \end{aligned}$$

which yields Bayesian Kinematics of Chance.

($\Leftarrow$) Suppose that Bayesian Kinematics of Chance holds true. For notational convention, assume that $K\big(ch_t, I^t_{t'}\big) = ch_{t'}$. Then,

$$\begin{aligned} ch_t(A) &= ch_t\left(\bigvee_{I^t_{t'} \in \mathcal{I}^t_{t'}} (A \wedge I^t_{t'})\right) \\ &= \sum_{I^t_{t'} \in \mathcal{I}^t_{t'}} ch_t\big(A \wedge I^t_{t'}\big) \\ &= \sum_{I^t_{t'} \in \mathcal{I}^t_{t'}} ch_t\big(I^t_{t'}\big) ch_t\big(A | I^t_{t'}\big) \\ &= \sum_{I^t_{t'} \in \mathcal{I}^t_{t'}} ch_t\big(I^t_{t'}\big) K\big(ch_t, I^t_{t'}\big)\big(A\big) \text{ (by Bayesian Kinematics of Chance)} \\ &= \sum_{ch_{t'} \in \mathcal{C}_{(t,t')}} \lambda_{ch_{t'}} K\big(ch_t, I^t_{t'}\big)\big(A\big) \text{ (for } \lambda_{ch_{t'}} = ch_t\big(I^t_{t'}\big)) \\ &= \sum_{ch_{t'} \in \mathcal{C}_{(t,t')}} \lambda_{ch_{t'}} ch_{t'}(A), \end{aligned}$$

as required. $\square$

Thus, given that the later chance function assigns chance 1 to the true element of the partition containing propositions about the complete intervening histories, Generalized Chance Expectation leads to Bayesian Kinematics of Chance, and vice versa. So whenever the relation between chance functions at different times is governed by Generalized Chance Expectation, Bayesian Kinematics of Chance follows.

The assumption used in Theorem 4.1 tells us that exactly one of the elements in $\mathcal{I}^t_{t'}$ will be the true complete intervening history of chance events between t and t'. This assumption, of course, amounts to Chance Certainty, as introduced in section 4.2. Theorem 4.1 thus can be read as showing that Chance Certainty and Generalized Chance Expectation are equivalent to Bayesian Kinematics of Chance. As it is easy to observe in the proof of Theorem 4.1, Bayesian Kinematics of Chance entails Generalized Chance Expectation without assuming Chance Certainty.

Interestingly, by using a similar train of thought, we can justify Chance Rigidity as a constraint on any transition from one chance function to another chance function. This result, which holds independently of Theorem 4.1, can be established by the truth of the following proposition:

Proposition 4.4. *Suppose that $ch_{t'}\big(I^t_{t'}\big) = 1$ for all $I^t_{t'} \in \mathcal{I}^t_{t'}$. Then, Generalized Chance Expectation $\Leftrightarrow$ Chance Rigidity.*

Proof. ($\Rightarrow$) Suppose that Generalized Chance Expectation holds true. For notational convention, assume that $K\big(ch_t, I^t_{t'}\big) = ch_{t'}$. Since $\mathcal{I}^t_{t'}$ is a partition and $ch_{t'}$ is a probability function over $\mathcal{F}$, it follows that if $ch_{t'}(I^t_{t'}) = 1$, then for all $J^t_{t'} \in \mathcal{I}^t_{t'}$ such that $J^t_{t'} \neq I^t_{t'}$, $ch_{t'}\big(J^t_{t'}\big) = 0$. Then,

$$
\begin{aligned}
ch_t\big(A|I^t_{t'}\big) &= \frac{ch_t\big(A \wedge I^t_{t'}\big)}{ch_t\big(I^t_{t'}\big)} \\
&= \frac{\sum_{ch_{t'} \in \mathcal{C}_{(t,t')}} \lambda_{ch_{t'}} ch_{t'}\big(A \wedge I^t_{t'}\big)}{\sum_{ch_{t'} \in \mathcal{C}_{(t,t')}} \lambda_{ch_{t'}} ch_{t'}\big(I^t_{t'}\big)} \\
&\quad \text{(by Generalized Chance Expectation)} \\
&= \sum_{ch_{t'} \in \mathcal{C}_{(t,t')}} \lambda_{ch_{t'}} ch_{t'}\big(I^t_{t'}\big) ch_{t'}\big(A|I^t_{t'}\big)
\end{aligned}
$$

(by the fact that $ch_{t'}(I^t_{t'}) = 1$ and the product rule of probability)
$= ch_{t'}\big(A|I^t_{t'}\big)$ (by the fact that $ch_{t'}(I^t_{t'}) = 1$)
$= K\big(ch_t, I^t_{t'}\big)\big(A|I^t_{t'}\big)$ (by assumption),

which yields Chance Rigidity.

(⇐) Suppose that Chance Rigidity holds true. For notational convenience, assume that $K\big(ch_t, I^t_{t'}\big) = ch_{t'}$. Then,

$$\begin{aligned}
ch_t(A) &= ch_t\left(\bigvee_{I^t_{t'} \in \mathcal{I}^t_{t'}} (A \wedge I^t_{t'})\right)\\
&= \sum_{I^t_{t'} \in \mathcal{I}^t_{t'}} ch_t\big(A \wedge I^t_{t'}\big)\\
&= \sum_{I^t_{t'} \in \mathcal{I}^t_{t'}} ch_t\big(I^t_{t'}\big)ch_t\big(A|I^t_{t'}\big)\\
&= \sum_{I^t_{t'} \in \mathcal{I}^t_{t'}} ch_t\big(I^t_{t'}\big)K\big(ch_t, I^t_{t'}\big)\big(A|I^t_{t'}\big) \text{ (by Chance Rigidity)}\\
&= \sum_{I^t_{t'} \in \mathcal{I}^t_{t'}} ch_t\big(I^t_{t'}\big)K\big(ch_t, I^t_{t'}\big)\big(A\big)\\
&\big(\text{by the fact that } K\big(ch_t, I^t_{t'}\big)\big(I^t_{t'}\big) = 1 \text{ for all } I^t_{t'} \in \mathcal{I}^t_{t'}\big)\\
&= \sum_{ch_{t'} \in \mathcal{C}_{(t,t')}} \lambda_{ch_{t'}} K\big(ch_t, I^t_{t'}\big)\big(A\big) \ \big(\text{for } \lambda_{ch_{t'}} = ch_t(I^t_{t'})\big)\\
&= \sum_{ch_{t'} \in \mathcal{C}_{(t,t')}} \lambda_{ch_{t'}} ch_{t'}\big(A\big),
\end{aligned}$$

which yields Generalized Chance Expectation. □

From this proposition, it follows that if both Chance Certainty and Generalized Chance Expectation are satisfied, then Chance Rigidity holds true. Therefore, Generalized Chance Expectation entails any kinematics rule for chance that obeys Chance Rigidity. Of course, it is easy to observe that this result follows already from Theorem 4.1 and Proposition 4.1. That is, since Generalized Chance Expectation and Chance Certainty are equivalent to Bayesian Kinematics of Chance, then, in view of Proposition 4.1, they are also equivalent to Chance Rigidity. But

Proposition 4.4 holds independently of the facts established by Theorem 4.1 and Proposition 4.1.

We have just shown that Bayesian Kinematics of Chance could be justified without appealing to the Principal Principle: under a fairly plausible assumption, it is entailed by a principle relating chance functions at different times. This sort of justification might be called *objective*, for it relies entirely on facts concerning the relation between physical probabilities, and not on facts about how physical probabilities constrain one's credences. In this way, the argument just given mitigates the problems surrounding Lewis's argument that were described in section 4.4. In the next section, I shall try to bolster the argument just given by providing additional motivation for Generalized Chance Expectation.

4.6 Motivating Generalized Chance Expectation

Why should Generalized Chance Expectation be a requirement for chances? While Lewis's Principal Principle is a widely acceptable constraint on any chance distribution, it seems that we lack a good motivation for accepting Generalized Chance Expectation. In this section, I shall try to put Generalized Chance Expectation on a firm footing by using considerations of resiliency similar to those presented in chapter 3.

The sort of resiliency considerations that are put forward in this section amount to chance's invariance across changes in a chance set-up. That is, we consider a chance set-up at time t and its possible changes at t', $t' > t$, described by complete intervening histories. We then ask to what extent a chance function at t over the chancy events produced by that chance set-up changes under variation of these intervening histories. Intuitively, if a mechanism producing chancy outcomes remains unaltered in time, the statistical laws should give chance distributions over these outcomes that stay unchanged in time. A similar thought underpins modern theory of stationary dynamical systems. For example, Patrick Billingsley (1965, pp. 1-2) wrote that "if time does not alter the roulette wheel, the gambler's fortunes fluctuate according to constant probability laws". And similarly, Donald Ornstein wrote that:

> A stationary random process can be thought of as a box that prints out one

> letter each unit of time, where the probability of printing out a given letter may depend on the letters already printed out, but is independent of time (that is, the mechanism in the box does not change). (Ornstein 1974, p. 2)

But if a chance set-up changes in time, the idea of chance distributions being unaltered in time cannot be fully maintained: we should expect that statistical laws give chance distributions that track these changes. However, we might still think that, for any time t, statistical laws give chance distributions that maximize resiliency under variation of complete intervening histories.

One way to unpack this idea is to think of resiliency as a quantity measured by the extent to which a chance function at t diverges from possible chance functions at t'. And each possible chance function at t' is a chance function that accommodates a possible change in the chance set-up at t'. While here we should not require a chance distribution at t to remain unaltered as the chance set-up changes at t', we however might reasonably require that statistical laws give the chance function at t that cannot be replaced by a chance function that is less divergent from the possible chance functions at t'. As will be apparent, within a resiliency framework to be proposed below, any chance function at t that satisfies Generalized Chance Expectation maximizes resiliency in the sense just given.

The first component of the resiliency framework is a particular resiliency measure. This measure is meant to express the resiliency of a chance function ch_t over a particular complete intervening history $I^t_{t'} \in \mathcal{I}^t_{t'}$. And this kind of resiliency is measured by the extent to which the chance function ch_t diverges from a possible later chance function that accommodates the information encoded by $I^t_{t'} \in \mathcal{I}^t_{t'}$. If that possible later chance function is $K\big(ch_t, I^t_{t'}\big)$, then the resiliency of ch_t over $I^t_{t'}$ is given by

$$\mathbf{R}\big(ch_t, I^t_{t'}\big) = D\big(K\big(ch_t, I^t_{t'}\big), ch_t\big), \tag{4.27}$$

where $D\big(K\big(ch_t, I^t_{t'}\big), ch_t\big)$ is a measure of the divergence between two chance functions. By convention, I take $D\big(K\big(ch_t, I^t_{t'}\big), ch_t\big)$ to be non-negative, i.e. $D\big(K\big(ch_t, I^t_{t'}\big), ch_t\big) \geq 0$ with equality if and only if $ch_t = K\big(ch_t, I^t_{t'}\big)$. Further, I assume that our divergence measure belongs to the class of Bregman divergences. This class encompasses a number of interesting measures of statistical divergence, e.g. the squared loss function, the Kullback-Leibler divergence, or the Mahalanobis distance. For the purposes of this chapter, we can introduce

Bregman divergence as follows. Let $\mathbf{x}$ and $\mathbf{y}$ be real-valued vectors in $[0,1]^n$, representing, respectively ch_t and $K\big(ch_t, I^t_{t'}\big)$. Then:

> **Bregman divergence:** Suppose that $[0,1]^n$ is a convex subset of $\mathbb{R}^n$. Let $\Phi : [0,1]^n \to \mathbb{R}$ be a strictly convex function whose gradient $\nabla\Phi$ is defined in the interior of $[0,1]^n$ and extends to a bounded, continuous function on $[0,1]^n$. Then, for all $\mathbf{x}, \mathbf{y} \in [0,1]^n$, the Bregman divergence $D_\Phi : [0,1]^n \times [0,1]^n \to \mathbb{R}$ from $\mathbf{y}$ to $\mathbf{x}$ corresponding to Φ is given by
>
> $$D_\Phi(\mathbf{y}, \mathbf{x}) = \Phi(\mathbf{y}) - \Phi(\mathbf{x}) - \nabla\Phi(\mathbf{x}) \cdot (\mathbf{y} - \mathbf{x}),$$
>
> where $\nabla\Phi(\mathbf{x})$ is the gradient vector of Φ evaluated at $\mathbf{x}$ and $\cdot$ denotes the inner product of two vectors. In words, the function D_Φ is the difference between the value of Φ at at $\mathbf{y}$ and the tangent to Φ evaluated at $\mathbf{x}$.

The second component of our framework is an explication of the idea of resiliency maximization. To this end, I first introduce two important notions. Let $\mathcal{C}_t$ be a set of chance functions at t that give rise to the same later chance functions in $\mathcal{C}_{(t,t')}$. Suppose that ch_t, ch'_t are in $\mathcal{C}_t$. Assume that $\mathbf{R}(ch_t, I^t_{t'}) = D_\Phi(\mathbf{y}, \mathbf{x})$. Then, we say that:

- ch'_t *strongly resiliency-dominates* ch_t if $\mathbf{R}(ch'_t, I^t_{t'}) < \mathbf{R}(ch_t, I^t_{t'})$ for all $I^t_{t'} \in \mathcal{I}^t_{t'}$,
- ch'_t *weakly resiliency-dominates* ch_t if $\mathbf{R}(ch'_t, I^t_{t'}) \leq \mathbf{R}(ch_t, I^t_{t'})$ for all $I^t_{t'} \in \mathcal{I}^t_{t'}$, and $\mathbf{R}(ch'_t, I^t_{t'}) < \mathbf{R}(ch_t, I^t_{t'})$ for at least $I^t_{t'} \in \mathcal{I}^t_{t'}$.

That is, ch'_t *strongly resiliency-dominates* ch_t if it is less divergent from every later chance function in $\mathcal{C}_{(t,t')}$. And ch'_t *weakly resiliency-dominates* ch_t if it is at least as divergent as ch_t is from every later chance function in $\mathcal{C}_{(t,t')}$, and it is less divergent from at least one later chance function in $\mathcal{C}_{(t,t')}$. Now, we can explicate the idea of resiliency maximization by the following norm imposed on every chance function in $\mathcal{C}_t$:

> **Maximizing Resiliency:** Suppose that ch_t and ch'_t are in $\mathcal{C}_t$. If

(i) ch'_t *strongly resiliency-dominates* ch_t, and

(ii) there is no other ch''_t in $\mathcal{C}_t$ that *weakly resiliency-dominates* ch'_t,

then ch_t is inadmissible with respect to the resiliency measure.

Thus, the norm of Maximizing Resiliency tells us which chance functions at t in $\mathcal{C}_t$ are inadmissible with respect to our resiliency measure. According to this norm, an inadmissible chance function at t is one that is strongly resiliency-dominated by some other chance function at t that itself is not weakly resiliency-dominated by any other chance function at t from $\mathcal{C}_t$.

Within our resiliency framework, we can establish the following result:

Theorem 4.2. *Let* $\mathbf{R}(ch_t, I^t_{t'}) = D_\Phi(\mathbf{y}, \mathbf{x})$. *Suppose that* $Conv\big(\mathcal{C}_{(t,t')}\big) \subseteq [0,1]^n$. *Then:*

(i) If ch_t *does not satisfy Generalized Chance Expectation, then there is* ch'_t *such that it satisfies Generalized Chance Expectation and* $\mathbf{R}(ch'_t, I^t_{t'}) < \mathbf{R}(ch_t, I^t_{t'})$ *for all* $I^t_{t'} \in \mathcal{I}^t_{t'}$.

(ii) If ch'_t *satisfies Generalized Chance Expectation, then there is no* $ch''_t \in \mathcal{C}_t$ *such that* $\mathbf{R}(ch''_t, I^t_{t'}) \leq \mathbf{R}(ch'_t, I^t_{t'})$ *for all* $I^t_{t'} \in \mathcal{I}^t_{t'}$, *and* $\mathbf{R}(ch''_t, I^t_{t'}) < \mathbf{R}(ch'_t, I^t_{t'})$ *for at least one* $I^t_{t'} \in \mathcal{I}^t_{t'}$.

Proof. For the proof:

- we assume that ch_t, ch'_t, ch''_t, and $K\big(ch_t, I^t_{t'}\big)$ are represented as real-valued vectors $\mathbf{v}$, $\mathbf{w}$, $\mathbf{x}$, $\mathbf{y}$ respectively.

- we use a property of Bregman divergence called the extended Pythagorean property. It can be stated as follows. Let $Conv\big(\mathcal{C}_{(t,t')}\big)$ be a closed convex non-empty set in $[0,1]^n$. Let $\mathbf{w}$ be the D_Φ-projection of $\mathbf{v} \in [0,1]^n - Conv\big(\mathcal{C}_{(t,t')}\big)$ into $Conv\big(\mathcal{C}_{(t,t')}\big)$. Then, for all $\mathbf{y} \in Conv\big(\mathcal{C}_{(t,t')}\big)$ and $\mathbf{v} \in [0,1]^n - Conv\big(\mathcal{C}_{(t,t')}\big)$,

$$D_\Phi(\mathbf{y}, \mathbf{w}) + D_\Phi(\mathbf{w}, \mathbf{v}) \leq D_\Phi(\mathbf{y}, \mathbf{v}). \tag{4.28}$$

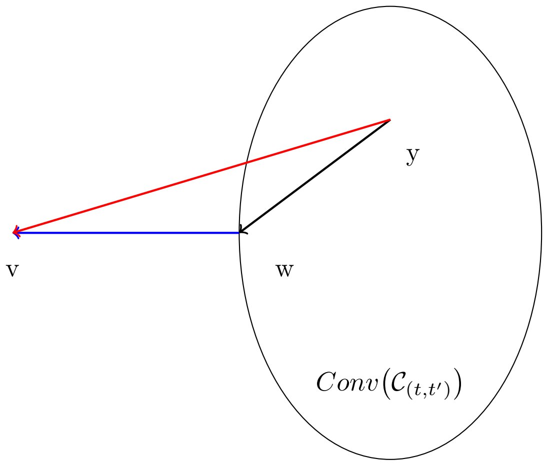

Figure 4.4: Visualizing the extended Pythagorean property: $D_\Phi(\mathbf{y}, \mathbf{w}) + D_\Phi(\mathbf{w}, \mathbf{v}) \leq D_\Phi(\mathbf{y}, \mathbf{v})$.

For the proof of Theorem 4.2 (i), suppose that ch_t does not satisfy Generalized Chance Expectation. This means that $\mathbf{v} \in [0,1]^n - Conv\big(\mathcal{C}_{(t,t')}\big)$. Since ch'_t satisfies Generalized Chance Expectation, we have that $\mathbf{w} \in Conv\big(\mathcal{C}_{(t,t')}\big)$. Hence, $\mathbf{v} \neq \mathbf{w}$ and so $D_\Phi(\mathbf{w}, \mathbf{v}) > 0$. Then, by (4.28), for all $\mathbf{y} \in Conv\big(\mathcal{C}_{(t,t')}\big)$, and hence for $\mathbf{y} \in \mathcal{C}_{(t,t')}$

$$D_\Phi(\mathbf{y}, \mathbf{w}) < D_\Phi(\mathbf{y}, \mathbf{v}),$$

and so

$$\mathbf{R}(ch'_t, I^t_{t'}) < \mathbf{R}(ch_t, I^t_{t'}),$$

as required.

For the proof of Theorem 4.2 (ii), suppose that ch'_t satisfies Generalized Chance Expectation. This means that $\mathbf{w} \in Conv\big(\mathcal{C}_{(t,t')}\big)$. Now, suppose that $D_\Phi(\mathbf{y}, \mathbf{x}) \leq D_\Phi(\mathbf{y}, \mathbf{w})$ for all $\mathbf{y} \in Conv\big(\mathcal{C}_{(t,t')}\big)$, and hence for all $\mathbf{y} \in \mathcal{C}_{(t,t')}$. Since $\mathbf{w} \in Conv\big(\mathcal{C}_{(t,t')}\big)$, it follows that $D_\Phi(\mathbf{y}, \mathbf{x}) \leq D_\Phi(\mathbf{w}, \mathbf{w}) = 0$. This implies that $D_\Phi(\mathbf{y}, \mathbf{x}) = 0$ and $\mathbf{x} = \mathbf{w}$. But this contradicts the assumption that $\mathbf{x} \neq \mathbf{w}$. $\square$

Theorem 4.2 says that any chance function at t that does not satisfy Generalized Chance Expectation will be ruled out as inadmissible with respect to our resiliency measure. And if it satisfies Generalized Chance Expectation, it will not be ruled out as inadmissible with respect to that resiliency measure. In the first part, the theorem says that if a chance function at t does not satisfy Generalized Chance Expectation, and so lies outside the convex hull of $\mathcal{C}_{(t,t')}$, then it can be replaced by some other chance function that satisfies Generalized Chance Expectation, and hence is in the convex hull of $\mathcal{C}_{(t,t')}$ and is more resilient over every complete intervening history. The second part of the theorem says that once the chance function at t satisfies Generalized Chance Expectation, its resiliency will not be improved by replacing it with any other chance function from $\mathcal{C}_t$.

Interestingly, there is a striking analogy between the resiliency-based approach to chance just applied and Joyce's accuracy-based approach to credence described in chapter 1. Like Joyce's norm of accuracy, the norm of Maximizing Resiliency is a dominance norm. That is, both norms tell us which options we should rule out as inadmissible with respect to a certain measure of divergence. Moreover, much like in Joyce's approach, the resiliency-based approach enables us to derive some constraints on chance functions, by appealing to the norm of Maximizing Resiliency.

The main result of this section shows that Generalized Chance Expectation enforces chance functions that maximize resiliency. If experimental arrangements, like the marble-selecting device described in section 4.2, change in time, resiliency requires chance functions attached to these arrangements to be maximally stable, or robust, under these changes. If we impose Generalized Chance Expectation as a constraint on any chance function at t, it follows that this chance function maximizes resiliency under possible changes that the experimental arrangements might undergo between t and t', $t' > t$. Hence, not only does Generalized Chance Expectation support Bayesian Kinematics of Chance; it is also resiliency-conducive in the sense explicated in our framework.

4.7 The Scope of Bayesian Kinematics of Chance

The foregoing shows that Bayesian Kinematics of Chance can be supported by a well-motivated principle relating prior and posterior chances. Still, however, one might wonder whether Bayesian Kinematics of Chance, and hence Generalized

Chance Expectation, can be applied to all cases in which chances evolve through time. This section makes some headway in addressing this issue.

In section 4.2, to illustrate how Bayesian Kinematics of Chance works, I have presented an example in which the changes in a chance distribution were due to changes in the categorical properties of the marble-selecting device. This prompts a more general observation: Bayesian Kinematics of Chance seems to work in cases where the distribution of chances over a set of outcomes depends upon the specifics of the chance set-up that produces these outcomes. For example, if we change the mass distribution of a coin, we will alter the parameter value of the binomial distribution associated with the coin flipping. Or, if we change the atomic number of a radioactive atom, we will change the parameter of the exponential distribution, and hence the atom's chance of decaying within t years.

But in recent years philosophers have advanced at least two different ways of thinking about how chance distributions depend on the instantiation of a chance set-up. These two ways are closely tied to two different metaphysical views about chance, to wit, reductionism and non-reductionism about chance. While the former takes facts about chances to be reducible to facts about categorical-property instantiations, the latter takes facts about chances to be ontologically on a par with facts about categorical-property instantiations. The question arises: could Bayesian Kinematics of Chance fit in with these two metaphysical views about chance? In what follows, I first show how Bayesian Kinematics of Chance fits in with, and moreover underpins, reductionism about chance. For concreteness, in discussing this issue I will focus on Lewis's version of reductionism. Second, I shall argue that Bayesian Kinematics of Chance could also work under some non-reductive accounts of chance: according to these accounts of chance, chances are not brute features of a chance set-up, but are closely tied to non-modal, physical properties of that chance set-up. Finally, I shall briefly point out some cases that are beyond the scope of Bayesian Kinematics of Chance.

According to Lewis's (1994, p. 478) version of reductionism about chance, a chance assignment at time t is logically entailed by the probabilistic laws of nature together with the history of categorical-property instantiations up to and including t. That is, for each proposition A in the domain of a chance function, the probabilistic laws of nature entail what Lewis calls a history-to-chance conditional. This conditional says that: if the history up to and including t is H_t, then $ch_t(A) = x$. And together with the history H_t, the conditional entails the fact that $ch_t(A) = x$. Lewis's view might well be read as follows: for any A, the chance of A at t is *determined* by the laws of nature and the history through time t. This view, in fact, is an expression of Lewis's reductionist view about chance: facts about chances are ontologically nothing more than facts about categorical-property instantiations. For the probabilistic laws of nature are also reducible to patterns of categorical-property instantiations. In Lewis's words:

> In general, probabilistic laws yield history-to-chance conditionals. For any given moment, these conditionals tell us the chance distribution over alternative future histories from that moment on, as a function of the previous history of particular facts up to and including that moment. The historical antecedents are of course given by the arrangement of qualities. The laws do the rest. (...). What pattern in the arrangement of qualities makes the chances? In part, features of history up to the moment in question. For the rest, it is the pattern that makes the probabilistic laws, whatever that is. (Lewis 1994, p. 478)

The question of how the probabilistic laws of nature reduce to the patterns of categorical property-instantiations is answered by Lewis's (1994) best-system analysis of laws and chances. That is, they are reducible to the patterns that are described by theorems of the best system for a given world. And this system is a system that strikes the best balance between informativeness, simplicity, and fit. But one needs not to adopt Lewis's best-system analysis of chance to be a reductionist about chance. For example, a frequency theorist takes chances to be reducible to patterns of categorical-property instantiations, but does not regard these patterns as being described by theorems of the best system.

Not only does Bayesian Kinematics of Chance fit well with Lewis's brand of reductionism, but also shows how the chance distribution depends on the history of categorical-property instantiations. To see this, suppose that the probabilistic laws of nature fix an "ur-chance" function, i.e. the chance function at the beginning of the history of categorical-property instantiations.[2] Then, Bayesian Kinematics of Chance tells us how a chance function at any time t relates to the

[2]The phrase "ur-chance" comes from Hall (2004).

ur-chance function and the history through t. That is,

$$ch_t(\cdot) = urch(\cdot|H_t), \tag{4.29}$$

providing $urch(H_t) > 0$. Thus, Bayesian Kinematics of Chance shows clearly that the chance distribution at any time t is a function of two things: the history through t and the ur-chance distribution conditional on that history. Since the ur-chance function is given by the laws of nature that are reducible to the patterns of categorical-property instantiations, the chance distribution at any particular time is a function of the categorical-property instantiations. And this is exactly what Lewis's reductionism about chance amounts to. A neat illustration of this idea comes from Hoefer (2007, pp. 564-565):

> For Lewis, a non-trivial time-indexed objective probability $Ch_{tw}(p)$ is, in effect, the chance of p occurring given the instantiation of a big setup: the entire history of the world up to time t.

Could Bayesian Kinematics of Chance play a similar role in the case of non-reductionism about chance? For concreteness, let us focus on propensity theories of chance. In general, under propensity theories, chance is a physical tendency, or disposition, of an experimental set-up or a physical situation to produce a certain kind of outcome. Although this disposition is not reducible to categorical-property instantiations, it is often taken to be closely tied to the chance set-up or experimental arrangement that produces chancy outcomes (e.g. Popper 1959; Milne 1987). After all, it is a disposition of a chance set-up to produce a certain kind of outcome. And though propensity theorists often disagree over what a chance set-up is, they agree that propensities should be relativized to chance set-ups. For example, Donald Gillies (2000, p. 126-129) distinguishes two understandings of the notion of a chance set-up. According to the first view, a chance set-up is the complete state of the universe at a given time. On the second view, a chance set-up is the complete set of (nomically and/or causally) relevant conditions at a given time. But, given this distinction, it seems plausible to think that propensities are relativized either to a global chance set-up pertaining to the entire history of the universe (presumably, the light cone) up to a given time, or to a more local chance set-up pertaining to some relevant test conditions at a given time.

Importantly, whatever characterization of the chance set-up we choose, it seems that we could use a slightly modified version of Bayesian Kinematics of Chance to fit in with the idea of chances as propensities being relativized to chance set-ups. That is, if we take S_t to denote a proposition describing a chance set-up at time t, then, under propensity theories, the chance distribution at t is given by

$$ch_t(\cdot) = ch(\cdot|S_t), \tag{4.30}$$

providing $ch(S_t) > 0$. That is, if chances can be conditioned on propositions describing chance set-ups, Bayesian Kinematics of Chance shows that the chance distribution at t is a function of the chance set-up instantiated at t and the chance distribution conditional on the instantiation of that chance set-up. And since the chance set-up changes as t changes, so will the chance distribution that results from conditioning on that chance set-up. Thus, Bayesian Kinematics of Chance seems to sit well with the idea that, under propensity theories, chances are intimately related to chance set-ups.

To sum up, not only does Bayesian Kinematics of Chances underpin reductionism about chance; it also resonates well with non-reductive accounts of chance once we recognize that propensity chances should be conditioned on propositions describing chance set-ups.

But it seems that Bayesian Kinematics of Chance cannot be the whole story about how chances evolve through time. There are cases where the posterior chance distribution does not depend on a prior chance distribution conditional on the instantiation of a chance set-up. For example, a chance set-up may undergo a discontinuous change, so that it might suddenly acquire some chance of being in a given state, having had no such chance before. Similarly, as it has been argued by Lange (2006), certain scenarios involving higher-order chances cannot be dealt with conditionalization on complete intervening histories. Scenarios of this sort involve what Fisher (2006) calls *free-floating* chances, i.e. chances that do not evolve in lock-step with complete intervening histories. Cases of this sort show that even if a given complete intervening history materializes at time t, the chance at t has no determinate value. Such cases purport to show that chances do not depend on changes in a chance set-up.

4.8 Concluding Remarks

Bayesian Kinematics of Chance is a particular application of the rule of conditionalization to the domain of chances or physical probabilities. Lewis showed how this kinematical model for chances follows from his Principal Principle. I have shown that in fact the Principal Principle is much more powerful: it entails each of the conditions, to wit Chance Certainty and Chance Rigidity, whose conjunction is equivalent to Bayesian Kinematics of Chance.

But one might reasonably ask what any kinematics of chance has to do with the way chances constrain one's credences. Or, one might ask whether there are other requirements for chances from which Bayesian Kinematics of Chance could follow. In reaction to these considerations, I have presented an alternative argument for Bayesian Kinematics of Chance, an argument that relies on Generalized Chance Expectation—a principle that relates chance functions at different times. In addition, I have shown that this principle places a well-motivated constraint on chances, in the sense that it is an essential feature of resilient chance functions.

It is an open question of what kind of kinematics rule for chances could be devised in the case of free-floating chances. Lange (2006), for example, has suggested that we could retain the essence of Bayesian Kinematics of Chance by providing a different account of the history upon which chances are conditioned: such history would include not only the instantiations of categorical properties, but also the instantiations of first-order chances of these properties. Consequently, Lange's Bayesian Kinematics of Chance allows second-order chances to be conditioned on information about first-order chances. It is a question for further research of whether, and if so how well, Lange's Bayesian Kinematics of Chance sits with principles relating chance functions at different times like Generalized Chance Expectation.

Chapter 5

Legal Proof, Naked Statistical Evidence, and Accuracy: Why Should Chances Constrain Judges' Credences?

This chapter introduces a Bayesian model to tackle the issue of using naked or bare statistical evidence in the context of legal proof, i.e. the problem of whether naked statistical evidence is adequate to support a verdict in a court of law. In doing so, the chapter highlights a way of coordinating a fact-finder's credences with evidence about chance, and shows how a particular chance-credence coordination policy can be supported by appealing to a fairly plausible idea of accuracy in legal fact-finding.

Within the model to be given, the chapter presents two accuracy-based arguments for the thesis that chances should constrain a fact-finder's credences about factual hypotheses discussed in courts of law. The first argument says that the fact-finder's credences informed by chances cannot lead to a decrease of subjective expected *verdict accuracy*: the subjective expected accuracy of verdicts cannot decrease when credences are expected to match chances. The second argument shows that the fact-finder's credences informed by chances maximize objective expected *credence accuracy*: it shows that when the fact-finder, who strives to hold objectively accurate credences, is asked what she should do with information about chance, her optimal answer is to set her credences equal to chances.

5.1 Introduction

The question of whether a fact-finder's (a judge's or juror's) belief informed by exclusively statistical evidence could support a verdict of liability or guilt has been a long-running controversy among legal scholars. Consider the following version of a hypothetical case known among legal scholars as the Blue Bus:[1]

> **The Blue Bus.** Late at night in some city Mrs. Smith was run over by a bus. As a matter of fact, 80 percent of the buses causing accidents in the city belong to the Blue Bus Company, 20 percent of them belong to the Red Bus Company, and no other companies operate bus lines in the city. Mrs. Smith appears to be able to establish 0.8 chance that the accident was caused by a blue bus. She sues the Blue Bus Company. For simplicity, let us assume that other elements of the case, that is, the fact of the injury, negligence, causation, are also established. Could the defendant be found liable solely on the basis of this statistical evidence?

[1] For slightly different versions of this case, see Tversky and Kahneman (1977), Tversky and Kahneman (1982), Thomson (1986), Schauer (2003). Although this case is typically regarded as hypothetical, it resembles some actual court cases. For example, Thomson (1986) draws a parallel between the Blue Bus case and the American court case of *Smith v. Rapid Transit, Inc*, in which the plaintiff was run over by a negligently-driven bus on Main Street and the defendant's bus company owned the entire franchise for operating buses on Main Street. Hamer (1994) shows that the Blue Bus case is similar to the American court case of *Kaminski v. Hertz*, in which the plaintiff was injured by a yellow truck with a Hertz logo on the side and Hertz owned 90% of the trucks with such logo.

In civil cases, the plaintiff, who has the burden of proof, has to establish her case by the "preponderance of the evidence" or on the "balance of probabilities". Explained in probabilistic terms, the legal standard of proof in civil cases means that, for the plaintiff to win, the fact-finder needs to have a subjective probability (a degree of belief or credence) of just over 0.5 that her case is true. And in the Blue Bus, since there is a 0.8 chance that the accident was caused by a blue bus, it appears that the fact-finder would believe that it is more probable than not that a blue bus was responsible for the accident, and so the plaintiff would win the case. But, as the vast majority of legal scholars and commentators claim, if the case reached a court, judges would regard their credences informed by statistical evidence as insufficient for imposing liability on Blue Bus Company, and so the plaintiff would not win her case. The point this hypothetical case makes is that though the fact-finder's credence in plaintiff's case based on statistical information satisfies the proof requirement, there is something intuitively wrong about ascribing liability on the basis of credences informed by *naked statistical evidence*, i.e. statistical evidence that is not accompanied by more conventional evidence like eyewitness testimony. But what is really wrong with basing legal decisions on naked statistical evidence?

According to the predominant view among legal scholars, the use of naked statistical evidence in legal settings is problematic because naked statistical evidence lacks an important quality that "individualized" evidence (e.g. eyewitness testimony) allegedly possesses. Just add the following story to the Blue Bus. Suppose that there is an eyewitness who identifies the bus as belonging to the Blue Bus Company. Even though the eyewitness is not perfectly reliable, say her visual identification ability is accurate eight times out of ten, advocates of this tradition hold that it is rather uncontroversial that a fact-finder's credence informed by such individualized evidence could license a verdict of liability. They claim that this could be so, even if the risk of erroneous finding for the plaintiff based on individualized and naked statistical evidence appears to be the same. But why should we treat statistical and individualized evidence differently? Several explanations have been proposed in the literature to answer this question. Most notably, it has been argued that statistical evidence, unlike individualized evidence, (i) lacks an appropriate causal connection with a given disputed fact at a legal trial (Thomson 1986; Wright 1988), (ii) does not have appropriate weight (Cohen 1977; Stein 2005; Nance 2008; Hamer 2012), (iii) is susceptible to the reference class problem (Colyvan et al. 2001; Allen and Pardo 2007), or (iv) does not make credences sensitive to the truth (Enoch et al. 2012).

In the first part of this chapter, I argue that the aforementioned explanations either fail to pin down the crucial difference between statistical and individualized evidence or fail to show why statistical evidence is problematic, even if they identify that difference correctly. Although the discussion in section 5.2 does not cover all the possible explanations, it casts doubt on whether we could provide a viable understanding of the crucial difference between statistical and individualized evidence. This discussion brings also reason to suggest that instead of focusing on how to explain the problematic value of naked statistical evidence in legal settings, we should rather try to explain why naked statistical evidence might be valuable in such settings.

In the second part of this chapter, I provide an answer to the aforementioned question. More specifically, I show that, within a particular framework and under certain assumptions, naked statistical evidence in courts of law is conducive to the achievement of accuracy, which appears to be a fundamental objective of legal fact-finding. A similar idea has already been expressed in the legal literature. Most notably, Jonathan Koehler and Daniel Shaviro (1990) have argued, albeit in an informal way, that using statistical evidence, regardless of how problematic it might be, enhances the accuracy of legal fact-finding. To make the idea of accuracy in legal proceedings more precise, I introduce, in section 5.3, a simple Bayesian model of the epistemology of legal fact-finding, and explicate a way in which a Bayesian fact-finder's credences are constrained by statistical evidence in the form of information about chance (physical probability). In section 5.4, I show how the accuracy in legal fact-finding could be understood within this framework: I distinguish between *verdict accuracy* and *credence accuracy*. Armed with these formulations, I present two arguments supporting the idea that the Bayesian fact-finder's credences should match chances. The first argument, presented in section 5.5, shows that credences informed by chances cannot lead to harmful verdicts: the expected accuracy of verdicts cannot decrease when credences are expected to match chances. The second argument, given in section 5.6, shows that the objective expected inaccuracy of credences informed by chances can only decrease. In particular, it shows that, with respect to objective expected credence accuracy, a particular policy for coordinating credences with chances cannot be inferior, and could be superior, to the policy of ignoring information about chance.

Before continuing, let me mention another robust tradition explaining why the use of purely statistical evidence in courts of law seems problematic. This tradition puts emphasis on the specifics of legal trial. Advocates of this tradition share the belief that while naked statistical evidence may be useful in contexts such as science and policy-making, its use in courts of law should be restricted because it conflicts with important features of legal context or with fundamental values of the law. For example, David Wasserman (1992) has argued at great length that statistical evidence fails to respect the individuality and the autonomy of the defendant. This is so because it treats the defendant as a predetermined mechanism or a randomly selected member of a given reference class. According to Wasserman, individuality and autonomy are two important moral dimensions of legal fact-finding: they emphasize a commitment of the legal system to "treat the defendant as an autonomous individual, free to determine and alter his conduct at each moment" (Wasserman 1992, p. 943). The merits of such arguments notwithstanding, the approach in this chapter is purely epistemological and does not concern law-specific considerations that might be invoked in the debate about the use of statistical evidence in legal context. It thus does not provide an exhaustive defence of the value of statistical evidence in courts of law. Furthermore, the approach is epistemological insofar as the notion of credence or subjective probability covers a fact-finder's epistemic attitude towards some propositions. But what I do hope to accomplish is to show that a fairly natural idea of accuracy in legal fact-finding, applied to the fact-finder's epistemic attitude covered by credences, can be invoked to support the use of statistical evidence in courts of law.

Concomitantly, I do not claim that considerations of accuracy alone show that credences based on bare statistical evidence suffice to license court verdicts. As it has been pointed out above, there are other central desiderata of a legal adjudication besides accuracy. There are also law-specific and ethical reasons that judges need to take into account when reaching a verdict. Importantly, as will be claimed in section 5.4, some of these reasons may stand in a marked tension with accuracy.

Also, two other disclaimers are in order. First, it is not my goal to argue that naked statistical evidence is no less valuable than individualized evidence when accuracy considerations are invoked. Arguments of this sort have been given by Ferdinand Schoeman (1987). In particular, he has argued that though statistical evidence is perhaps not so persuasive, it is no less reliable than individualized evidence: there is no principal way to differentiate statistical evidence that is 80 percent reliable from a witness testimony that is also 80 percent reliable. But accuracy-based arguments presented here provide a different rationale for using statistical evidence in courts of law: considerations of accuracy show that it is better to use statistical evidence than to ignore it.

Second, the idea of this chapter is orthogonal to the philosophical debate concerning the use of cases like the Blue Bus to argue against the so-called Lockean thesis, i.e. the thesis that one has a belief in a proposition just in case one has a sufficiently high credence in it. In recent years, some authors have argued that the Blue Bus-style cases indicate why belief cannot be reduced to a sufficiently high credence. For example, a recent analysis of Lara Buchak's (2013) shows that one's credence based on bare statistical evidence, though high, cannot justify ascriptions of blame, and so cannot be called one's belief. According to Buchak, it seems that I can't blame the Blue Bus Company for causing harm, even if 80 percent of the buses causing harm in the city belong to that company, and so my credence based on this information is quite high. But, arguably, I can blame this company if I believe that it caused harm. The approach presented in this chapter neither supports nor undermines such arguments.

5.2 The Legal Debate about Naked Statistical Evidence

The predominant view among legal scholars is that we can vindicate the problematic value of naked statistical evidence by showing how this sort of evidence differs from what they call individualized evidence, e.g. eyewitness testimony or confession.[2] That is, they claim that there is some fundamental difference between naked statistical evidence and individualized evidence, and the existence of this difference explains why fact-finders are so reluctant to rely on naked statistical evidence. In this section, I discuss some of the explanations advanced to pin down the difference between naked statistical evidence and individualized evidence. It is not my goal to present and discuss an exhaustive list of these explanations.[3] This section only discusses what I take to be the most influential explanations offered in the legal literature.

5.2.1 Causal Connection

According to Judith Jarvis Thomson (1986), the principal difference between naked statistical and individualized evidence concerns an appropriate causal connection to a given fact for which they are taken as evidence. While individualized evidence is causally linked in the appropriate way to that fact, there is no appropriate causal link between that fact and naked statistical evidence. For example, the fact that the bus causing the accident was blue is causally linked in the appropriate way to the fact that the witness testified that the bus looked blue to her. That is, as Thomson points out, the former fact causally explains the latter. But the fact that the bus causing the accident was blue cannot causally explain the fact that 80 percent of the buses involved in the accidents in the city belong to the Blue Bus Company. Thomson suggests that this difference between individualized and naked statistical evidence justifies why judges express resistance to the latter.

It is not my goal to fully assess Thomson's account. What is important in the context of this chapter is that Thomson's account does not seem to succeed in singling out the difference between individualized and naked statistical evidence. Let me give two reasons for this claim. First, there seems to be a sense in which a given fact causally explains naked statistical evidence. And, importantly, by using this understanding of causal explanation we can show that the fact that

[2] A good discussion of this view is to be found in Redmayne (2008), Pundik (2011), and Blome-Tillmann (2015).

[3] For a good survey of the existing explanations, see Koehler (1991) and Ho (2008, pp. 136-140).

the bus causing the accident was blue causally explains why 80 percent of the buses involved in the accidents in the city belong to the Blue Bus Company. Just consider a simple counterfactual account of causal explanation. According to this account, c causally explains e just in case had c not occurred, e would not have occurred. Now, we might say that if the bus causing the accident had not been blue, it would not have been the case that 80 percent of the buses involved in the accidents in the city belong to the Blue Bus Company. After all, the relative frequency of buses involved in the accidents in the city would have been slightly different.

Secondly, it is not evidently clear that, in cases like the Blue Bus, the fact that the bus causing the accident was blue constitutes a causal explanation of the fact that the witness testified that the bus looked blue to her. Since the witness is not fully reliable, it might well be that the fact of her testimony was caused by something else, say by the fact that she dislikes the Blue Bus Company for having high ticket prices. We might then think that if the blue bus had not caused the accident, the witness would still have testified that the blue bus caused the accident. But again, according to a simple counterfactual account of causal explanation, this possibility shows that the fact that the bus causing the accident was blue does not causally explain the fact about eyewitness testimony.

To conclude, unless Thomson's explanation is augmented with a substantive account of causal explanation, it seems that both individualized and naked statistical evidence might be causally linked in an appropriate way to a fact disputed at a legal trial. Moreover, it is quite plausible that individualized evidence may fail to be causally linked in an appropriate way to that fact.

5.2.2 Weight

Some authors argue that naked statistical evidence lacks an appropriate weight that individualized evidence possesses. For example, David Hamer (2012, p. 136) suggests that "proof at trial requires a body of evidence that not only delivers a high enough probability assessment, but one that is also sufficiently complete or weighty". He notes that, in cases like the Blue Bus one, "if there are obvious categories of missing evidence, one's existing level of certainty may appear fragile" (Hamer 2012, p. 150). That is, it might well be that further evidence such as an admission of a particular bus driver would tell us that it was in fact a red bus that caused the accident.

But while it is true that naked statistical evidence is often not sufficiently *resilient* relative to further evidence, it is hard to see in what sense individualized evidence might be immune to this problem. As more evidence is brought at trial, individualized evidence like eyewitness testimony is likely to become less weighty, and even become irrelevant to a given case. When judges gather evidence related to the credibility of the witness, say by cross-examination, it is quite plausible that the witness testimony might lose the appropriate level of weight. Hence, it seems that individualized evidence is not necessarily less prone to the problem of appropriate weight.

5.2.3 Sensitivity

According to David Enoch, Levi Spectre, and Talia Fisher (2012), the fundamental difference between naked statistical and individualized evidence can be explained by appealing to a particular requirement for belief called sensitivity. They define this requirement as follows:

> **Sensitivity:** A subject's belief that P is sensitive just in case had it not been the case that P, the subject would (most probably) not have believed that P.

Enoch, Spectre, and Fisher claim that when the requirement of sensitivity is satisfied, the subject's belief is appropriately connected to the truth. That is, its satisfaction rules out a situation in which the fact that the subject's belief is true is a matter of epistemic luck.

But how does the requirement of sensitivity explains the difference between naked statistical and individualized evidence? Enoch, Spectre and Fisher argue that while one's belief based on individualized evidence is sensitive, one's belief based naked statistical evidence is not, and thus is epistemically defective. Consider again the Blue Bus. Suppose that the judge finds against the Blue Bus Company based solely of the eyewitness testimony. Is judge's belief that a blue bus caused the accident sensitive? Armed with the requirement of sensitivity, we need to answer the following question: had it not been a blue bus, would the judge have believed that a blue bus caused the accident? Since the eyewitness is pretty reliable, we might reasonably think that had it not been a blue bus, she would not have testified seeing a blue bus. Hence, based on her testimony, the judge would not have believed that a blue bus caused the accident. Thus, the judge's belief is sensitive.

Now, suppose that the judge finds against the Blue Bus Company based solely on naked statistical evidence, i.e. on the fact that 80 percent of the buses causing accidents in the city belong to the Blue Bus Company. Again, we need to answer the following question: had it not been a blue bus, would the judge have believed that a blue bus caused the accident? It is plausible to think that had it not been a blue bus, the statistical evidence concerning the relative frequency of accident-causing blue buses would not have been significantly different. If so, the judge would still have believed that a blue bus caused the accident. Hence, her belief is not sensitive.

Although Enoch, Spectre, and Fisher's sensitivity account appears to be initially promising, it fails to pin down the fundamental difference between individualized and naked statistical evidence. To see this, consider two versions of another hypothetical case known among legal scholars as the Gatecrasher:

> **The Gatecrasher:**[4] John, a rodeo enthusiast, is accused of gatecrashing. It is known that 100 people paid for admission to the rodeo while 1,000 spectators were counted on the seats, of whom John is one. Suppose that there is no testimony as to whether John paid for admission or climbed over the fence. In this scenario, the only available evidence against John is statistical, and concerns the relative frequency of those who did not pay for admission to rodeo among all the spectators, which equals 0.9.

Suppose that the judge finds against John. Then, it seems that the judge's belief that John crashed the gate, based solely on the relative frequency of those spectators who crashed the gate, is not sensitive: if John had not crashed the gate, then the judge would still have believed that John did it. This is so because if John had not crashed the gate, still 899 spectators would have crashed the

[4] This classic case has been presented in Cohen (1977).

gate, and the relative frequency of gatecrashers would stay almost unchanged. But consider another version of the Gatecrasher:[5]

> **The Gatecrasher*:** Bob, a rodeo enthusiast, decides to crash the gate of a rodeo stadium. It is well known that he has the power to inspire others to follow him. As he climbs the fence, a great majority of people in the ticket line get the same idea and follow him in climbing the fence. The organizers of the rodeo decide to count the people in the stadium. It turns out that only 100 people paid for admission to the rodeo while 1,000 spectators were counted on the seats, of whom Bob is one. Bob is accused of gatecrashing. There is no testimony as to whether Bob paid for admission or climbed over the fence. The only available evidence against him is statistical, and concerns the relative frequency of those who did not pay for admission to rodeo among all the spectators, which equals 0.9.

Now, the question arises: is the judge's belief that Bob crashed the gate, which is based solely on naked statistical evidence, sensitive? The answer seems to be: Yes, it is. Notice that if Bob had not crashed the gate, others in the ticket line would most probably not have climbed the fence, and so nobody would have crashed the gate. Hence, the judge would not have believed that Bob crashed the gate. So we seem to have a clear case in which naked statistical evidence makes one's belief sensitive in the same way as individualized evidence does.

5.2.4 Reference Class Problem

Consider the following version of the Blue Bus case:

> **The Blue Bus*:** Late at night in some city Mrs. Smith was run over by a bus. The following three facts can be established. First, 80 percent of the buses causing accidents in the city belong to the Blue Bus Company, 20 percent of them belong to the Red Bus Company, and no other companies operate bus lines in that city. Second, 50 percent of the buses causing accidents on the road where Mrs. Smith

[5] A similar example is to be found in Blome-Tillmann (2015).

> was run over belong to the Blue Bus Company, 50 percent of them belong to the Red Bus Company, and no other companies operate bus lines on that road. Third, 30 percent of the buses causing accidents at night on the road where Mrs. Smith was run over belong to the Blue Bus Company, 70 percent of them belong to the Red Bus Company, and no other companies operate bus lines at night on that road. Mrs. Smith appears to be able to establish 0.8, 0.5, and 0.3 chance that the accident was caused by a blue bus. But which information about chance should she use?

The Blue Bus* illustrates the so-called reference class problem. The relative frequency of accident-causing blue buses changes with the reference class. That is, a blue bus causing the accident may be classified as belonging to the class of buses causing accidents in the city, or to the class of buses causing accidents on the road where Mrs. Smith was run over, or to the class of buses causing accidents at night on the road where Mrs. Smith was run over. But each way of classifying the blue bus gives a different relative frequency of causing the accident. The question arises: which reference class is correct? There seems to be no clear answer to this question. The observation that there is no uniquely correct reference class appears to be particularly troubling when the relative frequencies differ significantly, as in the Blue Bus*.

Even if we endorse the view that the correct reference class should be the narrowest one[6] or the broadest homogeneous one[7], we might still face the problem of whether such classes will be always available.

To see this, assume again that 80 percent of the buses causing accidents on the road where Mrs. Smith was hit belong to the Blue Bus Company, 20 percent of them belong to the Red Bus Company, and no other companies operate bus lines on that road. But suppose that it is also true that 30 percent of the buses causing accidents at night belong to the Blue Bus Company, 70 percent of them

[6] According to Hans Reichenbach (1949), we should choose the narrowest reference class for which reliable statistics is available.

[7] This view was defended by Wesley Salmon (1967, pp. 91-124). He argued that we should choose not the narrowest, but the broadest, available reference class. In addition, the reference class should be homogeneous. And a reference class R is homogeneous with respect to some property if there is no set of properties in terms of which R can be partitioned in a way that would change the relative frequency of that property in R.

belong to the Red Bus Company, and no other companies operate bus lines at night on that road. Mrs. Smith appears to be able to establish 0.8 and 0.3 chance that the accident was caused by a blue bus. But which information about chance should she use?

It seems that neither the narrowest reference class nor the broadest homogeneous reference class could be selected in the scenario given above. How can we compare the "narrowness" of the class of buses causing accidents on the road where Mrs. Smith was hit and the class of buses causing accidents at night? Even if they were comparable, they would, arguably, be equally narrow. Similarly, neither class of the buses seems to be the broadest homogeneous one: if we take the intersection of the two classes, the relative frequency of accident-causing blue buses would change significantly.

Importantly, the reference class problem does not only afflict naked statistical evidence concerning the relative frequency of an event or attribute in some reference class. Consider, for example, Laurence J. Cohen's (1981) suggestion that naked statistical evidence in a legal context should concern evidence about propensities or causal tendencies. But since propensities are features of chance set-ups or experimental arrangements, it is quite plausible that the value of propensity would change, depending on how we describe a given chance set-up. For example, the propensity of a blue bus operating in the city and the propensity of a blue bus operating on a particular road in that city to cause accidents might be entirely different. Here the reference class problem is again fully present: there is no principled way of choosing the correct description of the chance set-up.

Though naked statistical evidence is susceptible to the reference class problem, and individualized evidence seems to be immune to it, it does not seem this difference could explain why naked statistical evidence is so problematic in the legal context. Let me give two reasons for this claim. First, it has to be noticed that individualized evidence might face a problem that closely resembles the reference class problem. As it has been argued by Frederick Schauer (2003, chapter 3), any evidence in courts of law should be supported by some generalization. For example, the eyewitness testimony can be used to support a finding against the Blue Bus Company via a generalization that most eyewitness are reliable. But, typically, we would have various classes of eyewitnesses with different levels of reliability. Which class should we choose as a basis for the generalization? Clearly, we seem to face a problem similar to the reference class problem. But though judges often face it, it is clear that the fact that there might be different classes of eyewitnesses does not block the generalizations that judges make in courts of law.

Secondly, rather than seeing the reference class problem as a problem, we could embrace it. A view of this sort has been defended at length by Alan Hájek (2007). Its core idea is that instead of seeking some privileged reference class, we should accept that there are only relative probabilities, i.e. probabilities relativized to certain conditions or reference classes. Seen in this light, we may say that what the Blue Bus* really illustrates is that the relative frequencies of accident-causing blue buses are, by their very nature, relativized to a reference class. The judge thus does not need to choose the correct relative frequency assignment, for all the relative frequency assignments in that case are equally correct.

Naturally, the judge will have to give some sort of justification for why she has chosen a particular relative frequency assignment as a guide to her belief. But again this does not show why naked statistical evidence should differ from individualized evidence. After all, the judge will also have to give a justification for why she thinks the eyewitness testimony is reliable.

The issue of which chance assignment should guide a fact-finder's credence will resurface in section 5.6, where I introduce the notion of objective expected credence accuracy. There, it will be pointed out that not every conditional chance assignment could be used to determine the objective expected credence accuracy in legal proceedings.

5.2.5 Summary

Neither of the views discussed in this section seems to succeed in explaining why the use of naked statistical evidence is so problematic in the legal context. The explanations appealing to causal connection, weight, and sensitivity fail to pin down the fundamental difference between naked statistical and individualized evidence. Although the explanation that appeals to the reference class problem identifies that difference correctly, it does not succeed in showing that the reference class problem explains why fact-finders are reluctant to rely on naked statistical evidence.

Of course, my examination of the existing explanations of why naked statistical evidence is problematic cannot be complete. Most notably, I have not discussed the views that invoke some law-specific considerations to explain why the use of naked statistical evidence in courts of law is problematic. At best, then, the discussion in this section shows *how* the search for a crucial distinction between naked statistical and individualized evidence might fail to succeed.

In the next part of this chapter, I suggest that instead of seeking a crucial difference between naked statistical and individualized evidence, we should try to show why the use of naked statistical evidence might be valuable for legal adjudication. To this end, I introduce a Bayesian model of legal fact-finding and then show how a fact-finder's credences informed by naked statistical evidence in courts could have some value that is important to the law.

5.3 Bayesianism in Legal Fact-Finding

Bayesianism is a popular philosophy of probability employed in the philosophy of science, epistemology, decision theory, and statistics. Bayesians endorse the degree-of-belief interpretation of mathematical theory of probability as contrasted with various frequency or propensity interpretations. But Bayesianism has also a long tradition in theorizing about evidence law and fact-finding at legal trial. In particular, Bayesian statistics has been used to evaluate and interpret various types of evidence at legal trial (e.g. evidence relating to DNA or character evidence used to impeach a witness), and Bayesian decision theory has been employed to argue that the "the preponderance of the evidence" standard in civil case (understood as a degree of belief of just over 0.5 in the plaintiff's claim) and the "beyond a reasonable doubt" (understood as a degree of belief of just over 0.9 in the defendant's guilt) standard in criminal case minimize, respectively, the expected number of errors made by the fact-finder and the expected cost of wrongful convictions.[8]

[8] A good analysis of the uses of Bayesianism in the legal context is to be found in Kaye (1988). Most specifically, Kaye discusses three main uses of Bayesianism in legal fact-finding: (i) Bayesian interpretation of probability, (ii) Bayesian statistical or inductive inference, and (iii) Bayesian decision theory.

In this section, I introduce a simple Bayesian model of the epistemology of legal fact-finding.[9] The model portrays a fact-finder's, e.g. a judge's, doxastic attitude towards propositions, whose truth is disputed at legal trial, as credences or degrees of belief suitably constrained. The fact-finder's credence in a proposition measures the strength of her belief in that proposition. The constraints imposed on her credences are meant to govern her credences in propositions about the facts at issue, given the evidence presented in the courtroom. With this model in mind, I explicate (i) a way in which fact-finders evaluate the truth of hypotheses in the light of evidence and (ii) a way in which naked statistical evidence bears on the truth of hypotheses.

5.3.1 A Bayesian Model

For simplicity's sake, I consider a civil case of a single defendant and plaintiff, and a single factual hypothesis (hypothesis about a single disputed fact) at trial. Let P stand for the plaintiff's factual hypothesis, say, the proposition that a bus belonging to the Blue Bus Company caused the accident. Let $\neg P$ stand for the defendant's factual hypothesis (the negation of the plaintiff's hypothesis), that is, the proposition that a bus belonging to the Blue Bus Company did not cause the accident. We assume that the fact-finder's epistemic attitude towards the two-element set $\{P, \neg P\}$ is represented by a credence function cr: it takes each proposition in that set and returns a number in $[0, 1]$ that measures her credence in that proposition.

Bayesianism provides precise constraints under which the fact-finder's credence function at legal trial should be regarded as rational. A minimal version of Bayesianism requires the fact-finder to satisfy two constraints. The first constraint runs as follows:

> **Probabilism*:** At every stage of legal trial, it ought to be the case that the fact-finder's credence function over the two-element set $\{P, \neg P\}$ is such that
>
> $$cr(P) + cr(\neg P) = 1.$$

[9] Similar Bayesian models of legal fact-finding are given in Lempert (1977), Dawid (2002), and Redmayne (1998).

This constraint may also be read as follows: if the fact-finder has credence 1 in one proposition, she should have credence 0 in the other. Notice that Probabilism is a consequence of a more general constraint that applies to the case in which the propositions P and $\neg P$ are elements of a set $\mathcal{F}$ which is an algebra of propositions. In such a case, the fact-finder's credence function takes each proposition in that algebra and returns a number in $[0, 1]$. This more general constraint then may be given as follows:

> **Probabilism:** At every stage of legal trial, it ought to be the case that the fact-finder's credence function over $\mathcal{F}$ is such that
>
> 1. cr is normalized, i.e. $cr(\top) = 1$ and $cr(\bot) = 0$, for any tautological proposition $\top$ and any contradictory proposition $\bot$,
> 2. cr is additive, i.e. $cr(X \vee Y) = cr(X) + cr(Y)$ for any mutually exclusive $X, Y \in \mathcal{F}$.

It is easy to see that Probabilism entails Probabilism*. Notice, first, that because $P \vee \neg P$ is a tautological proposition, we have that $cr(P \vee \neg P) = 1$ by the fact that cr is normalized. And since P and $\neg P$ are mutually exclusive, we have that $cr(P \vee \neg P) = cr(P) + cr(\neg P)$ by the fact that cr is additive. So $cr(P) + cr(\neg P) = 1$, as required. In order to model a way in which the fact-finder responses to the evidence at trial, I will focus mainly on Probabilism. That is, I will consider situations in which the fact-finder assigns credences not only to P and $\neg P$.

The second constraint endorsed by minimal Bayesianism relates the fact-finder's credence function at the beginning of a legal trial (her *initial* credence function) to her credence function after the total evidence E at that trial has been introduced (her *end-point* credence function). This constraint, known as Bayesian Conditionalization or Bayes's rule, tells the fact-finder how she should update her initial credence function upon the receipt of total evidence at trial. It can be stated as follows:

> **Bayesian Conditionalization:** If E is the total evidence presented at legal trial and $E \in \mathcal{F}$, then the fact-finder's end-point credence in

every $X \in \mathcal{F}$, denoted by $cr'(X)$, should be given by

$$cr'(X) = cr(X|E),$$

provided that $cr(E) > 0$.

That is, the fact-finder's end-point credence in X should be set equal to her prior conditional credence in X given E. Using Bayes' theorem—a theorem of the probability calculus—we can calculate her prior conditional credence in X given E as follows:

$$cr(X|E) = \frac{cr(E|X)}{cr(E)} cr(X), \tag{5.1}$$

where $cr(E|X)$ is the subjective likelihood of X relative to E, $cr(X)$ is the initial credence in X, and $cr(E)$ is the initial credence about the total evidence in court.

Various arguments can be given to show that the two constraints presented above make the fact-finder's credences rational. Specifically, so-called *synchronic* Dutch book arguments (see, e.g. Ramsey 1931) show that if an agent violates Probabilism, she is not pragmatically rational: she is susceptible to a collection of bets ensuring a negative net payoff, come what may. Similarly, so-called diachronic Dutch book, or Dutch strategy, arguments show that any agent who violates Bayesian Conditionalization is vulnerable to a set of bets which ensure that she suffers a net loss (see Teller 1973; Skyrms 1987b; Lewis 1999). An important assumption of these arguments says that what is for the agent to be pragmatically irrational is to assign credences that leave her willing to accept a collection of bets—a Dutch book—that guarantees to produce a loss. Of course, this sort of arguments is not directly applicable to legal settings: legal scholars are rather reluctant to the idea of identifying a fact-finder's credence in a proposition about some disputed fact with her willingness to accept a bet.[10] But still these arguments could retain their force in the legal context in the form of "de-pragmatized" Dutch book arguments. Most notably, Brian Skyrms (1984) and Brad Armendt (1993) have claimed that Dutch book vulnerability flags an underlying inconsistency (called by Armendt the "divided mind" inconsistency): it

[10]For example, it might be argued that if a judge's credence is identified with her willingness to accept a bet, then the legal system runs the risk of having the judge that takes her duty less seriously than desirable.

manifests the fact that an agent whose credences violate Probabilism really does two different evaluations of the same betting option. For example, a credence in $X \vee Y$ determines one evaluation as a fair bet on $X \vee Y$ and credences in X and Y determine another evaluation of essentially the same betting option. Now, we get compatible evaluations only if credences satisfy the additivity axiom: if the credence in $X \vee Y$ is the sum of the credences in X and Y.[11]

What is, however, more important in the context of this chapter is that various accuracy-based arguments can be devised to show that a fact-finder's credences satisfying these two constraints maximize actual or expected accuracy, and thus are epistemically rational (see Joyce 1998; Leitgeb and Pettigrew 2010b). Exactly how these arguments work will be discussed in more detail in section 5.6, when I introduce the notion of accuracy for the fact-finder's credences.

Could the minimal version of Bayesianism just given tell us how the fact-finder's credences should incorporate information about chances, in cases like the Blue Bus? It is easy to observe that the model presented so far does not discriminate between various types of evidence that the fact-finder might receive in the course of a trial. It just says that she should conditionalize her credences upon receipt of her total evidence whatever that might be. But if the fact-finder knows the chance of P's coming out true, and knows that this evidence is most pertinent to P's truth, how should this information constrain her credence in P? To answer this question, we can extend our version of Bayesianism by adding another constraint on the fact-finder's credences. In general, this constraint says how the fact-finder's credence concerning the chance of P ought to relate to her credence in P. Below I introduce two constraints of this type: one applying to the fact-finder's credences at the initial stage of legal trial, and the other applying to her credences at the end-point stage of that trial.

Let cr be the fact-finder's credence function at the initial stage of legal trial, and let cr' be her credence function at the end-point of that trial. Further, let C_{ch} be the proposition that chances over $\mathcal{F}$ are assigned by the chance function ch. Call this proposition the *chance hypothesis*. That is, ch takes each proposition in

[11] For a defence of "depragmatized" Dutch book arguments, see also David Christensen (2004). He argues that assigning a credence to a proposition *commits* an agent to viewing as fair certain bets on that proposition. So Dutch book vulnerability shows that the agent has doxastic attitudes that commit her to accept a bad combination of bets.

$\mathcal{F}$ and returns a number in $[0, 1]$ which is the chance of that proposition to come out true. The chance hypothesis might concern both propositions about a type- and token-event. For example, it might assign a chance to the proposition that a blue bus caused an accident or to the proposition that the accused crashed the gate. It is assumed that ch is a probability function. Further, it is assumed that $C_{ch} \in \mathcal{F}$ and C_{ch} is the only evidence at the initial stage of legal trial. Given these assumptions, the following two constraints can be introduced:

> **Initial Chance-Credence Principle (IP):** At the initial stage of legal trial, the fact-finder ought to have a credence function such that for all $X \in \mathcal{F}$ and all ch,
>
> $$cr(X|C_{ch}) = ch(X).$$
>
> **End-Point Chance-Credence Principle (EP):** At the end-point stage of legal trial, the fact-finder with total evidence E ought to have a credence function such that for all $X \in \mathcal{F}$ and all ch,
>
> $$cr'(X|C_{ch}) = ch(X|E),$$
>
> provided $ch(E) > 0$.

That is, IP applies to the fact-finder who knows the chances and nothing more at the initial stage of a legal trial. It says that in such a situation her credences should match those chances. This constraint is a close cousin of David Lewis's Principal Principle (Lewis 1986), which is, arguably, the most developed chance-credence principle. Recall that Lewis's constraint applies to an agent's "reasonable initial credence function"—her credence function at the beginning of her epistemic life. The Principal Principle says that if the agent started with a reasonable initial credence in some proposition X and learned that the chance of X at time t is x, and if the rest of her evidence is admissible at t, then one would have credence in X equal to x. By analogy, IP shows how chances constrain the fact-finder's credences at the initial stage of a legal trial—prior to acquiring any evidence pertinent to the disputed facts in the courtroom.

By contrast to IP, EP applies to the fact-finder's epistemic situation at the end-point of a legal trial. It is a requirement for the fact-finder who knows chances and who has total evidence E presented at that trial. It says that in that case the fact-finder's credence in X should match the conditional chance of X given evidence E, provided that the chance of E is positive. Of course, this constraint can be applied if the conditional chance is defined.

There is also another angle from which one might look at the two constraints introduced above. Recall Hall's (2004) distinction between a database- and an analyst-expert function given in chapter 1. That is, while we defer to database-experts because of the evidence they have, we defer to analyst-experts because they are good at analysing evidence they are given. With this distinction in mind, we may think of IP as saying that the fact-finder ought to defer to chance because chance conveys more information than she has at the initial stage of a legal trial. Intuitively, when little evidence is available at the beginning of a trial, chance conveys more information than the fact-finder has, and so is worthy of deference. Similarly, one may think of EP as saying that the fact-finder ought to defer to chance because chance comes to the right conclusions once it is provided with the fact-finder's total evidence at that trial. For example, suppose that the total evidence at trial concerns the age, gender, and ethnicity of the defendant. Given this evidence, the fact-finder may defer to the so-called propensity-for-crime evidence, which concerns the relative frequency of those who commit the crime with which the defendant is charged in the class of people who share the age, gender, and ethnicity that also characterize the defendant. This is so because it is likely that the information about relative frequency is correct, given the evidence.

Three observations concerning the two chance-credence principles just given are important to notice. First, if the fact-finder satisfies IP and Bayesian Conditionalization, then she satisfies EP. That is, if the fact-finder's credences at the initial stage match the chances, and they are updated in the course of a legal trial by conditionalization on her total evidence, then at the end-point of that trial they match the conditional chances, given the total evidence. Formally:

$$\begin{aligned} cr'(X|C_{ch}) &= \frac{cr'(X \wedge C_{ch})}{cr'(C_{ch})} \\ &= \frac{cr(X \wedge C_{ch}|E)}{cr(C_{ch}|E)} \text{ (by Bayesian Conditionalization)} \end{aligned}$$

$$
\begin{aligned}
&= \frac{\frac{cr(X \wedge C_{ch} \wedge E)}{cr(E)}}{\frac{cr(C_{ch} \wedge E)}{cr(E)}} \\
&= \frac{cr(C_{ch})cr(X \wedge E|C_{ch})}{cr(C_{ch})cr(E|C_{ch})} \\
&= \frac{cr(C_{ch})ch(X \wedge E)}{cr(C_{ch})ch(E)} \text{ (by IP)} \\
&= \frac{ch(X \wedge E)}{ch(E)} \\
&= ch(X|E),
\end{aligned} \tag{5.2}
$$

as required.

The second observation is straightforward: if C_{ch} is the only evidence gathered at a legal trial, then EP reduces to IP. This is so because E is a tautological proposition, and hence $cr(E) = 1$ and $cr'(X) = cr(X|E) = cr(X)$. The reduction of EP to IP can be presented more formally as follows:

$$
\begin{aligned}
cr'(X|C_{ch}) &= \frac{cr'(X \wedge C_{ch})}{cr'(C_{ch})} \\
&= \frac{cr(X \wedge C_{ch}|E)}{cr(C_{ch}|E)} \text{ (by Bayesian Conditionalization)} \\
&= \frac{cr(X \wedge C_{ch})}{cr(C_{ch})} \text{ (by the fact that } cr(E) = 1) \\
&= cr(X|C_{ch}) \\
&= ch(X) \text{ (by IP).}
\end{aligned} \tag{5.3}
$$

The third observation concerns the situation in which, though there are other items of evidence pertinent to the truth of X, C_{ch} is the most pertinent among them. That is, other pieces of evidence are "trumpable" by evidence about chance: they give no information about the truth of X that does not go through the chance of X. This situation can be understood in our framework as follows: E and X are stochastically independent according to ch. Then, $cr'(X|C_{ch}) = ch(X|E) = ch(X)$. That is, the fact-finder's credences at the end-point of legal trial should match not conditional, but unconditional chances.

How can we use our Bayesian model to give precise expression of the Blue Bus? To answer this question, some preliminary clarifications are needed. We need to acknowledge that the notion "excusively statistical evidence" used in this case is ambiguous. It seems that two quite different things might be meant by this phrase. First, it could mean "the only available evidence pertinent to the truth of P". Second, it could mean "the evidence most pertinent to the truth of P". In the first case, E is a tautological proposition: there is no further evidence except statistical evidence. In the second one, A and E are stochastically independent according to ch. For example, the fact of the injury in the Blue Bus is trumped by chance, for it gives information about P that does not go beyond the chance of P. But the eyewitness testimony stating that a blue bus caused the accident would not be trumpable by chance: since this evidence entails P, it conveys information that does not go through the chance of P. If so, then the statistical evidence in such a case would not be exclusive.

Importantly, EP can encompass the two interpretations of exclusively statistical evidence in the Blue Bus. If C_{ch} is the only available evidence pertinent to the truth of P, EP reduces to IP. And if C_{ch} is the evidence which is most pertinent to the truth of P, EC recommends the fact-finder to set her credence in P equal to the unconditional chance of P. More precisely, given that the chance of P is 0.8, EP requires the fact-finder's end-point credence in P to be

$$cr'(P|C_{ch}) = ch(P) = 0.8. \tag{5.4}$$

In what follows, I provide a more general support to the thesis that the Bayesian fact-finder at trial should obey EP. That is, I show that the Bayesian fact-finder who is an expected accuracy maximizer at trial should use naked statistical evidence in the manner prescribed by EP.

5.4 Legal Fact-Finding and the Goal of Accuracy

Numerous authors have claimed that accuracy is a principal goal of legal fact-finding (Dworkin 1985, chapter 3; Stein 2005; Goldman 1999, chapter 9; Goldman 2002). Some authors have even tried to explicate what accuracy at a legal trial could mean. One robust tradition explains accuracy as a certain value that should be maximized at the trial process. For example, Edmund Morgan (1948) writes that the court should try to get

> as close an approximation of the truth as is possible. (Morgan 1948, pp. 184-185)

In a similar vein, Alvin Goldman (1999) claims that

> it is a vital and central desideratum of a legal adjudication system that it promotes the rendering of accurate verdicts. No system can be perfect, in part because parties to law suits are commonly prone to deception, and deception is hard to detect. Nonetheless, accuracy is certainly to be sought, as far as is feasible and subject to other constraints. (Goldman 1999, p. 279)

And in a more elaborate way, Koehler and Shaviro (1990) write:

> Verdict accuracy is one of the principal goals of the trial process. Even in the absence of separte policy concerns that infuence the conduct of trials, however, accuracy cannot be guaranteed. Gaps and mistakes in fact-finding inevitably will occur in some cases and thus lead to inaccurate verdicts. Given this problem, along with the lack of a truth criterion even after trial, the best that can be accomplished in relation to verdict accuracy is to minimize the number of inaccurate verdicts that one reasonably expects. (Koehler and Shaviro 1990, p. 250)

In adducing considerations of accuracy to legal fact-finding, it is important not to commit a confusion. As stated explicitly in Koehler and Shaviro (1990), accuracy is attributed to verdicts or findings of facts that are formal decisions made by the fact-finders. In the context of this chapter, verdicts can be understood as public judgements. This type of accuracy might be called *verdict accuracy*. But accuracy can well be attributed to fact-finders' doxastic attitudes towards factual hypotheses disputed in courts. Since the basic doxastic attitudes we focus on are fact-finders' credences, we might call this type of accuracy *credence accuracy*. Both types of accuracy may be understood as *cognitive* or *epistemic* values attached, respectively, to decisions like verdicts and to belief states like credences.

Typically, verdicts as findings for the plaintiff or defendant are based on the fact-finder's credences in parties' factual hypotheses. But it does not mean that the accuracy of credences is always correlated with the accuracy of verdicts. It is perfectly possible that credences that maximize accuracy have no impact on the accuracy of verdicts. The first reason for thinking this is so is that an investigation may lead to a more accurate credence in some proposition, even if it does not lead to judging this proposition as true. The second reason is that credence and verdict accuracy are importantly different when we think of them as measurable cognitive values. That is, these quantities are measured by using different scales. Consider the following observation made by Ho (2008):

> 'Accuracy' is a measure of proximity to the truth. The closer the estimate is to the real age or weight, the nearer it approaches correctness. But a positive finding by the court carries a categorical assertion. It is not an estimate. It is either true or false (Ho 2008, p. 66)

Ho's point may be explained as follows: the accuracy of verdicts neither increases nor decreases, but verdicts are either accurate or inaccurate. However, Ho's observation does not preclude a gradual notion of accuracy in legal proceedings. Moreover, there seems to be a natural place to locate this notion within legal context. That is, while verdict accuracy is measured on a *categorical* scale, credence accuracy can be measured on a *gradational* scale. A measure of credence accuracy tells us how "distant" credences are to the truth-values of factual hypotheses: accuracy increases with the fact-finder's credences in truths and decreases with her credences in falsehoods. A measure of verdict accuracy tells us about the cognitive consequence of a verdict, given the truth of a factual hypothesis: if a verdict is for the plaintiff and the facts are such that the plaintiff deserves to win, then the verdict is accurate. And if the verdict is for the plaintiff and the facts are such that the defendant deserves to win, then the verdict is inaccurate. Thus, verdict accuracy neither increases nor decreases. However, this is not to say that the accuracy of a verdict for the plaintiff when the plaintiff deserves to win is as valuable as the accuracy of a verdict for the defendant when the defendant deserves to win. Likewise, the accuracy of a verdict for the plaintiff when the defendant deserves to win may not be as valuable as the accuracy of a verdict for the defendant when the plaintiff deserves to win. Typically, in criminal cases these accuracies are not equally valuable (or equivalently, the *in*accuracies are not equally regrettable). The most famous illustration of this thought is Blackstone's maxim which says that "it is better that ten guilty persons escape than that one innocent should suffer" (Blackstone 2002, p. 358). But civil litigation is importantly different: the situation of the parties is symmetrical. That is, the parties at civil litigation are equal before the law: it is as easy for the plaintiff to prove her case as it is for the defendant to disprove it. So the verdict accuracies should be equally valuable. Since this chapter deals only with civil litigation, I will consider, in section 5.5, a model in which these accuracies are equally valuable.

Although accuracy might be regarded as a fundamental goal of trial process, we need to emphasize that it is neither ultimate nor even fully attainable in all models of legal trial. It is not ultimate, since legal trials also aim at achieving justice, reaching evidence-responsive verdicts, protecting defendant's rights, or procuring and preserving public acceptance of verdicts. It seems plausible that some of these goals can be reconciled with accuracy. Just consider procedural justice that seems to be important for any legal adjudication. We may say that a legal decision is procedurally just if it satisfies some standard for procedural rightness. Now, it seems that one appropriate standard is the accuracy standard. That is, one may argue that a legal decision is procedurally just to the extent that it promotes accurate judgements (see Goldman 1999, chapter 9). However, some of these goals may stand in marked tension with accuracy. Just consider the rule, common in legal systems, according to which an accused person cannot be compelled to testify, even if his testimony could maximize verdict accuracy. Here accuracy is suppressed in order to protect the defendant's rights. Thus, if accuracy does not have a merit that elevates it over the other goals, legal trials would typically aim at striking the optimal balance between accuracy and these other objectives.

Also, accuracy may not even be fully attainable in all models of legal trial. Crucial in this context is the distinction between *inquisitorial* (dominant in the civil law tradition) and *adversarial* (dominant in the common law tradition) models of legal trial.[12] In the former model, a neutral fact-finder undertakes the task of managing the legal process, developing and presenting the evidence. The fact-finder, a neutral inquisitor, is motivated solely by the goal of accuracy rather than by parties' interests in winning. In the latter model, it is parties not the fact-finder that control investigation and presentation of the evidence. Parties,

[12]For a more elaborate analysis of these two models, see Damaska (1975).

motivated by a desire to win rather than a quest for approximating the truth, present their evidence and versions of a case. The fact-finder plays a passive role in receiving the evidence and evaluating parties' versions of the case. Unlike the inquisitorial model, the adversary model is prone to suppress accuracy for the sake of a definite winner in a courtroom battle. Clearly, the adversarial model has accuracy-obstructing features: for strategic reasons, each adversary may withhold evidence that provides strong support for a given hypothesis.

In sum, accuracy is one of the fundamental goals of legal trial. But it also appears to be a multi-faceted notion. I have distinguished two senses in which accuracy might be considered in legal settings: these are verdict and credence accuracy. Also, accuracy should not be regarded as the ultimate goal of legal trial. There are models of legal trial that suppress accuracy for the sake of different law-specific values.

5.5 A Bayesian Fact-Finder and Verdict Accuracy

How could the idea of verdict accuracy be represented in the Bayesian model given in section 5.3? In this section, I extend this model by adding to it a simple decision-theoretic component whose basic idea is that a Bayesian fact-finder ought to maximize subjective expected verdict accuracy. Thus, in this extended framework, the Bayesian fact-finder not only adopts credences, but also makes judgements or finds for the parties under uncertain circumstances.

With this extended Bayesian model in mind, I first show how the civil standard of proof, called the proof by a preponderance of the evidence, can be supported by appealing to the idea of expected verdict accuracy maximization. Second, I show that the Bayesian fact-finder's credences informed by chances in the manner prescribed by EP cannot result in a decrease of her expected verdict accuracy.

5.5.1 Verdict Accuracy

Our extended Bayesian model employs the following assumptions:

- $\mathcal{F}$ is a finite set of factual hypotheses describing possible states of the world.

The elements of $\mathcal{F}$ are mutually exclusive and jointly exhaustive.

- $cr' : \mathcal{F} \to [0,1]$ is the credence function of a Bayesian fact-finder with total evidence E. It takes each member of $\mathcal{F}$ and returns a credence $cr'(X)$.

- $\mathcal{V}$ is a finite set of propositions describing verdicts that are understood as public judgements.

- a is a verdict accuracy function. It takes each conjunction $V \wedge X$, where $V \in \mathcal{V}$ and $X \in \mathcal{F}$, and returns the accuracy $a(V \wedge X)$ of the verdict V if X describes the true state of the world. More generally, $a(V \wedge X)$ captures the cognitive or epistemic value of each conjunction $V \wedge X$. Conjunctions of the form $V \wedge X$ may be called outcomes relative to $\mathcal{V}$ and $\mathcal{F}$.

- the Bayesian fact-finder with total evidence E is an expected verdict accuracy maximizer. That is, she judges V that maximizes her expected accuracy relative to cr' and a given by

$$\mathrm{Exp}_{a,cr'}(V) = \sum_{X \in \mathcal{F}} cr'(X)a(V \wedge X). \tag{5.5}$$

  That is, the subjective expected accuracy of judging V is a weighted sum of its accuracies in every possible state of the world, where the weight assigned to a particular state is given by the credence function cr'.

The extended Bayesian model of legal fact-finding just given can be used to evaluate acts understood as fact-finder's verdicts or public judgements. As this model assumes, verdicts can be evaluated in terms of their cognitive or epistemic consequences. That is, it is assumed that the fact-finder has a verdict accuracy function that assigns an epistemic value called verdict accuracy for each outcome $V \wedge X$. This epistemic value depends only on (i) which factual hypothesis is true and (ii) what the fact-finder's verdict is. Accuracy, so understood, has nothing to do with practical gain or cost. Moreover, it is attached to the act of judging, and not to the act endorsing a particular belief state.

For concreteness, recall the Blue Bus. In this case, $\mathcal{F} = \{P, \neg P\}$, where P describes a state of the world in which a bus belonging to the Blue Bus Company caused the accident, and $\neg P$ describes a state of the world in which a bus belonging to the Blue Bus Company did not cause the accident. The set of actions that the fact-finder considers is given by $\mathcal{V} = \{V_p, V_{\neg p}\}$, where V_p is a verdict for the plaintiff and $V_{\neg p}$ is a verdict for the defendant. Then, the accuracies assigned to the fact-finder's possible judgements are given by:

- $a(V_p \wedge P)$ is the accuracy of a verdict for the plaintiff when a bus belonging to the Blue Bus Company caused the accident.
- $a(V_p \wedge \neg P)$ is the accuracy of a verdict for the plaintiff when a bus belonging the Blue Bus Company did not cause the accident.
- $a(V_{\neg p} \wedge \neg P)$ is the accuracy of a verdict for the defendant when a bus belonging the Blue Bus Company did not cause the accident.
- $a(V_{\neg p} \wedge P)$ is the accuracy of a verdict for the defendant when a bus belonging to the Blue Bus Company caused the accident.

It is important to emphasize that the accuracy of a verdict for the plaintiff when the plaintiff deserves to win is as valuable as the accuracy of a verdict for the defendant when the defendant deserves to win. Likewise, the accuracy of a verdict for the plaintiff when the defendant deserves to win is as valuable as the accuracy of a verdict for the defendant when the plaintiff deserves to win. The main reason for thinking that these accuracies are equally valuable is that the plaintiff and the defendant in a civil case have an equal stake in the proceedings.

Now, in order to calculate the expected accuracy of V_p, the fact-finder needs to weigh the epistemic risk of judging V_p when P is false against the epistemic gain of judging V_p when P is true. Similarly, to calculate the expected accuracy of $V_{\neg p}$ she has to weigh the epistemic risk of judging $V_{\neg p}$ when P is true against the epistemic gain of judging $V_{\neg p}$ when P is false.

5.5.2 Preponderance of the Evidence and Verdict Accuracy

Legal fact-finding is a rule-governed process. In the tradition of common law, one important category of rules governing legal fact-finding are standards of proof. Typically, a standard of proof is understood as a requirement that a party ought to satisfy in order to win its case. The standard of proof in civil cases says that in order to win the plaintiff has to establish her case by the preponderance of the evidence. Roughly, it means that the plaintiff, who bears the burden of proof, must prove that her case is more probable than not.

There are at least two robust traditions in analysing the civil standard of proof: the *external* and the *internal* analysis. According to the external analysis, the civil standard of proof is a decisional threshold imposed on the fact-finder's credences at the end-point stage of the process of evaluating a given factual hypothesis in the light of evidence. If the fact-finder's credence in P meets this threshold, she must accept P; otherwise, she must reject P. According to the internal analysis, the civil standard of proof is an instruction on the fact-finder's deliberative attitude in the whole process of evidential evaluation. On this view, the civil standard is not a decisional threshold, but rather a standard of caution which reflects resistance to persuasion on the truth of a factual hypothesis.[13] In order to fit in with the Bayesian model of legal fact-finding given in this chapter, I will focus on the external analysis of the civil standard of proof.

In our Bayesian model, the civil standard of proof can be introduced as follows:

> **Preponderance of The Evidence (PE):** The plaintiff's factual hypothesis P meets the preponderance of the evidence standard just in case the fact-finder's credence in P is
>
> $$cr'(P) > \frac{1}{2},$$
>
> where $cr'(P) = cr(P|E)$ and E is the total evidence presented at a legal trial.

That is, the plaintiff's factual hypothesis P satisfies the preponderance of the evidence standard just in case the fact-finder's end-point credence in P conditional of the total evidence E is greater than $\frac{1}{2}$.

Why should we believe that PE imposes the correct decisional threshold on the fact-finder's credence in P? Interestingly, it can be shown that PE follows from the injunction to maximize subjective expected verdict accuracy. That is,

[13] For a thorough analysis of external and internal approaches to the standards of proof, see Ho (2008, chapter 4).

if the fact-finder is a subjective expected accuracy maximizer, then the threshold imposed on $cr'(P)$ that warrants acceptance of P is the one given by PE.

To show this, first notice that the fact-finder who maximizes her expected verdict accuracy will judge V_p if

$$\mathrm{Exp}_{a,cr'}(V_p) > \mathrm{Exp}_{a,cr'}(V_{\neg p}). \tag{5.6}$$

By expanding inequality (5.6), we get

$$cr'(P)a(V_p \wedge P) + cr'(\neg P)a(V_p \wedge \neg P) > cr'(P)a(V_{\neg p} \wedge P) + cr'(\neg P)a(V_{\neg p} \wedge \neg P). \tag{5.7}$$

By rearranging the terms, inequality (5.7) comes down to

$$\frac{cr'(P)}{cr'(\neg P)} > \frac{a(V_{\neg p} \wedge \neg P) - a(V_p \wedge \neg P)}{a(V_p \wedge P) - a(V_{\neg p} \wedge P)}. \tag{5.8}$$

Notice that the numerator of the right side of inequality (5.8) reflects a difference between the accuracy of a right verdict for the defendant and the accuracy of a wrong verdict for the plaintiff. Similarly, the denominator of the right side of inequality (5.8) reflects a difference between the accuracy of a right verdict for the plaintiff and the accuracy of a wrong verdict for the defendant. I assume that the ratio of these differences has a positive value. Now, after noting that $cr'(\neg P) = 1 - cr'(P)$, inequality (5.8) boils down to

$$cr'(P) > \frac{1}{1 + \frac{a(V_p \wedge P) - a(V_{\neg p} \wedge P)}{a(V_{\neg p} \wedge \neg P) - a(V_p \wedge \neg P)}}. \tag{5.9}$$

Since we deal with a civil case, and so $a(V_p \wedge P) = a(V_{\neg p} \wedge \neg P)$ and $a(V_{\neg p} \wedge P) = a(V_p \wedge \neg P)$, we have that

$$\frac{a(V_p \wedge P) - a(V_{\neg p} \wedge P)}{a(V_{\neg p} \wedge \neg P) - a(V_p \wedge \neg P)} = 1, \tag{5.10}$$

and so, by inequality (5.9), we get

$$cr'(P) > \frac{1}{2}, \tag{5.11}$$

as required.

Note, however, that the result given above holds if (i) the accuracy of a right verdict for the plaintiff is equal to the accuracy of a right verdict for the defendant, and (ii) the accuracy of a wrong verdict for the plaintiff is equal to the accuracy of a wrong verdict of the defendant. If we drop the assumption and allow for some accuracies to be greater than other, a different decisional threshold may follow from the injunction to maximize subjective expected verdict accuracy. This stems from the fact that, as shown in inequality (5.9), $cr'(P)$ is a function of verdict accuracies.

5.5.3 Verdict Accuracy and End-Point Chance-Credence Principle

How could the idea of expected verdict accuracy maximization support the use of naked statistical evidence in courts? In this subsection, I show that if the fact-finder coordinates her credences with chances in accordance with EP, then her expected verdict accuracy after such coordination never decreases, and could increase. The result to be given may be understood as follows: the act of judging (reaching a verdict) before coordinating credences with chances cannot have greater expected accuracy than the act of judging after coordinating credences with chances.

To begin with, suppose that the fact-finder with total evidence E has to decide between two options: either ignore information about chance and judge, or coordinate her credences with chances in accordance with EP and then judge. Importantly, the fact-finder values these two options from her current perspective, i.e. relative to her end-point credence function cr' and to a finite partition of possible chance hypotheses, $\{C_{ch_1}, ..., C_{ch_n}\}$, with $cr'(C_{ch_i} \wedge C_{ch_j}) = 0$ for all $i \neq j$, and $cr'(C_{ch_1} \vee ... \vee C_{ch_n}) = 1$.

In order to compare these two options, let us first consider a situation in which the fact-finder decides to coordinate her credences with chances in accordance with EP, and then judges verdict V. That is, she sets her credences in such a way that, for all ch, $cr'(X|C_{ch}) = ch(X|E)$, and then judges V. Since she is an expected verdict accuracy maximizer, she judges V with the value given by

$$\max_{V\in\mathcal{V}}\sum_{X\in\mathcal{F}} cr'(X|C_{ch})a(V\wedge X)=\max_{V\in\mathcal{V}}\sum_{X\in\mathcal{F}} ch(X|E)a(V\wedge X). \tag{5.12}$$

But she values option (5.12) from her current perspective, and hence her expectation of (5.12) relative to cr' is given by

$$\sum_{ch} cr'(C_{ch})\max_{V\in\mathcal{V}}\sum_{X\in\mathcal{F}} ch(X|E)a(V\wedge X), \tag{5.13}$$

which is the prior expectation of the act of judging V that has the highest expected verdict accuracy after coordinating credences with chances.

Now, consider a situation in which the fact-finder decides to judge V before coordinating her credences with chances. Again, since she is an expected accuracy maximizer, she judges V that has the value given by

$$\max_{V\in\mathcal{V}}\sum_{X\in\mathcal{F}} cr'(X)a(V\wedge X). \tag{5.14}$$

By probability theory and EP, (5.14) comes down to

$$\begin{aligned}\max_{V\in\mathcal{V}}\sum_{X\in\mathcal{F}} cr'\left(\bigvee_{ch}(X\wedge C_{ch})\right)a(V\wedge X) &= \max_{V\in\mathcal{V}}\sum_{X\in\mathcal{F}}\sum_{ch} cr'(X\wedge C_{ch})a(V\wedge X)\\ &=\max_{V\in\mathcal{V}}\sum_{X\in\mathcal{F}}\sum_{ch} cr'(X|C_{ch})cr'(C_{ch})a(V\wedge X)\\ &=\max_{V\in\mathcal{V}}\sum_{ch}\sum_{X\in\mathcal{F}} ch(X|E)cr'(C_{ch})a(V\wedge X),\end{aligned} \tag{5.15}$$

which is the maximum of the prior expectation of the expected verdict accuracy after coordinating credences with chances.

Importantly, it can be shown that (5.13) is greater or equal to (5.15). To show this, let us use the following mathematical result:

Jensen's inequality: For any random variable X and any convex function ϕ,

$$\mathrm{Exp}\,[\phi(X)] \geq \phi(\mathrm{Exp}\,[X]),$$

where Exp stands for expectation.

With this inequality in mind, let us write

$$\phi(V) = \max_{V \in \mathcal{V}} \sum_{X \in \mathcal{F}} ch(X|E)a(V \wedge X), \tag{5.16}$$

where ϕ is a convex function of V. Also, we can write

$$\mathrm{Exp}\,[\phi(V)] = \sum_{ch} cr'(C_{ch}) \max_{V \in \mathcal{V}} \sum_{X \in \mathcal{F}} ch(X|E)a(V \wedge X), \tag{5.17}$$

and

$$\phi(\mathrm{Exp}\,[V]) = \max_{V \in \mathcal{V}} \sum_{ch} \sum_{X \in \mathcal{F}} ch(X|E)cr'(C_{ch})a(V \wedge X). \tag{5.18}$$

Then, it follows, by Jensen's inequality, that

$$\begin{aligned} &\sum_{ch} cr'(C_{ch}) \max_{V \in \mathcal{V}} \sum_{X \in \mathcal{F}} ch(X|E)a(V \wedge X) \geq \\ &\max_{V \in \mathcal{V}} \sum_{ch} \sum_{X \in \mathcal{F}} ch(X|E)cr'(C_{ch})a(V \wedge X), \end{aligned} \tag{5.19}$$

as required. This result shows that the expected verdict accuracy after coordinating credences with chances is not less than the expected verdict accuracy before coordinating credences with chances, and may be greater. So if we consider the consequences of fact-finder's credences on her verdicts, then EP cannot lead the fact-finder to expect to reach less accurate verdicts, and may lead her to expect to reach more accurate ones. Since EP captures a way of using naked statistical evidence in courts of law, one may understand this result as saying that expecting this sort of evidence at trial should not lead the fact-finder to get less accurate verdicts.

The argument just given needs qualifying in a number of ways. First, whether or not the fact-finder coordinates her credences with chances, the elements of $\mathcal{F}$, $\mathcal{V}$, and the accuracies assigned to conjunctions of the form $V \wedge X$ stay unchanged.

That is, the fact of coordinating credences with chances does not by itself alter the states of the world nor the verdicts. Second, the fact of coordinating credences with chances does not by itself alter the fact-finder's credences about the states of the world. But setting credences equal to chances in accordance with EP would typically alter the fact-finder's credence about the states of the world. Third, the act of judging does not give any information about the truth of factual hypotheses. Third, the argument concerns the fact-finder who expects or plans to obey EP, and not the fact-finder who actually satisfies this constraint. The result does not say that information about chance is never harmful to the accuracy of verdicts: it only says that this information is never harmful with respect to the fact-finder's prior expectation.

5.6 A Bayesian Fact-Finder and Credence Accuracy

Naked statistical evidence can affect both verdict and credence accuracy. In the previous section, I have shown how the idea of expected verdict accuracy maximization lends support to the use of naked statistical evidence in courts of law. In this section, I explore a connection between the use of naked statistical evidence and the idea of expected credence accuracy maximization. First, I show how the idea of credence accuracy maximization could be represented precisely in our Bayesian model of legal fact-finding. To this end, I employ the resources of epistemic utility theory, developed by James M. Joyce (1998) and Richard Pettigrew (2013a; 2013c). Second, I present an argument showing that a Bayesian fact-finder's credences that match chances in accordance with EP minimize objective expected inaccuracy (or equivalently, maximize objective expected accuracy). More precisely, the result shows that the Bayesian fact-finder's credences that obey EP minimize objective expected inaccuracy relative to any possible chance function and a proper inaccuracy measure. So when the Bayesian fact-finder is asked what she should do with information about chance, this argument shows that EP is an optimal response for her.

5.6.1 Credence Accuracy

According to the epistemic utility theory, the accuracy of cr' at some possible world w is its "distance" from the truth at w. Accuracy, so understood, is a measurable property: an accuracy measure tells us how close cr' is to the "ideal" credence function at w. And the ideal credence function at w, denoted by v_w, is the one that assigns credence 1 to propositions that are true at w, and credence 0 to propositions that are false at w. Following Joyce (1998), I will talk about the inaccuracy measure and will treat the accuracy measure as the negative of the inaccuracy measure.

The idea of inaccuracy measure then can be understood more precisely as follows. For our Bayesian fact-finder, we consider two possible worlds: w_1 in which P is true (and hence $\neg P$ is false), and w_2 in which P is false (and hence $\neg P$ is true). Then, for each of these worlds, the inaccuracy of her credence function cr' at that world is some function of the values that cr' assigns to P and $\neg P$ and the truth values of P and $\neg P$ at that world. More precisely, the inaccuracy of cr' at world w, denoted by $I(cr', w)$, is given by

$$I(cr', w) = \sum_{X \in \mathcal{F}} s(cr'(X), v_w(X)), \tag{5.20}$$

where $s : [0,1] \times \{0,1\} \to [0, \infty]$ is a scoring rule. It is important to notice that while $s(cr'(X), v_w(X))$ measures the divergence of a particular credence $cr'(X)$ from a particular truth value $v_w(X)$ at w, $\sum_{X \in \mathcal{F}} s(cr'(X), v_w(X))$ measures the divergence of a credence function cr' over $\mathcal{F}$ from a truth-value distribution v_w over $\mathcal{F}$ at w. That is, while the former measures the "local" inaccuracy, the latter measures the "global" inaccuracy. In what follows, I will be concerned only with the measure of "global" inaccuracy.

It appears that the idea of inaccuracy measure captures crucial features of legal fact-finding. Fact-finders adopt doxastic attitudes towards factual hypotheses discussed in court under circumstances of uncertainty: evidence presented in court does no provide conclusive reasons for believing or disbelieving hypotheses. Rather, when evidence accumulates during the trial the fact-finder's degree of uncertainty fluctuates. Along that, the accuracy of these degrees increases or decreases. It might well be that these degrees of uncertainty do not reach the categorical level of belief or disbelief. Rather, if a fact-finder's degree of uncertainty in a proposition reaches some appropriate level, the substantive law tells us that the fact-finder should accept that proposition, by appealing to the legal standards of proof.

Since there are many plausible inaccuracy measures, I will focus only on a particular class of these measures called *proper* inaccuracy measures. To introduce these measures, let me first define the notion of expected inaccuracy of a credence function cr'' relative to the inaccuracy measure I and the credence function cr'. Suppose that $\{w\}$ is the singleton proposition that is true only at world w, and let $\mathcal{W}$ be a finite set of possible worlds. Then, the expected inaccuracy of cr'' relative to I and cr' can be defined as follows:

$$\mathrm{Exp}_{I,cr'}(cr'') = \sum_{w \in \mathcal{W}} cr'(\{w\})I(cr'', w). \tag{5.21}$$

That is, the expected inaccuracy of cr'' relative to I and cr' is the sum of its inaccuracies at each possible world w, weighted by the credence cr' assigned to that world. With this notion in mind, we can define the proper inaccuracy measure as follows:

> **Propriety:** An inaccuracy measure I is proper if, for any cr' and cr'',
>
> $$\mathrm{Exp}_{I,cr'}(cr') \leq \mathrm{Exp}_{I,cr'}(cr'').$$

That is, I is proper if cr' does not expect that any other cr'' is better at minimizing inaccuracy. The rationale behind this condition is obvious: a fact-finder's ought not to have a credence function that expects itself to be worse at minimizing inaccuracy than it expects any other credence function to be. Someone who holds credences that expect themselves to be epistemically inferior to some other credences undermines her own epistemic state, for she expects that she could do better. Moreover, credences scored by a proper inaccuracy measure are protected from unjustified changes. For if I were not a proper inaccuracy measure, the fact-finder could change, whenever she wants, her end-point credence function cr' to some cr'' such that $\mathrm{Exp}_{I,cr'}(cr'') < \mathrm{Exp}_{I,cr'}(cr')$.

Armed with these notions, my task is to provide an answer to the following question: how could the idea of credence accuracy be utilized to argue for EP? But before doing so, I will briefly show how the idea of credence accuracy can be used to argue for Probabilism and Bayesian Conditionalization—the two constraints that characterize the minimal Bayesian model of legal fact-finding introduced in subsection 5.3.1.

As it has been argued in Joyce (1998; 2009), we can derive Probabilism by appealing to the idea of credence accuracy and a particular dominance norm which says that one should not have a credence function that is accuracy-dominated by some other credence function, which itself cannot be accuracy-dominated. The crucial result of Joyce's shows that, for a certain class of inaccuracy measures, (i) if one's credence function violates Probabilism, then there is another credence function that satisfies Probabilism and is strictly less inaccurate in every possible world, and (ii) if one's credence function satisfies Probabilism, then there is no other credence function that is less inaccurate in at least one possible world. By combining this result with the dominance norm, Joyce concludes that one's credence function should obey Probabilism, on pain of being accuracy-dominated.

Similarly, Hilary Graves and David Wallace (2006) show how Bayesian Conditionalization follows from the idea of credence accuracy and a norm which says that one should have a credence function that maximizes expected accuracy given a situation in which one is to receive a piece of information. Their crucial result is that if an agent updates her credences by any rule other than Bayesian Conditionalization, she does not maximize her expected accuracy relative to her current credence function. That is, she expects that credences updated by some other rule are less accurate than credences updated by Bayesian Conditionalization. They conclude that any expected accuracy maximizer should obey Bayesian Conditionalization.

In what follows, I give an argument for EP that also appeals to the idea of credence accuracy, and utilizes a particular norm which says what credences the fact-finder should have.

5.6.2 Credence Accuracy and End-Point Chance-Credence Principle

To begin with our credence accuracy-based argument for EP, let us define the notion of objective expected inaccuracy of cr' relative to an inaccuracy measure I and any conditional chance function $ch(\cdot|E)$ as follows:

$$\mathrm{Exp}_{I,ch(\cdot|E)}(cr') = \sum_{w\in\mathcal{W}} ch(\{w\}\,|E)I(cr', w). \tag{5.22}$$

That is, the objective expected inaccuracy of a fact-finder's credence at the end-point stage of a legal trial is the sum of its accuracies at each possible world, weighted by the chance function conditional on the total evidence. Note that the expectation is objective, since it is calculated relative to the chance of $\{w\}$ conditional on E.

Importantly, it does not seem right to think that every chance of $\{w\}$ conditional on E can be used to calculate the objective expected inaccuracy of cr'. As discussed in subsection 5.2.4, even if we endorse the view that chances are always relativized to the total evidence, we still need to decide which of them are well suited for calculating the objective expected inaccuracy. By way of suggesting, we may require such chances to be sufficiently resilient in the sense that no further evidence could alter their values. It also seems true that such chances should be conditioned on sufficiently specific total evidence.

Now, we can introduce the following norm:

> **Minimizing Objective Expected Inaccuracy:** A Bayesian fact-finder with total evidence E ought to have a credence function cr' such that for any other cr'' and all possible ch,
>
> $$\mathrm{Exp}_{I,ch(\cdot|E)}(cr') \leq \mathrm{Exp}_{I,ch(\cdot|E)}(cr'').$$

This norm requires the Bayesian fact-finder with total evidence E not to adopt a credence function that is objectively expected to be worse at minimizing inaccuracy than some other credence function. In other words, the fact-finder who adopts credences that do not minimize objective expected inaccuracy is irrational.

This norm appears to be fairly plausible: if the chance function ch conditional on E "expects" cr'' to have a lower inaccuracy than it "expects" cr' to have, then the fact-finder should not adopt cr' as her credence function.

To run the credence accuracy-based argument for EP, I need first to emphasize that EP is not the only answer to the question of how credences should be coordinated with chances. Clearly, there are other chance-credence principles that could tell us how to coordinate credences with chances. Just consider the following:

No-Chance-Credence Principle (NP): At the end-point stage of legal trial, the fact-finder with total evidence E ought to have a credence function such that for all $X \in \mathcal{F}$ and all ch,

$$cr'(X|C_{ch}) = cr'(X).$$

Anti-Chance-Credence Principle (AP): At the end-point stage of legal trial, the fact-finder with total evidence E ought to have a credence function such that for all $X \in \mathcal{F}$ and all ch,

$$cr'(X|C_{ch}) = 1 - ch(X|E),$$

provided $ch(E) > 0$.[14]

The two principles do not recommend the fact-finder to set her credences equal to chances. While NP requires the Bayesian fact-finder to ignore the information about chance and stick to her credences, AP recommends to set her credence in X equal to the difference between the truth value 1 of X and the chance of X. Thus, their recommendations are different from the one given by EP. But which one of these principles should govern the fact-finder's credences?

If we assume that the fact-finder adopts credences in accordance with Minimizing Objective Expected Inaccuracy, then it is EP that should govern her credences. To show this, suppose that the accuracy of her credences is scored by a proper scoring rule I. Further, suppose that $cr'(\cdot|C_{ch}) = ch(\cdot|E)$ and

[14] This policy is a slight reformulation of the Anti-Principal Principle presented in Loewer (2004).

$cr''(\cdot|C_{ch}) = ch^*(\cdot|E)$, where $ch(\cdot|E) \neq ch^*(\cdot|E)$. In particular, $cr''(\cdot|C_{ch})$ might obey NP, and so $ch^*(\cdot|E) = cr''(\cdot)$. Then, since I is proper, we have that, for all possible ch,

$$\text{Exp}_{I,ch(\cdot|E)}(ch(\cdot|E)) \leq \text{Exp}_{I,ch(\cdot|E)}(ch^*(\cdot|E)). \tag{5.23}$$

And by the assumption that $cr'(\cdot|C_{ch}) = ch(\cdot|E)$ and $cr''(\cdot|C_{ch}) = ch^*(\cdot|E)$, we get

$$\text{Exp}_{I,ch(\cdot|E)}(cr'(\cdot|C_{ch})) \leq \text{Exp}_{I,ch(\cdot|E)}(cr''(\cdot|C_{ch})). \tag{5.24}$$

But the left-hand side of (5.24) is the objective expected inaccuracy of a credence function that obeys EP, and the right-hand side is the objective expected inaccuracy of a credence function that obeys any other chance-credence principle. Hence, we conclude that EP is a chance-credence principle that recommends credence functions that are optimal with respect to objective expected inaccuracy. In particular, if the fact-finder were to ignore information about chance in accordance with NP, she would adopt credences that do not minimize objective expected inaccuracy.

To bring this argument home, recall the Blue Bus. For concreteness, let us apply a particular proper inaccuracy measure called the Brier score. In the context of this chapter, the Brier score can be presented as follows:

Brier score: The inaccuracy I of credence function cr' at world w is

$$I(cr', w) = \frac{1}{|\mathcal{F}|} \sum_{X \in \mathcal{F}} (cr'(X) - v_w(X))^2.$$

That is, the Brier score measures the inaccuracy of cr' at w by taking a weighted average of the sum over the squared differences between credences and truth-values assigned to each proposition in $\mathcal{F}$ at w.

In the Blue Bus, there are two relevant worlds: world w_B at which a bus belonging to the Blue Bus Company caused the accident, and world w_R at which a bus belonging to the Red Bus Company caused the accident. Suppose that $cr'(\cdot|C_{ch})$ obeys EP and, for simplicity, assume that E is a tautological proposition. Then, $cr'(\{w_B\}\,|C_{ch}) = ch(\{w_B\}\,|E) = 0.8$ and $cr'(\{w_R\}\,|C_{ch}) = ch(\{w_R\}\,|E) = 0.2$. Further, let us stipulate that $cr''(\{w_B\}\,|C_{ch}) = cr''(\{w_R\}) =$

0.5 and $cr''(\{w_R\}\,|C_{ch}) = cr''(\{w_R\}) = 0.5$. Thus, $cr''(\cdot|C_{ch})$ obeys NP. Then,

$$\begin{aligned}
\mathrm{Exp}_{I,ch(\cdot|E)}(cr'(\cdot|C_{ch})) &= ch(\{w_B\}\,|E)I(ch(\cdot|E), w_B) \\
&+\ ch(\{w_R\}\,|E)I(ch(\cdot|E), w_R) \\
&= 0.8 \times 0.5\left[(0.8-1)^2 + (0.2-0)^2\right] \\
&+\ 0.2 \times 0.5\left[(0.8-0)^2 + (0.2-1)^2\right] \\
&= 0.16. \qquad (5.25)
\end{aligned}$$

And,

$$\begin{aligned}
\mathrm{Exp}_{I,ch(\cdot|E)}(cr''(\cdot|C_{ch})) &= ch(\{w_B\}\,|E)I(cr'', w_B) \\
&+\ ch(\{w_R\}\,|E)I(cr'', w_R) \\
&= 0.8 \times 0.5\left[(0.5-1)^2 + (0.5-0)^2\right] \\
&+\ 0.2 \times 0.5\left[(0.5-0)^2 + (0.5-1)^2\right] \\
&= 0.25. \qquad (5.26)
\end{aligned}$$

Since $\mathrm{Exp}_{I,ch(\cdot|E)}(cr'(\cdot|C_{ch})) < \mathrm{Exp}_{I,ch(\cdot|E)}(cr''(\cdot|C_{ch}))$, the Bayesian fact-finder, who minimizes objective expected inaccuracy, will choose to coordinate her credences with chances in accordance with EP, rather than in accordance with the alternative principle NP.

5.7 Conclusions

In this chapter, I have shown how the use of naked statistical evidence might be valuable in legal settings. In the first part of this chapter, I have criticized those views, according to which there is some fundamental difference between naked statistical evidence and individualized evidence, and the existence of this difference explains why legal fact-finders are so reluctant to rely on naked statistical evidence.

In the second part of this chapter, I have developed a connection between accuracy and the use of naked statistical evidence in legal fact-finding. I have extricated two possible ways of understanding the notion of accuracy in legal fact-finding, to wit, verdict accuracy and credence accuracy. These notions have been then formalized within a simple Bayesian model of legal fact-finding. Also, within this model, I have formulated a particular constraint on the fact-finder's credences that requires them to match chances. I have shown that with this constraint in place we can capture an interesting way of using exclusively statistical evidence in courts of law. As it has been shown in sections 5.5 and 5.6, the Bayesian model of legal fact-finding can be used to provide accuracy-based arguments for the claim that the fact-finder's credences should line up with chances. The linchpin of these arguments is that a particular chance-credence principles could always be conducive to the achievement of both credence and verdict accuracy. Hence, there is something epistemically defective about the fact-finder's credences that do not line up with chances.

Importantly, I have not claimed that the Bayesian model *describes* adequately the process of legal fact-finding. As it is well known, there are many difficult problems associated with the use of Bayesian models in legal proof.[15] In particular, one might doubt whether jurors or judges assign precise numerical credences to factual hypotheses, and whether they actually employ policies like Bayesian Conditionalization and EP, or employ them correctly. Our Bayesian model, however, may be understood as a *regulative* ideal: insofar as the fact-finders depart from this model, this is to be taken as a departure from some ideal.

[15] A good survey of these conceptual problems is to be found in Allen (1997).

Chapter 6

Maximum Relative Entropy Updating and the Value of Learning

This chapter examines the possibility of justifying the principle of maximum relative entropy (MRE), considered as an updating rule, by looking at the value of learning theorem established in classical decision theory. This theorem captures an intuitive requirement for learning: learning should lead to new credences that are expected to be helpful, and never harmful in making decisions. This requirement is called *the value of learning*. The chapter analyses the extent to which learning ruled by MRE could satisfy this requirement, and so could be a rational means for pursuing practical goals. First, by representing MRE updating as a conditioning model, it is shown that MRE satisfies the value of learning in cases where learning prompts a complete redistribution of one's credences over a partition of propositions. Second, it is argued that the value of learning may

This chapter is based on the paper "Maximum Relative Entropy Updating and the Value of Learning" (Dziurosz-Serafinowicz 2015) that appeared in *Entropy*.

not be generally satisfied by MRE updates in cases of updating on a change in one's conditional credence. The analysis given in this chapter points towards a more general moral: that the justification of MRE updating in terms the value of learning may be sensitive to the context of a given learning experience. In addition, it lends support to the idea that MRE is not a universal or mechanical updating rule, but rather a rule whose application and justification may be context-sensitive.

The chapter covers one of the main themes of this thesis, to wit, the use of higher-order probabilities. Specifically, it exploits a particular condition which relates an agent's first- and second-order credences in the context of probabilistic updating.

6.1 Introduction

Let the functions cr and cr' represent, respectively, an agent's prior and posterior credence functions over an algebra of propositions $\mathcal{F}$ generated by a set of possible worlds $\mathcal{W}$. Assume that both functions are probability functions over $\mathcal{F}$. A rule for changing the agent's prior credence function cr over $\mathcal{F}$ in light of new evidence (hereafter, an updating rule) aims to provide an answer to the following problem: given cr and some constraint χ imposed on cr', which cr' should the agent choose from the set of her posterior credence functions that satisfy χ? A given constraint χ imposed on cr' is supposed to represent a learning experience, and we associate with every learning experience a set $\mathcal{C}_\chi$ of posterior credence functions singled out by χ, i.e. $\mathcal{C}_\chi = \{cr' : cr' \text{ satisfies } \chi\}$. We take it that $\mathcal{C}_\chi$ is a closed convex set, i.e. it is determined by a constraint χ such that if cr'_1 and cr'_2 satisfy χ, then also any convex combination of them, $\lambda cr'_1 + (1-\lambda)cr'_2$ with $\lambda \in [0,1]$, will satisfy χ. This type of constraint is called affine.

An updating rule that is subject to considerable discussion among philosophers is the principle of maximum relative entropy (MRE), also known as the rule of minimizing cross-entropy, the principle of minimum discrimination information, or Kullback–Leibler divergence. It says that, given P, the partition $\{S_i\}$ of minimal elements in $\mathcal{F}$, and some constraint χ on cr', the agent should choose cr', so as to satisfy χ, while minimizing the relative entropy with respect to cr

as measured by the following function:

$$\mathrm{RE}(cr, cr') = \sum_i cr'(S_i)\log\frac{cr'(S_i)}{cr(S_i)}. \tag{6.1}$$

That is, by MRE, an updater should adopt as her posterior credence function, from those defined over $\mathcal{F}$ and satisfying χ, the one that is RE-closest to her prior credence function defined over $\mathcal{F}$. RE thus can be seen as a measure of the "distance" between cr and the possible cr''s that satisfy χ. Additionally, RE $= 0$ just in case $cr = cr'$. Of course, RE is not a distance measure in the mathematical sense, for it is not symmetric.

Much of the controversy surrounding MRE concerns its status. At least four main views on this issue can be distinguished. According to the first view (Williams 1980), MRE is a generally valid rule of updating one's credences from which the two well-known conditionalization rules, to wit, Bayes's rule and Jeffrey's rule, can be derived. The second view denies the very idea of MRE's universal validity. Within this camp, some (Friedman and Shimony 1971; Shimony 1985; Seidenfeld 1986) argue that in certain situations, it conflicts with Bayes's rule; others (van Fraassen 1981; van Fraassen et al. 1986) argue that it leads to counter-intuitive consequences in the Judy Benjamin case, which is a case of updating on a conditional proposition; and some (Bradley 2005; Douven and Romeijn 2011) argue, quite generally, that MRE is just one of many updating rules and, as such, is applicable in the right circumstances. On the third view (Skyrms 1985), MRE can be regarded, under certain conditions, as a special case of Bayes's rule. Finally, on the fourth view (Skyrms 1987b), MRE is not a rule for updating one's credences, but rather a rule for statistical supposing. These views have their merits, although none have achieved widespread acceptance.

However, there is yet another foundational question concerning MRE, a question that might be posed independently of the aforementioned concern. This is the question of whether, and if so, how, MRE can be justified as a method of updating one's credences. Surprisingly, there have been relatively few attempts to answer this question. The most notable among them are Edwin T. Jaynes's (1957) attempt to show that this method gives a probability assignment that is maximally non-committal with regard to missing information, John E. Shore and

Rodney Johnson's (1980; 1981) justification by consistency, and Peter Grünwald's (2000) minimax decision-theoretic justification. In contrast, there are several existing justifications of the two most prominent updating rules, to wit, Bayes's rule and Jeffrey's rule. Bayes's rule is justified on the grounds that it is both a pragmatically and epistemically rational way of updating. The pragmatic rationality of this rule is established by the diachronic Dutch book argument, which shows that if you update your degrees of beliefs other than by Bayes's rule, then you are susceptible to a collection of bets ensuring a negative net pay-off, come what may (Teller 1973; Lewis 1999). Various accuracy-based arguments show that Bayes' rule is also epistemically rational. In particular, they show that Bayesian updating minimizes the expected inaccuracy (Leitgeb and Pettigrew 2010b). Similarly, various Dutch book arguments support Jeffrey's rule by establishing its pragmatic rationality (Skyrms 1987b).

The aim of this chapter is to examine the possibility of justifying MRE updating by linking it to the value of learning theorem, introduced to the philosophical literature by Leonard Savage (1954) and Irving J. Good (1967). The value of learning theorem may be viewed as capturing an intuitive requirement of rationality for learning. The requirement says that learning should lead to new degrees of belief that are expected to be helpful and never harmful in making decisions. Call this requirement *the value of learning*. The notion of rationality that it alludes to is essentially pragmatic: we consider whether an opinion shift ruled by MRE is rational for an agent who always chooses that act that maximizes her expected utility. However, as recently argued in Huttegger (2014), we can also think of the value of learning as a necessary requirement for one's opinion shift to count as genuine learning. Of course, on this view, there might be other features of genuine learning that are not captured by the value of learning, e.g. that genuine learning leads to more accurate credences. Therefore, it might not be a sufficient condition. Importantly, it has been shown that the value of learning holds for both Bayes's rule (Good 1967) and Jeffrey's rule (Graves 1989).

I show that updating by MRE satisfies the value of learning in cases where the constraint reporting one's learning experience concerns a complete redistribution of one's credences over a partition of propositions. My strategy will be to exploit a link between a particular generalized model of Bayesian conditioning and updating by MRE on a partition of propositions. The generalized model of conditioning allows us to assign second-order credences to propositions about first-order ones, and to condition the former on propositions concerning the latter. If we interpret the second-order credences as one's priors and the first-order ones as one's posteriors, then the model allows for conditioning prior credences on propositions about the posterior ones. In this set-up, we can represent, under certain conditions, updating by MRE on a partition as a form of conditioning on a proposition specifying posterior credences for each member of that partition. However, there are other types of constraints to which MRE updating can be applied. In particular, these might involve a constraint to the effect that one should assign a conditional posterior credence to some proposition, given another proposition. I show that whether or not MRE updating leads to the value of learning theorem in response to such a constraint crucially depends on how broadly the constraint is described. If this constraint can be described effectively as a complete redistribution of one's credences over a partition of propositions, the value of learning theorem holds. However, if it cannot be so formulated, then the value of learning theorem cannot be established. I explain why this is so: contrary to what the value of learning theorem requires, in such cases, the MRE updater's prior credences are not equal to the expectation of her possible posterior credences.

There is yet another angle from which we might look at the main result of this chapter. It is often said that MRE is an updating rule that prescribes modesty or minimal revision for the agent's opinion shifts. As characterized in van Fraassen (1981, p. 376), MRE is "the rule that one should not jump to unwarranted conclusions, or add capricious assumptions, when accommodating one's belief state to the deliverances of experience". Minimizing RE under some constraints imposed on posterior credences is a way, but by no means the only way, to make the idea of modesty more precise: the agent adopts the posterior credence function that meets the constraints reporting her learning experience and is RE-closest to her prior credence function. Under this procedure, the existence of a uniquely maximally modest cr' satisfying a given constraint is guaranteed, since $\mathcal{C}_\chi$ is a closed convex set. But why should we value such modest opinion shifts? Of course, modesty might itself be a virtue that does not require further justification.

Be that as it may, modesty might also be viewed as a rational tool for pursuing other goals. What this chapter shows is that it is not always true that revising credences by dint of MRE leads to modest new credences that are expected to be helpful and never harmful for one's decisions.

6.2 The Value of Learning and Bayes's Rule

It is rather uncontroversial to say that a change in one's credences may bring consequences for one's decisions. Suppose that you have to decide now whether to act on the basis of your current information or to perform a cost-free experiment to obtain further information, update your credences, and then act. For example, you have to decide whether to submit your paper to a journal now or to pursue some line of research, update your credences about the content of your paper, and then decide whether to submit it. What should you do?

There is a striking result in decision theory, due originally to Frank P. Ramsey and revived by Savage (1954) and Good (1967), that gives an answer to the aforementioned concern. Informally put, the theorem states that the prior expectation of making an informed decision is at least as great as the expected utility of making an uninformed decision, and is strictly greater if it is not the case that the maximum expected utility of an act is the same for all possible experimental results (or equivalently, if at least one of the experimental results could alter the choice of one's actions). This theorem is known in the literature as the *value of knowledge theorem* or the *value of learning theorem.*

In its original form, the theorem has been proven in the context of Bayes's rule of conditioning. As shown by Good, Bayes's rule implies the value of learning theorem. To present Good's argument, let us introduce the following assumptions:

- Let $\mathcal{A} = \{A_1, ..., A_m\}$ be a finite set of actions, and let $\mathcal{S} = \{S_1, ..., S_n\}$ be a finite set of states of the world.
- For each combination of A_i and S_j, we assign a utility $U(A_i \wedge S_j)$.
- Let $\mathcal{E} = \{E_1, ..., E_k\}$ be a finite partition of experimental outcomes.

- Assume that the agent is an expected utility maximizer, that is, she chooses the act A_i that maximizes her expected utility given by

$$\sum_j cr(S_j)U(A_i \wedge S_j),$$

where $cr(S_j)$ is the agent's prior credence in S_j.

- The agent's learning experience is reported by the constraint χ saying that one should assign the posterior credence 1 to some E_k. Then, the associated set of posterior credence functions is $\mathcal{C}_\chi = \{cr' : cr'(E_k) = 1\}$. Bayes's rule prescribes you to choose from that set the posterior credence function cr' that satisfies the constraint and is defined as follows:

  Bayes's rule: For all j,

  $$cr'(S_j) = cr(S_j|E_k),$$

  provided that $cr(E_k) > 0$.

  That is, your posterior credence in S_j equals your prior credence in S_j conditional on E_k.

- The experiment is costless.

For simplicity's sake, we consider only finite sets of states. It is worth noticing that the value of learning theorem carries over to infinite sets of states if the credence function is countably additive.

Suppose that the agent is faced with the following decision problem. She has to decide whether to act now or to wait until the experiment is performed, update her degrees of belief by Bayes's rule, and then act. Since the agent is an expected utility maximizer, the present value of her deciding now, without performing the experiment, is:

$$\begin{aligned}\max_i \sum_j cr(S_j)U(A_i \wedge S_j) &= \max_i \sum_k \sum_j cr(S_j|E_k)cr(E_k)U(A_i \wedge S_j)\\ &= \max_i \sum_k \sum_j \frac{cr(E_k|S_j)cr(S_j)}{cr(E_k)} cr(E_k)U(A_i \wedge S_j)\end{aligned}$$

$$= \max_i \sum_k \sum_j cr(S_j)cr(E_k|S_j)U(A_i \wedge S_j), \tag{6.2}$$

which is the expected value of act A_i with the highest expected utility.

The present value of making an informed decision is given as follows. Suppose that E is the true member of $\mathcal{E}$. Then, the posterior value of making a decision informed by E is the value of act A_i with the highest expected utility with respect to the conditional credence $cr(S_j|E)$:

$$\max_i \sum_j cr(S_j|E)U(A_i \wedge S_j). \tag{6.3}$$

Given (6.3), the present value of making a decision conditional on E is calculated by

$$\sum_k cr(E_k) \max_i \sum_j cr(S_j|E_k)U(A_i \wedge S_j) = \sum_k \max_i \sum_j cr(S_j)cr(E_k|S_j)U(A_i \wedge S_j), \tag{6.4}$$

which is the prior expectation of the posterior value of making an informed decision.

To see that equation (6.4) is at least as great as equation (6.2), note first that equations (6.2) and (6.4) differ only in the order of the $\max_i$ and the $\sum_k$ operations. Now, observe that for any real-valued function $f(k, i)$ of k and i:

$$\begin{aligned} \sum_k \max_i f(k,i) &\geq \sum_k f(k,i^*) \\ &\geq \max_i \sum_k f(k,i), \end{aligned} \tag{6.5}$$

where i^* is the value of i that maximizes $\sum_k f(k, i)$. Then, it follows that

$$\sum_k \max_i \sum_j cr(S_j)cr(E_k|S_j)U(A_i \wedge S_j) \geq \max_i \sum_k \sum_j cr(S_j)cr(E_k|S_j)U(A_i \wedge S_j), \tag{6.6}$$

which establishes the value of learning theorem.

The value of learning theorem carries an important philosophical message for someone who evaluates learning and updating rules in terms of their potential consequences for decisions. The message is that, from the perspective of maximizing expected utility, a change in one's credences could make one's decisions better and never worse. That is, acquiring information by way of an update is expected to be helpful and never harmful. Of course, this result does not hold unconditionally. It rests on a few substantial assumptions. First of all, it is set up in the framework of Savage's decision theory in which states of the world and acts are stochastically independent in the sense that choosing an act does not give you information about which state of the world is true. Likewise, one's decision whether to perform an experiment is stochastically irrelevant to the states of the world. Notice, however, that updating on experimental outcomes may alter your credences about the states. Second, the states, acts, and utilities are the same before and after updating your credences. Third, it is assumed that you are an expected utility maximizer before and after updating.

It is important to recognize that the agent assesses the value of making an informed decision from her current perspective, without knowing which of the experimental outcomes is true. To assess this value, she takes the expectation of equation (6.3) with respect to the unknown E_k. This, in turn, shows how her prior credences must be related to her possible posterior credences. Since she knows that she will update by Bayes's rule, it follows that, for each j, her prior credence in S_j must be equal to the expectation of her conditional prior credence, $cr(S_j|E_k)$; that is,

$$cr(S_j) = \sum_k cr(E_k)cr(S_j|E_k), \tag{6.7}$$

where the sum extends over all k such that $cr(E_k) > 0$. This is an elementary observation. However, what happens if Bayes's rule is not assumed? In the next section, I will suggest a more general answer to the question of how the agent's prior credence function should be related to her possible posterior credence functions for the value of learning to be satisfied. This answer requires to focus on Brian Skyrms's condition M, which relates one's prior and posterior credences.

6.3 Condition M and the Value of Learning

Does the value of learning imply a particular way in which one's prior and one's possible posterior degrees of belief are related? In this section, we give an affirmative answer to this question by exploring Skyrms's condition M. I will present this condition within the framework of an unstructured and opaque degrees-of-belief change called by Skyrms (1990) the *black-box learning*. It is unstructured in the sense that we do not know how the agent updates her credences, i.e. we do not know what rule she adopts as her updating policy and what the constraint that prompts the shift in her credences is. The only thing we know is the effect of her learning experience on her posterior credences.

Black-box learning is a generalized model of learning. According to it, an epistemic agent starts with a prior credence function, passes through a black-box learning experience, and ends up with a posterior credence function. Thus, the agent only knows the input (prior credence function) and the output (posterior credence function). Here the learning process is not transparent: the agent cannot go into the black-box and see what is inside. In particular, she cannot say whether she learned a proposition with certainty or redistributed her degrees of belief over a partition of propositions. That is, she cannot specify a constraint that prompts the shift in her degrees of belief. Likewise, she cannot specify a rule of updating that would deal with her learning episode. For example, she does not expect that she would learn a proposition as a result of her interaction with the environment, yet she might think about this experience and revise her opinion on the basis of her thoughts. More precisely, black-box learning may be described as follows. Let an agent's degrees-of-belief space be a triple $(\mathcal{W}, \mathcal{F}, cr)$, where $\mathcal{W}$ is a set of worlds that the agent considers possible, the elements in $\mathcal{F}$ are propositions about which the agent has an opinion, and cr is the agent's credence function. Suppose that the agent is in a learning situation, where she expects her credence function over $\mathcal{F}$ to change from cr to one of the posterior credence functions in the set $\{cr'\}$, resulting from her interaction with the environment. Since her learning is described only by the effect on her possible posterior credence functions, we can enlarge her degrees-of-belief space by adding the posterior credence function as a random variable. As a result, the agent might have second-order degrees of belief over propositions about the first-order ones, where the first-order credences are her possible posterior credences. By doing so, we get a higher-order probability structure in the sense proposed in Gaifman (1988). Such a structure may be represented by $(\mathcal{W}, \mathcal{F}, cr, cr')$, where $\mathcal{F}$ is an algebra of propositions, subsets from $\mathcal{W}$, cr is one's prior credence function over $\mathcal{F}$, and cr' is a measurable function defined as $cr' : \mathcal{F} \times [0, 1] \to \mathcal{F}$. Let the proposition about one's posterior credences be denoted by $X_{cr'}$. The proposition says that the posterior credence function over $\mathcal{F}$ is given by cr'.

Could a black-box learner satisfy the value of learning? Recall that the black-box learner has no updating rule at his disposal and no constraint that prompts his degrees-of-belief shift. One might thus be suspicious as to whether black-box learning could be even justified. After all, we deal with a situation where one expects one's credences will change as a result of an interaction with the environment without being confident that the change will be prompted by something learned. Additionally, a black-box learning situation does not exclude the possibility that reasons other than learning might prompt one's degrees-of-belief change. In particular, one might expect that one's credences will change by taking a drug that makes one confident that one can fly, by memory loss, or by being brainwashed.

Skyrms (1990) shows convincingly that a sufficient condition for one's degrees-of-belief change in black-box learning to satisfy the value of learning is the following:

> **M:** An agent's prior credence function ought to be such that, for all j and for any possible posterior credence function cr',[1]
>
> $$cr(S_j|X_{cr'}) = cr'(S_j),$$
>
> providing that $cr(X_{cr'}) > 0$.

That is, condition M requires one's prior credence in S_j conditional on the proposition about S_j's posterior degree of belief to be equal to that posterior degree of belief. A similar condition, known as the *Reflection Principle*, has been defended in van Fraassen (1984).

[1] In Skyrms (1990), M stands for Martingale.

Let me now show that a degrees-of-belief shift that satisfies condition M leads to the value of learning theorem. The agent's present value of deciding now is the maximum of her prior expectation of posterior expected utility. In symbols,

$$\begin{aligned} \max_i \sum_j cr(S_j)U(A_i \wedge S_j) &= \max_i \sum_{cr'} \sum_j P(S_j|X_{cr'})cr(X_{cr'})U(A_i \wedge S_j) \\ &= \max_i \sum_{cr'} \sum_j cr'(S_j)cr(X_{cr'})U(A_i \wedge S_j). \end{aligned} \tag{6.8}$$

The posterior value of making a decision informed by $X_{cr'}$ is given by

$$\max_i \sum_j cr(S_j|X_{cr'})U(A_i \wedge S_j) = \max_i \sum_j cr'(S_j)U(A_i \wedge S_j). \tag{6.9}$$

Now, we can calculate the present value of making an informed decision as one's prior expectation of the value given by equation (6.9). That is,

$$\sum_{cr'} cr(X_{cr'}) \max_i \sum_j cr'(S_j)U(A_i \wedge S_j). \tag{6.10}$$

To see that the value given by equation (6.10) is at least as great as the value given by equation (6.8), let me first introduce the following formulation of Jensen's inequality:

> **Jensen's inequality:** For any random variable X and any convex function ϕ,
> $$\text{Exp}\,[\phi(X)] \geq \phi(\text{Exp}\,[X]),$$
> where Exp stands for expectation.

With this inequality in mind, let us write

$$\phi(A_i) = \max_i \sum_j cr'(S_j)U(A_i \wedge S_j), \tag{6.11}$$

where ϕ is a convex function of A_i. Also, we can write

$$\text{Exp}\,[\phi(A_i)] = \sum_{cr'} cr(X_{cr'}) \max_i \sum_j cr'(S_j)U(A_i \wedge S_j), \tag{6.12}$$

and

$$\phi(\mathrm{Exp}\,[A_i]) = \max_i \sum_{cr'} \sum_j cr'(S_j) cr(X_{cr'}) U(A_i \wedge S_j). \tag{6.13}$$

Then, it follows, by Jensen's inequality, that

$$\sum_{cr'} cr(X_{cr'}) \max_i \sum_j cr'(S_j) U(A_i \wedge S_j) \geq \max_i \sum_{cr'} \sum_j cr'(S_j) cr(X_{cr'}) U(A_i \wedge S_j), \tag{6.14}$$

as required. Hence, condition M satisfies the value of learning.

What happens if condition M does not hold? Skyrms (1997) shows that if the black-box learner fails to satisfy condition M, then the expected utility of her informed decision could be lower than the expected utility of her uninformed decision. Thus, condition M is both sufficient and necessary for the value of learning to hold. Similarly, Huttegger (2014) argues that condition M and the value of learning are in fact equivalent. Assuming Skyrms's result, Huttegger shows, quite generally, that if updating one's credences satisfies the value of learning, then condition M must hold. Thus, condition M is all we need for the value of learning to hold.

To explain the necessity of condition M, suppose that $cr(S_j|X_{cr'}) = \frac{1}{3}$ and $cr'(S_j) = \frac{2}{3}$. Hence, you violate condition M. Consider a bet on S_j conditional on the proposition that $cr'(S_j) = \frac{2}{3}$; it costs you \$5 and pays you \$5 if both S_j and the proposition that $cr'(S_j) = \frac{2}{3}$ are true. Since you violate condition M, you are vulnerable to a Dutch book, i.e. a set of bets that guarantee you a net loss, come what may. You have to decide now whether to accept this bet or to update your credence in S_j and then decide. Since your decision to reject this bet now has greater expected utility than your decision to act later, and possibly to risk acceptance of this bet, the value of learning theorem fails to hold.

Now, if condition M alone is all that is required for the value of learning to hold, we can determine, by focusing solely on that condition, the way in which one's prior and posterior credences should be related for one's opinion shift to satisfy the value of learning. Additionally, since we deal with a black-box learning situation, this way of relating priors and posteriors must be independent of which updating rule the agent endorses as her updating policy.

It is an immediate consequence of condition M that one's prior credences are the expectation of one's anticipated posterior credences, i.e. for all j,

$$cr(S_j) = \sum_{cr'} cr'(S_j) cr(X_{cr'}). \tag{6.15}$$

In other words, the agent's prior credence in S_j is a convex combination of her possible posterior credences in S_j. Given that equation (6.15) is a consequence of condition M, if equation (6.15) fails to hold, then condition M cannot be satisfied, and hence the value of learning theorem cannot be established. Note that equation (6.15) does not tell us how the agent arrives at her posterior credences. After all, equation (6.15) characterizes a black-box learner. The basic idea behind equation (6.15) is that no matter how the agent arrives at her posterior credences, her prior credences are required to be the expectation of her posterior ones.

It is not hard to observe that a Bayesian conditionalizer satisfies equation (6.15). If you know that you will update by dint of Bayes's rule, your prior credences are the expectation of your anticipated posterior ones that are given by the conditional prior credences. Of course, the important question here is: how could one's conditional credences, the $cr(S_j|E_k)$'s, capture one's anticipated credences that figure in equation (6.15)? Two interesting answers to this question are given in the literature. First, as pointed out in Weisberg (2007), one might believe with credence one that one will update by Bayes's rule on E_k. Then, one's anticipated future credences are just the $cr(S_j|E_k)$'s. Second, following Easwaran (2013), one might view the $cr(S_j|E_k)$'s as "plans" to update one's credences after learning which member of $\mathcal{E}$ is true. Then, the agent's anticipated future credences are simply her credences that she plans to have. In my view, both of these answers are plausible ways to find a bridge between one's conditional credences and one's anticipated future credences.

In what follows, I show that updating by MRE on a constraint prompting a complete redistribution of credences over a partition of propositions agrees with a Bayesian model of learning from experience that satisfies equation (6.15). This, in turn, leads straightforwardly to the value of learning theorem for MRE. However, I also show that MRE updates on a constraint prompting a change in one's conditional credences might not lead to the value of learning theorem. I explain that this is because such MRE updates might not coincide with a model of learning that satisfies equation (6.15).

6.4 The Value of Learning and MRE

In general, MRE updating can be applied to a learning situation reported by an affine constraint on posterior credences. An affine constraint can always be formulated as saying that one's expectation of a random variable, computed relative to one's posterior credence function, has a given value. Examples of such constraint include: (i) a constraint to the effect that one should assign posterior credences to a partition of propositions without conferring certainty on any of them, or (ii) a constraint to the effect that one should assign a conditional posterior credence for some proposition given another proposition.

For example, to see how constraint (i) can be expressed as one's expectation of a random variable, suppose that X is a $\mathcal{F}$-measurable random variable, i.e. a function from $\mathcal{W}$ to the real numbers $\mathbb{R}$. Suppose that the elements of a partition $\mathcal{E} = \{E_1, ..., E_k\}$ of W are represented as 0,1-valued random variables or indicator functions. The indicator function of E_i, denoted by $I_{E_i}(w)$, can be understood as the truth value of E_i at world w, that is, $I_{E_i}(w) = 1$ if $w \in E_i$, and $I_{E_i}(w) = 0$ otherwise. Since posterior credences over the members of that partition are equal to the posterior expectations of the indicator functions, (i) may be reformulated as a constraint to the effect that the expectations of these indicator functions, computed with respect to the posterior credence function, get some values in $\mathbb{R}$. In this section, I show that an MRE update in response to constraint (i) leads to the value of learning theorem.

To this end, I first introduce the following well-known result. Suppose that the agent's learning experience is reported by the following constraint. Let $\mathcal{E} = \{E_1, ..., E_k\}$ be a partition of $\mathcal{W}$, and let $q_1, ..., q_k \in \mathbb{R}^+$ be such that $q_1 + ... + q_k = 1$. Then, χ is a constraint to the effect that upon learning experience, the agent redistributes her credences over $\{E_1, ..., E_k\}$ such that $cr'(E_i) = q_i$, for $i = 1, ..., k$. The agent's set of posterior credences that satisfy this constraint is given by $\mathcal{C}_\chi = \{cr' : cr'(E_i) = q_i, i = 1, ..., k\}$, which is a closed and convex set. Given that the agent updates her credences by MRE, she chooses from the set

$\mathcal{C}_\chi$ her posterior credence function that minimizes the distance measured by RE. There is a result showing that if the constraints on posterior credences concern a whole partition of propositions, RE is uniquely minimized just in case the agent's posterior credence function comes by Jeffrey's rule on the partition $\{E_1, ..., E_k\}$ (see Williams 1980; Diaconis and Zabell 1982). That is, cr' should be such that, for all j,

$$cr'(S_j) = \sum_i cr(S_j|E_i)q_i. \tag{6.16}$$

That is, cr' is a weighted average of the agent's prior conditional credence in S_j given E_i, for all i, where the weights are the values of posterior credences for the E_i's. This result may be summarized by the following proposition:

Proposition 6.1. *Suppose that* $\mathcal{C}_\chi = \{cr' : cr'(E_i) = q_i, i = 1, ..., k\}$. *Then,* $\mathrm{RE}(cr, cr'') \geq \mathrm{RE}(cr, cr(\cdot|E_i)q_i)$ *for all* $cr'' \in \mathcal{C}_\chi$*, with equality just in case* $cr'' = cr(\cdot|E_i)q_i$.

As shown by Jeffrey (1983), the agent's posterior credence function is equal to the one given by formula (6.16) if and only if the following condition holds:

Rigidity: For all j and all i,

$$cr'(S_j|E_i) = cr(S_j|E_i).$$

Rigidity says that the agent's conditional credences given members of $\{E_1, ..., E_k\}$ remain intact as she shifts her credences from cr to cr'. Since MRE updating on a whole partition $\{E_1, ..., E_k\}$ is also rigid, there is no surprise that it coincides with Jeffrey's rule. We may look at Rigidity in the case of MRE updating as follows: under RE-minimization, for each member E of $\{E_1, ..., E_k\}$, the ratios of one's posterior to one's prior credences about propositions that imply E do not change, i.e. if S_i and S_j, $i \neq j$, imply E, then $\frac{cr'(S_j)}{cr'(S_i)} = \frac{cr(S_j)}{cr(S_i)}$.

With this result in hand, we can introduce a way to represent MRE updating in response to constraint (i) as Bayesian conditioning in an enlarged degrees-of-belief space. This move is mobilized by a general result, due to Persi Diaconis and Sandy Zabell (1982), which says when a shift from cr to cr' in the original smaller space agrees with Bayesian conditioning in some bigger space. A related result, though somewhat different in detail, is defended by Peter Grünwald and Joseph Halpern (2003). For a two-element partition of propositions, a similar result is given in Skyrms (1980b). Roughly, the idea is as follows. Suppose that the agent shifts from cr to cr' by MRE updating on a partition of propositions. Given the agent's learning experience reported by a complete redistribution of her credences over that partition, we can enlarge the original space by adding the proposition that describes the agent's learning experience and the proposition that describes its absence. The proposition that describes the agent's learning experience is about the values that her posterior credence function assigns to each member of the partition. Then, under certain conditions, we can show that the MRE update in the original smaller space agrees with Bayesian conditioning in the bigger space.

More precisely, to enlarge the agent's degrees-of-belief space, we add to the algebra $\mathcal{F}$ a proposition X_{q_i} for each member i of the partition $\mathcal{E}$. Thus, we require that the underlying space $(\mathcal{W}, \mathcal{F})$ is sufficiently rich. In fact, each element of $\mathcal{W}$ specifies a value for q_i, which, in turn, may be regarded as a random variable. X_{q_i} says that the agent's posterior credence assigned to the i-th member of $\mathcal{E}$ equals q_i. This proposition may be understood as a set of worlds from $\mathcal{W}$ at which the posterior credence in E_i equals q_i. Denote the algebra extended by adding such propositions by $\mathcal{F}^*$. The agent's prior credence function cr over $\mathcal{F}^*$ may be viewed as a second-order credence function, since it assigns credences to her other credences assigned to the propositions in the smaller original algebra $\mathcal{F}$. Propositions about which the agent has an opinion and that belong to the extended algebra are the propositions that describe her learning experience reported by constraint (i), to wit, a learning experience that prompts a complete redistribution over the partition $\mathcal{E}$. Such propositions specify the agent's credences for every member of the partition $\mathcal{E}$. They may be understood as conjunctions, the $\bigwedge_{i=1}^{k} X_{q_i}$'s, of the X_{q_i}'s. For ease of exposition, denote such a conjunction by D.

Now, if the agent learns D with certainty, she can Bayes condition in the enlarged algebra. In fact, when she conditions in the enlarged algebra, she assigns second-order credences to propositions about her first-order ones. Denote such Bayesian conditioning in the enlarged algebra by BC*. It can be put as follows:

BC*: For all j and any $D \subseteq \mathcal{W}$,

$$cr'(S_j) = cr(S_j|D),$$

provided that $cr(D) > 0$.

The following theorem states that under certain conditions, updating by MRE on a partition $\mathcal{E}$ is representable as BC*.

Theorem 6.1. *Suppose that the agent's prior credence function cr obeys the following two conditions:*

(1) For all i, $cr(E_i|D) = q_i$, provided that $cr(D) > 0$.

(2) For all j and all i, $cr(S_j|E_i \wedge D) = cr(S_j|E_i)$, provided that $cr(E_i \wedge D) > 0$.

Then, for all j, $cr(S_j|D) = \sum_i cr(S_j|E_i)q_i$.

Proof. Suppose that cr satisfies conditions (1) and (2), $D \subseteq \mathcal{W}$, and $E_i \subseteq \mathcal{W}$ for all i. Then,

$$\begin{aligned} cr(S_j|D) &= \sum_i cr(S_j|E_i \wedge D)cr(E_i|D) \\ &= \sum_i cr(S_j|E_i \wedge D)q_i \text{ (by condition (1))} \\ &= \sum_i cr(S_j|E_i)q_i \text{ (by condition (2))}, \end{aligned}$$

as required. □

In fact, the theorem says that Bayesian conditioning in the enlarged algebra of propositions is in agreement with updating by MRE on a whole partition of propositions that belongs to some subalgebra of the enlarged one. This agreement rests on two conditions, originally introduced in Skyrms (1980b). Condition (1) is an application of condition M, whereas condition (2) is a kind of probabilistic independence called by Skyrms sufficiency. Both conditions have an intuitive appealing. Condition (1) says that the agent's prior credence in E_i, conditional on the proposition specifying posterior credences over the members of $\mathcal{E}$, should be equal to the posterior credence in E_i. This condition can be understood as saying that learning described by D is legitimate or justified. For example, it indicates that such a learning is not a result of memory loss. Sufficiency tells us that S_j is conditionally independent of D given each member of $\mathcal{E}$. Intuitively, if the agent knows which member of $\mathcal{E}$ is true, then her knowledge about credences assigned to each member of that partition should have no bearing on her credence in A.

However, we should not regard the conditions given above as universally correct. Clearly, condition (1) does not hold in epistemically "pathological" situations. Just consider the example of Ulysses and the sirens. Before hearing the siren's song, Ulysses has a high credence that sailing among the rocks is dangerous. However, he is also sure that after hearing the sirens, he would cease to believe (wrongly as he now thinks) that sailing among the rocks is dangerous. If he were to obey condition (1), he would have to cease to believe now that sailing among the rocks is dangerous. However, he now believes that this is not so, and so condition (1) is violated. Likewise, sufficiency does not hold in situations where S_j is a proposition X_{q_i}. Then, since D implies X_{q_i}, it is E_i, not D, that is irrelevant to S_j. However, whenever these two conditions hold, which seems to be fairly common, Bayesian conditioning in an enlarged degrees-of-belief space yields the same result as the MRE shift over a whole partition in the original smaller degrees-of-belief space.

Where does this result leave us vis-à-vis the question of whether a MRE shift on a whole partition satisfies the value of learning? To address this question, we first need to face a potential difficulty. Recall that in Good's argument, the experiment is represented by a finite partition of propositions, whose members are measurable subsets in $\mathcal{W}$. However, the outputs of learning experiences represented by constraint (i) are the values of posterior degrees of belief, not propositions. If this is so, how could the MRE updater assign credences to them? Furthermore, how could she determine the values of informed and uninformed decisions? By virtue of the representation introduced above, this difficulty can be mitigated by acknowledging that such values of posterior credences can be expressible as proposition D, which is a measurable subset in $\mathcal{W}$. That is, from the point of view of the enlarged degrees-of-belief space, what we learn from the experiment reported by constraint (i) is a proposition about the values of posterior credences over the members of a partition. Now, by moving to an enlarged degrees-of-belief space, we can think of a cost-free experiment as r possible results prompting r possible redistribution of the agent's credences over $\mathcal{E}$. Denote the m-th redistribution of the kind by the proposition D_m.

Now, it is easy to observe that by virtue of the representation captured in Theorem 6.1, the MRE updater on $\mathcal{E}$ satisfies condition M, and thus the value of learning theorem can be established. Since she can be represented as a Bayesian conditionalizer in the enlarged degree-of-belief space, in which the D_m's are measurable subsets, her prior credence in each state of the world S will be the expectation of her posterior credence in each S. These posteriors are given by the conditional prior credences, the $cr(A|D_m)$'s, defined in the enlarged degrees-of-belief space. More precisely, a demonstration that such an MRE update satisfies the value of learning may proceed as follows. The present value of making an uninformed decision is:

$$\begin{aligned}\max_i \sum_j cr(S_j)U(A_i \wedge S_j) &= \max_i \sum_m \sum_j cr(S_j|D_m)cr(D_m)U(A_i \wedge S_j)\\ &= \max_i \sum_m \sum_j \frac{cr(D_m|S_j)cr(S_j)}{cr(D_m)} cr(D_m)U(A_i \wedge S_j)\\ &= \max_i \sum_m \sum_j cr(S_j)cr(D_m|S_j)U(A_i \wedge S_j).\end{aligned} \tag{6.17}$$

The posterior value of making a decision informed by D_m is given by

$$\max_i \sum_j cr(S_j|D_m)U(A_i \wedge S_j). \tag{6.18}$$

Given Equation (6.18), the present value of making a decision conditional on D_m is calculated by

$$\begin{aligned}&\sum_m cr(D_m) \max_i \sum_j cr(S_j|D_m)U(A_i \wedge S_j)\\ =&\sum_m cr(D_m) \max_i \sum_j \frac{cr(D_m|S_j)cr(S_j)}{cr(D_m)}\end{aligned}$$

$$= \sum_m \max_i \sum_j cr(S_j) cr(D_m|S_j) U(A_i \wedge S_j), \tag{6.19}$$

which is the prior expectation of the posterior value of making an informed decision. Now, it is easy to notice that, on the same mathematical grounds as in Good's argument, the value given by equation (6.19) is at least as great as the value given by equation (6.17). Hence, MRE updating on $\mathcal{E}$ represented as BC* is expected to be helpful, and never harmful to one's decisions.

6.5 When the Value of Learning May Not Hold for MRE Updating

In this section, I examine the question of whether MRE updating in response to constraint (ii) leads to the value of learning theorem. As will be apparent, the answer to this question is: it depends on how broadly one's learning experience reported by constraint (ii) is described. More specifically, I show that whether the value of learning can be established in this case may be dependent on whether or not the contextual information, not reported by constraint (ii), is taken into account in addition to the explicit information. If only the explicit information is taken into account in this case, then the value of learning theorem may not hold. By taking the contextual information into account, the constraint is made complete in the sense explicated below, and the value of learning theorem holds.

In general, consider a learning experience in which the agent learns the following conditional information "If A, then the odds for B are $\sigma/(1-\sigma) : 1$", for $\sigma \in [0, 1]$. This information may prompt a change in the agent's conditional prior credences. That is, after learning this conditional information, her conditional prior credence, $cr(B|A)$, changes to her conditional posterior credence, $cr'(B|A)$, which should be set equal to σ. With this constraint, we associate a closed and convex set of posterior credence functions $\mathcal{C}_\chi = \{cr' : cr'(B|A) = \sigma\}$. In order to answer the question of whether a shift from cr to cr', which belongs to this set and minimizes RE, leads to the value of learning theorem, we examine whether such a shift satisfies condition M.

For concreteness, we focus on the famous Judy Benjamin case, originally introduced in van Fraassen (1981). In this case, private Judy Benjamin is dropped in an area that is divided into two territories, the red territory (R) and the blue territory ($\neg R$). Each of these territories is further divided into the second company area (S) and headquarters company area ($\neg S$). These divisions form four quadrants. Initially, Judy assigns to each of the four quadrants a credence of one quarter: $cr(R \wedge S) = cr(R \wedge \neg S) = cr(\neg R \wedge S) = cr(\neg R \wedge \neg S) = \frac{1}{4}$. Judy, then, receives the following radio message: "I don't know where you are. If you are in the red territory, the odds are 3:1 that you are in the headquarters company area". That is, the radio message prompts a change in one of Judy's conditional credences by setting $cr'(\neg S|R) = \frac{3}{4}$. Suppose further that Judy is an MRE updater, and let $R \wedge S, R \wedge \neg S, \neg R \wedge S, \neg R \wedge \neg S$ be the minimal elements of $\mathcal{F}$. Now, we may distinguish two ways of describing the constraint on Judy's posterior credence function:

(i*) The constraint pertains to *all* propositions in $\{R \wedge S, R \wedge \neg S, \neg R\}$,

(ii*) The constraint pertains to *some* propositions $\{R \wedge S, R \wedge \neg S, \neg R\}$.

Let us consider each of these in turn. Case (i*) rests on the assumption that the MRE updater can obtain additional information about her posterior credences over the members of the entire partition by looking at the context of the Judy Benjamin case. The only explicit information she gets is the information about her posterior conditional credence in $\neg S$ given R, i.e. $cr'(\neg S|R) = \frac{3}{4}$. Given this information, she knows how to set her posterior conditional credence in S given R: since all of her conditional credences must sum to one, we have that $cr'(S|R) = \frac{1}{4}$. However, this does not yet provide a redistribution over the entire partition. What about her shift from $cr(\neg R)$ to $cr'(\neg R)$? This information is not given explicitly. However, this information can be gleaned from the context of the case: since the radio message does not say whether Judy is in the red or in the blue territory, it follows that her credence in $\neg R$ remains unchanged, i.e. $cr'(\neg R) = cr(\neg R) = cr(\neg R \wedge S) + cr(\neg R \wedge \neg S) = \frac{1}{2}$. This completes her redistribution over the entire partition of propositions. Let us assume that Judy's learning experience does not lead her to the revision of her credence in R. We thereby assume a condition called *independence* (see Bradley 2005). Then, the sum of Judy's posterior credences in $R \wedge \neg S$ and $R \wedge S$ equals her prior credence in R, i.e. $cr'(R \wedge \neg S) + cr'(R \wedge S) = cr'(\neg S|R)cr(R) + cr'(S|R)cr(R)$. Now, Judy's task is to find the posterior credence function $cr' \in \left\{cr' : cr'(R \wedge \neg S) = \frac{3}{8}, cr'(R \wedge S) = \frac{1}{8}, cr'(\neg R) = \frac{1}{2}\right\}$ that minimizes RE relative to cr. As shown in Douven and Romeijn (2011) in a more general setting, RE is minimized if and only if, for all $A \in \mathcal{F}$:

(1) $cr'(A|R \wedge \neg S) = cr(A|R \wedge \neg S)$,

(2) $cr'(A|R \wedge S) = cr(A|R \wedge S)$,

(3) $cr'(A|\neg R) = cr(A|\neg R)$.

That is, Judy's new credence function minimizes RE relative to her prior credence function if and only if the shift in her credences is rigid, and thus goes in accord with Jeffrey's rule on the partition $\{R \wedge S, R \wedge \neg S, \neg R\}$.

As emphasized in Douven and Romeijn (2011), by using the contextual information in the Judy Benjamin case, we can complete the constraint reporting Judy's experience in a way that allows us to redistribute her credences over the entire partition of propositions and to apply Jeffrey's rule. Where does this result leave us vis-à-vis the question of whether an MRE shift in response to constraint (ii) leads to the value of learning theorem? If constraint (ii) pertains to the entire partition of propositions to which Jeffrey's rule can be applied, then, in view of the representation given in section 6.4, the MRE updater may be represented as a Bayesian conditionalizer in a degrees-of-belief space in which this constraint is a measurable subset of $\mathcal{W}$. Consequently, she satisfies condition M, and thus the value of learning holds for this case.

Things change if we turn to case (ii*). Here the radio message received by Judy prompts an incomplete redistribution of her credences over $\{R \wedge S, R \wedge \neg S, \neg R\}$. Here I assume that no information that makes the redistribution complete can be gleaned from the context of this case. The radio message is the sole constraint imposed on her posterior credence function. This explicit constraint causes her redistribution over $R \wedge S$ and $R \wedge \neg S$, leaving her posterior credence in $\neg R$ unknown. However, as shown in van Fraassen (1981), by using MRE updating, we can determine Judy's posterior credence in this proposition. However, this determination leaves us with a highly counter-intuitive consequence: $cr'(\neg R) > \frac{1}{2}$, and hence $cr'(\neg R) > cr(\neg R)$. That is, Judy's new degree of belief in $\neg R$ is greater than her prior credence $\neg R$, even if the radio message yields no information relevant to whether she is in the red rather than in the blue territory. More generally, for any value of σ, one's posterior credence in $\neg R$ that minimizes RE would be greater than one's prior credence in $\neg R$, and it remains unchanged only if $\sigma = \frac{1}{2}$. However, apart from being counter-intuitive, this observation shows that the MRE updater cannot satisfy condition M.

To show this, I explore a result, due to Seidenfeld (1986) and rehearsed by Uffink (1996), which shows that MRE updating cannot be represented as Bayesian conditioning in an enlarged space in which an incomplete constraint (ii) is a measurable subset of $\mathcal{W}$ unless the constraint is irrelevant to one's prior credence in $\neg R$. Suppose that Γ_σ (in the Judy Benjamin case, $\sigma = \frac{3}{4}$) is a measurable subset of $\mathcal{W}$. Since, for any value of σ, the posterior credence in $\neg R$ increases unless $cr(\neg R) = cr'(\neg R)$, we have that in the enlarged degrees-of-belief space:

$$cr(\neg R) \geq \int_0^1 cr(\neg R|\Gamma_\sigma)cr(\Gamma_\sigma)\,\mathrm{d}\sigma, \tag{6.20}$$

with strict inequality when there is some probability mass function on Γ_σ for $\sigma \neq 1/2$. That is, the prior credence in $\neg R$ cannot be a convex combination of the conditional credences, the $P(\neg R|\Gamma_\sigma)$'s, for $\sigma \neq 1/2$. Not only does it show that MRE updating in case (ii*) cannot be represented as Bayesian conditioning in the enlarged space, but also it shows that MRE updating in that case fails to satisfy condition M unless $cr(\neg R) = cr'(\neg R)$. If the conditional credences, the $cr(\neg R|\Gamma_\sigma)$'s, are understood as possible posterior credences, the $cr'(\neg R)$'s, then we have that

$$cr(\neg R) \geq \sum_{cr'} P'(\neg R)cr(X_{cr'}). \tag{6.21}$$

Consequently, $cr(\neg R|X_{cr'}) \geq cr'(\neg R)$, and so condition M does not hold in general. Additionally, given that condition M is both necessary and sufficient for the value of learning to hold, it follows that MRE updating does not in general lead to the value of learning theorem. That is, MRE updating may lead to a decrease in expected utility.

The above analysis has an interesting philosophical import. Whether MRE updating leads to the value of learning theorem in the case of constraint (ii) crucially depends on whether or not the agent takes into account the contextual information. However, this should not strike us as odd, for there is nothing in the machinery of MRE updating that could determine the unique way of describing one's learning experience. This opens the possibility of using both explicit and contextual information in order to determine a given constraint. More to the point, MRE does not suffice to guarantee the value of learning when the new information comes as constraints over conditional credences. It has been shown that to guarantee the value of learning, MRE must be supplemented by some additional rule, which tells us how to add extra constraints gleaned from the context.

Note, however, that case (ii*) also points towards another notion of context sensitivity. This has to do with how MRE determines the lacking information about one's posterior credence in $\neg R$. Though this information is not given explicitly, MRE could fill in the blanks for us. However, whether it does this adequately depends on the details of a given learning situation, which also include the context. On the widespread view, in the Judy Benjamin case, MRE does not fill in the blanks adequately, for it leads to counter-intuitive results: after updating, Judy's credence in $\neg R$ increases, while intuitively it should remain unchanged. However, it is perfectly possible to add to the Judy Benjamin case a story indicating that the choice of the blue or red territory is dependent on the choice of the red headquarters company area or the red second company area. However, this type of context-sensitivity should be distinguished from the one described above. For whatever story we plot in the Judy Benjamin case, MRE may provide us with the lacking information in a way that violates condition M, as indicated in Grünwald (2000). In contrast, the type of context sensitivity we alluded to above has consequences for whether or not condition M is satisfied by the MRE updater.

Let me point out some consequences of our analysis. The fact that, in some cases, the application of MRE and its justification in terms of the value of learning is context-sensitive both lend credence to the idea that updating rules are essentially tools in the "art of judgment", rather than universally valid inductive rules. In this spirit, Bradley (2005, p. 362) points out that even Bayes's rule "should not be thought of as a universal and mechanical rule of updating, but as a technique to be applied in the right circumstances, as a tool in what Jeffrey terms the 'art of judgment' ". Similarly, Douven and Romeijn (2011, p. 660) stress that adopting an updating rule based on minimizing distance between credences to cover updating on conditional information "may be an art, or a skill, rather than a matter of calculation or derivation from more fundamental epistemic principles". The analysis just given shows that even a justification of MRE updating in terms of the value of learning cannot proceed mechanically. Rather, it requires a careful consideration of the entire learning experience that the agent undergoes.

It is important to emphasize that our analysis should not be regarded as providing a support to yet another idea, widely discussed within the degrees-of-belief dynamics, called by van Fraassen (1989) *voluntarism*. According to this idea, deliverances of experience should be understood as *commands* that constrain the agent's posterior credences. These commands reflect the agent's decision to accept whatever her learning experience reveals. It is not hard to observe that voluntarism may lead to the idea that belief change is sensitive to what the agent accepts as her constraint. After all, two agents may accept different constraints on their posterior degrees of belief, even if they undergo the same learning experience. However, this is different from saying that the way in which we respond to a constraint depends on the context of our learning experience; for the context is not a feature of the agent's epistemic attitudes, but rather, it is a part of the learning experience that bears on the agent's epistemic attitudes. Hence, whether or not the context of a given case contributes to one's learning experience is not a matter of one's voluntary decision. Of course, according to voluntarism, the agent might voluntarily decide not to take the contextual information as her constraint. However, our analysis does not force us to accept this possibility.

6.6 Concluding Remarks

Clearly, the analysis given in this chapter is not a full story on the justification of MRE in terms of the value of learning. I have discussed this issue with respect to only two types of constraints: the first pertaining to a redistribution of one's degrees of belief over the entire partition of propositions; the second pertaining to a change in one's conditional degrees of belief. Despite this limitation, I have shown that the justification of MRE updating is not so simple a task as one might think. By fitting MRE updating and Bayesian conditioning together in an enlarged space, I have shown that in cases involving the first constraint, MRE leads to the value of learning. However, I have argued that this might not be so in cases involving the second type of constraint. In such cases, whether or not the value of learning holds crucially depends on whether the context of one's learning experience is taken into account.

We may transfer the insights of our analysis to the discussion about the status of MRE updating. Recall that initially, I have distinguished, from various views on this issue, the view on which MRE updating is universally valid and the views that deny its universal validity. It is tempting to think that if this rule of updating were universally valid, it would be neutral with respect to how a given learning experience is described. Moreover, it seems that if it were universally valid, its justification would not depend on whether or not the contextual information is reported by a given constraint. The findings of this chapter show that neither the application of MRE nor its justification are so neutral. Hence, they lend credence to the claim that MRE is not a universal or mechanical updating rule.

Chapter 7

Conclusions

The thesis is by no means the whole story about the double life of probability. My goal has not been that of those who set out to draw maps of the seas and who have to record all their shipping routes, ports, and islands. Rather, the goal has been to look for some well navigable channels to reach some destinations. As the navigable channels, I have used various principles relating chances to credences, prior chances to posterior chances, and prior credences to posterior credences. Although the results established in this thesis do not settle once and for all the questions that were listed in chapter 1, I believe that some interesting answers to them have been given by employing these principles.

Having reached the destinations, it is worth to summarize the main achievements and to look at some unexplored territories.

7.1 Chance and its Roles

One of the overarching themes of this thesis is that we can master the concept of chance by trying to understand what chance does, or what functional roles it plays. Unlike various frequency and propensity theories, this view does not force us to analyse the concept of chance in terms of allegedly more graspable concepts. Although I have not given an exhaustive list of the roles that chance plays in our life, I hope I have shown how this view provides a viable understanding of chance. As shown in chapter 2, focusing on the expert role of chance, along with the condition that chances, conceived as experts, should be weakly predictively accurate, allows us to show that chance is a finitely additive probability function.

It is worth emphasizing that there need not be a conflict between the view that focuses on the roles of chance and the traditional philosophical theories of chance like frequency and propensity theories. That is, one might ask the following question: which, if any, of the philosophical theories of chance gives an analysans of chance that satisfies the principles capturing various roles of chance?

A question of this sort was asked by Lewis (1986) in connection with his Principal Principle. Initially, Lewis claimed that only a reductionist about chance, for example a proponent of a frequency theory like Lewis's best-system theory, could show that chance should constrain an agent's credence, on pain of irrationality. For Lewis, it was unacceptably mysterious how a non-reductionist about chance, for example a propensity theorist, could show that chance should constrain the agent's credence. He expressed this claim by writing:

> Be my guest—posit all the primitive unHumean whatnots you like. (...) But play fair in naming your whatnots. Don't call any alleged feature of reality "chance" unless you've already shown that you have something, knowledge of which could constrain rational credence. I think I see, dimly but well enough, how knowledge of frequencies and symmetries and best systems could constrain rational credence. I don't begin to see, for instance, how knowledge that two universals stand in a certain special relation N^* could constrain rational credence about the future coinstantiation of those universals. (Lewis 1994, p. 489)

Lewis, however, quickly realized that even a reductionist about chance cannot provide an account of chance that perfectly occupies the expert role of chance captured by the Principal Principle. For Lewis, the Big Bad Bug, which has been discussed in chapter 3, was a way to show that reductionism about chance contradicts the Principal Principle.

In exploring the expert role of chance, I left unanswered the question of whether both reductionist and non-reductionist accounts of chance could show that chance plays the expert role given by the Principal Principle. But whether or not the result established in chapter 2 will be adjacent to this issue is a matter of further research.

7.2 Resilient Chances

I have argued that it is not only the chance-credence principles that inform our understanding of chance. The thesis has also defended and put to work principles relating prior and posterior chances. But while it is widely agreed that the chance-credence principles capture the epistemic role of chance in guiding one's credences, it might initially be hard to see what role of chance is captured by these chance-chance principles. The result established in chapter 3 shows that these principles follow from a plausible norm for chances, viz. the norm of maximizing resiliency. If, as I have argued there, resilient chances play a prominent role in probabilistic explanation and prediction, then these chance-chance principles appear to tell us, albeit implicitly, a great deal about the role of chance in explanation and prediction.

Although the idea of resilient chance was not fully exploited in this thesis, I have shown already that it provides some interesting applications. It sheds new light on the debate concerning the plausibility of Humean accounts of chance (chapter 3), and also helps us justify a kinematics of chance based on Bayesian conditionalization (chapter 4).

We might reasonably expect that there is a close connection between the idea of chances as experts, explored in chapter 2, and the idea of resilient chances put forward in chapters 3 and 4. Of course, this topic demands a much fuller discussion than I can enter into here. By way of suggesting, one might try to reconcile these two ideas by requiring chances to be resilient over the agent's evidence. Then, by means of some sort of chance-credence principle, the resiliency of chances will be carried over to the resiliency of credences. That is, if resilient chances constrain the agent's credences, then they "lock" one's credences in a robust way: if one knows that the chance of a coin landing heads is $\frac{1}{2}$, then one's credence in the coin landing heads should be $\frac{1}{2}$, and should remain unchanged upon the receipt of additional evidence (Lyon 2010). A similar idea was already suggested by Lewis:

> If the evidence somehow fails to diminish your certainty that the coin is fair, then it should have no effect on the distribution of credence about outcomes that accords with that certainty about chance. To the extent that uncertainty about outcomes is based on certainty about their chances, it is a stable, resilient sort of

> uncertainty—new evidence won't get rid of it. (Lewis 1986, p. 85)

While this suggestion opens a venue for further exploration, I also believe that the idea of resilient chances and the idea of chances as experts can be understood as independent navigable channels leading to distinct places in the philosophy of probability.

In employing a particular resiliency measure for chances, I have left unanswered the question of how we can justify it. One might want to tackle this issue by isolating certain postulates for a resiliency measure and showing that the Bregman divergence, which underpins the resiliency measure, satisfies them. A similar strategy has been pursued in Joyce (1998) and in Pettigrew (2016, chapter 4) in relation to the accuracy measure for credences. Naturally, a project of this sort would require a careful consideration of what the postulates for a resiliency measure should be. I leave this project for another time.

7.3 Legal Bayesianism

Chapter 5 has shown how we can utilize a Bayesian model of legal fact-finding in order to shed new light on the topic of "naked" statistical evidence in legal proceedings. In so doing, my ambition has not been to answer the vexing question of whether this sort of evidence could license verdicts in courts of law. My view is that this question is immersed in complex issues pertaining not only to epistemology and philosophy of probability, but also to ethics and policy making. Instead, I have only argued, within a particular probabilistic model of legal proof process, that there is something epistemically defective about the fact-finder's credences that do not line up with chances.

Although it may be seen that this limitation reduces the force of the results offered in chapter 5, I believe that it enables us to make the discussion and the results far more precise. In particular, it allows us to provide a precise account of credence and verdict accuracy, and it helps us model formally the idea of using "naked" statistical evidence in legal settings.

In devising a Bayesian model in legal settings, I have left open the question of whether judges actually are Bayesians. Perhaps they are on Mondays, Wednesdays, and Fridays, but are more critical of Bayesianism on the remaining days of the week.[1] Or, it actually might be the case that Bayesianism is not a good model, for it does not encompass all the intricacies of legal fact-finding (see, e.g. Cohen 1977 and Allen 1997). To mitigate this issue, I have used the Bayesian model as a regulative ideal in the sense that the Bayesian norms should be regarded as goals toward which legal fact-finders should strive. However, it still remains to be specified how much of this ideal can by attained in actual legal proceedings.

7.4 Probabilistic Updating

In chapter 6, I have asked whether the principle of maximum relative entropy, conceived as an updating rule, leads to new credences that are expected to be helpful, and never harmful in making decisions. In answer to this question, we are left with a rather weak "it depends". Although this is not the sort of answer one might have hoped for, it points to some general lessons concerning probabilistic updating.

Among these general lessons is the thought that often a more detailed analysis of the learning situation is required to apply probabilistic updating rules. Whether or not probabilistic updating leads to the value of learning theorem depends to a large extent on how we think of our learning experience. Concomitantly, the results of chapter 6 caution against a mechanical and uncritical application of maximum relative entropy updating. Even if this updating rule promises to cover a variety of learning experiences, it should be applied only in the right circumstances.

Throughout the discussion in chapter 6, I have focused on the principle of maximum relative entropy as applied only to some types of learning experience. Naturally, one might want to obtain a more general result covering other types of learning. Specifically, one might want to define the conditions under which the principle of maximum relative entropy leads to the value of learning theorem in the most general learning setting suitable for that principle.

[1] Much like Earman (1992, p. 1).

Samenvatting

Dit proefschrift betreft een filosofische studie van twee concepten van waarschijnlijkheid en hun onderlinge relatie. Het gaat om een subjectief/persoonlijk concept genaamd ‘overtuiging’ en een objectief/fysisch concept genaamd ‘kans’. In dit proefschrift worden verschillende principes en condities geïntroduceerd en benut die gaan over relaties tussen kansen en overtuigingen, a priori kansen en a posteriori kansen, en a priori overtuigingen en a posteriori overtuigingen. Het hoofddoel is om aan te tonen dat een studie van deze principes een vruchtbare manier is om over kansen en overtuigingen na te denken. Het tweede doel is om aan te tonen hoe deze principes gecombineerd kunnen worden met enkele gevestigde argumentatieve strategieën, om zodoende inzicht te verschaffen in beide concepten van waarschijnlijkheid.

In hoofdstuk 2 wordt een antwoord ontwikkeld op de vraag of kansen ‘formeel adequaat’ zijn. Dit is het criterium dat kansen moeten voldoen aan bepaalde axioma’s van waarschijnlijkheid. Het hoofddoel van dit hoofdstuk is om te laten zien hoe beschouwingen van kans-overtuigingsinteracties gebruikt kunnen worden om de formele adequaatheid van kansen te rechtvaardigen. Hiertoe wordt een kader geïntroduceerd waarbinnen, onder redelijke aannames, aangetoond kan worden dat de expertrol, die op zichzelf bewerkstelligd wordt door een kans-overtuigingsrelatie vereist dat kansen gerepresenteerd worden door een genormeerde eindig-additieve maat.

Hoofdstuk 3 handelt over twee principes betreffende de relatie tussen a priori en a posteriori kansen. Hiertoe wordt een benadering van kansen geïntroduceerd op basis van robuustheidsprincipes. Het basisidee achter deze benadering is dat iedere kansverdeling maximaal invariant moet zijn onder wisselende experimentele factoren. Deze benadering wordt gebruikt om de twee kans-kans-principes te verdedigen. Er wordt aangetoond dat iedere kansfunctie die in strijd is met de twee principes vervangen kan worden door een robuustere kansfunctie die wel voldoet aan de principes. Vervolgens wordt er aangetoond dat, onder een Humeaanse interpretatie van kansen, deze principes gevolgen hebben die moeilijk aanvaardbaar zijn. Uitgaande van de opvatting van kansen gebaseerd op robuustheid, is het zeer de vraag of deze principes behouden dienen te worden binnen de Humeaanse benadering.

Hoofdstuk 4 handelt over de dynamica van kansen: hoe dienen kansen te veranderen over tijd? Ten eerste worden de condities onderzocht die noodzakelijk zijn voor ieder dynamisch model voor kans om te voldoen aan de Bayesiaanse dynamica voor kansen. Ten tweede wordt Lewis' argument voor Bayesiaanse dynamica bestudeerd en wordt er aangetoond dat deze dynamica volgt uit Lewis' 'principal principle' (PP). Ten derde wordt een alternatief argument voor Bayesiaanse dynamica geïntroduceerd dat geen gebruik maakt van het PP, maar van een principe dat een relatie legt tussen a priori en a posteriori kansen. Dit principe wordt gemotiveerd door een beschouwing van kansen op basis van robuustheidsprincipes, verwant aan die uit hoofdstuk 3.

Hoofdstukken 5 en 6 gaan over het overtuigingsconcept van waarschijnlijkheid. In hoofdstuk 5 wordt een eenvoudig Bayesiaans model toegepast op gerechtelijke besliskunde om te beargumenteren dat statistisch bewijs bijdraagt aan het bereiken van nauwkeurigheid, hetgeen een fundamenteel doel is binnen het recht. Er worden twee argumenten geïntroduceerd, beide op basis van het vereiste dat overtuigingen accuraat zijn, voor de stelling dat die kansen beperkingen opleggen aan de overtuigingen aangaande feitelijke hypothesen, zoals die bediscussieerd worden in de rechtszaal. Het eerste argument stelt dat wanneer de overtuigingen, na geïnformeerd te zijn over kansen, niet een lagere subjectieve verwachting van de nauwkeurigheid van een oordeel kan hebben dan vooraf. Het tweede argument laat zien dat de overtuigingen, na geïnformeerd te zijn over kansen, de objectieve verwachte nauwkeurigheid van overtuigingen maximaliseert. De begrippen 'subjectieve verwachting van de nauwkeurigheid van een oordeel' en 'objectieve verwachte nauwkeurigheid van overtuigingen' worden nauwkeurig uitgelegd binnen een Bayesiaans model voor gerechtelijke beslissingen. Dit model induceert bovendien een kans-overtuigingsprincipe dat het idee om statistisch bewijs in een rechtszaak te gebruiken van een kader voorziet.

In hoofdstuk 5 wordt geconcentreerd op synchronische eisen voor overtuigingen (dat wil zeggen, er wordt gekeken naar verscheidene beperkingen op overtuigingen van een agent op een gegeven punt in de tijd). In hoofdstuk 6 wordt daarentegen de nadruk gelegd op de dynamica van overtuigingen (dat wil zeggen, er wordt gekeken naar de verscheidene manieren waarop een een agent na verloop van tijd haar overtuigingen zal moeten aanpassen). De mogelijkheid voor een rechtvaardiging van het principe van maximale relatieve entropie als een regel om overtuigingen aan te passen op basis van nieuwe informatie wordt onderzocht. Hiertoe wordt gekeken naar een stelling uit de klassieke beslistheorie over het bepalen van conditionele overtuigingen. Deze stelling behelst een intuïtieve eis voor leren: leren zou moeten leiden tot nieuwe overtuigingen waarvan verwacht wordt dat ze behulpzaam zijn bij het maken van keuzes. In dit proefschrift wordt deze eis de ‘waarde van leren’ genoemd. Er wordt onderzocht in hoeverre het principe van maximale relatieve entropie voldoet aan de waarde van leren en of het een rationele methode is voor het navolgen van praktische doelen. Gaandeweg wordt een langdurig dispuut in de waarschijnlijkheidstheoretische kenleer besproken, namelijk, de vraag of er een universele regel is voor het aanpassen van overtuigingen op basis van nieuwe informatie.

Curriculum Vitae

Patryk Dziurosz-Serafinowicz was born in Złotów, Poland, on the 31st of October in 1982. He graduated in philosophy and law at the Jagiellonian University in Kraków. In 2011, he obtained a doctorate in law, with specialization in legal theory, at the Jagiellonian University in Kraków. Since 2011 he has worked on this thesis to obtain a doctorate in philosophy at the University of Groningen.

Bibliography

Allen, R. J. (1997). Rationality, Algorithms and Juridical Proof: A Preliminary Inquiry. *The International Journal of Evidence & Proof*, 1(Special Issue):254–275.

Allen, R. J. and Pardo, M. S. (2007). The Problematic Value of Mathematical Models of Evidence. *The Journal of Legal Studies*, 36(1):107–140.

Armendt, B. (1993). Dutch Books, Additivity, and Utility Theory. *Philosophical Topics*, 21(1):1–20.

Arntzenius, F. and Hall, N. (2003). On What We Know about Chance. *British Journal for the Philosophy of Science*, 54(2):171–179.

Banerjee, A., Merugu, S., Dhillon, I. S., and Ghosh, J. (2005). Clustering with Bregman Divergences. *Journal of Machine Learning Research*, 6:1705–1749.

Bigelow, J., Collins, J., and Pargetter, R. (1993). The Big Bad Bug: What are the Humean's Chances? *British Journal for the Philosophy of Science*, 44(3):443–462.

Billingsley, P. (1965). *Ergodic Theory and Information.* New York: John Wiley and Sons.

Blackstone, W. ([1765]/2002). *Commentaries on the Laws of England*, volume IV. Chicago: University of Chicago Press.

Blattenberger, G. and Lad, F. (1985). Separating the Brier Score into Calibration and Refinement Components: A Graphical Exposition. *The American Statistician*, 39(1):26–32.

Blome-Tillmann, M. (2015). Sensitivity, Causality, and Statistical Evidence in Courts of Law. *Thought: A Journal of Philosophy*, 4(2):102–112.

Bradley, R. (2005). Radical Probabilism and Bayesian Conditioning. *Philosophy of Science*, 72(2):342–364.

Braithwaite, R. (1966). Why is It Reasonable to Base a Betting Rate Upon an Estimate of Chance. In Bar-Hillel, Y., editor, *Proceedings of the 1964 International Congress for Logic,Methodology and Philosophy of Science*, pages 263–273. Amsterdam: New Holland.

Brier, G. (1950). Verification of Forecasts in Terms of Probability. *Monthly Weather Review*, 78(1):1–3.

Briggs, R. (2009a). The Anatomy of the Big Bad Bug. *Noûs*, 43(3):428–449.

Briggs, R. (2009b). The Big Bad Bug Bites Anti-Realists about Chance. *Synthese*, 167(1):81–92.

Brown, P. M. (1976). Conditionalization and Expected Utility. *Philosophy of Science*, 43(3):415–419.

Buchak, L. (2013). Belief, Credence, and Norms. *Philosophical Studies*, 169(2):1–27.

Carnap, R. (1962). *Logical Foundations of Probability.* Chicago: University of Chicago Press.

Christensen, D. (2004). *Putting Logic in its Place: Formal Constraints on Rational Belief.* Oxford: Oxford University Press.

Cohen, L. J. (1977). *The Probable and the Provable.* Oxford: Clarendon Press.

Cohen, L. J. (1981). Subjective Probability and the Paradox of the Gatecrasher. *Arizona State Law Journal*, 12(2):627–634.

Colyvan, M., Regan, H. M., and Ferson, S. (2001). Is It a Crime to Belong to a Reference Class. *Journal of Political Philosophy*, 9(2):168–181.

Damaska, M. (1975). Presentation of Evidence and Factfinding Precision. *University of Pennsylvania Law Review*, 123(1):1083–1106.

Dawid, A. P. (2002). Bayes's Theorem and Weighing Evidence by Juries. *Proceedings of the British Academy*, 113:71–90.

De Finetti, B. (1970). *Theory of Probability*, volume 1. New York: John Wiley.

DeGroot, M. H. and Fienberg, S. E. (1983). The Comparison and Evaluation of Forecasters. *The Statistician*, 32(1/2):14–22.

Diaconis, P. and Zabell, S. L. (1982). Updating Subjective Probability. *Journal of the American Statistical Association*, 77(380):822–830.

Domotor, Z. (1981). Higher Order Probabilities. *Philosophical Studies*, 40(1):31–46.

Douven, I. and Romeijn, J.-W. (2011). A New Resolution of the Judy Benjamin Problem. *Mind*, 120(479):637–670.

Dworkin, R. (1985). *A Matter of Principle*. Cambridge, MA: Harvard University Press.

Dziurosz-Serafinowicz, P. (2015). Maximum Relative Entropy Updating and the Value of Learning. *Entropy*, 17(3):1146–1164.

Eagle, A. (2004). Twenty-One Arguments Against Propensity Analyses of Probability. *Erkenntnis*, 60(3):371–416.

Earman, J. (1992). *Bayes or Bust?* Cambridge, MA: Bradford MIT.

Easwaran, K. (2011). Bayesianism I: Introduction and Arguments in Favor. *Philosophy Compass*, 6(5):312–320.

Easwaran, K. (2013). Expected Accuracy Supports Conditionalization—and Conglomerability and Reflection. *Philosophy of Science*, 80(1):119–142.

Eells, E. (1983). Objective Probability Theory Theory. *Synthese*, 57(3):387–442.

Elga, A. (2007). Reflection and Disagreement. *Noûs*, 41(3):478–8211.

Enoch, D., Spectre, L., and Fisher, T. (2012). Statistical Evidence, Sensitivity, and the Legal Value of Knowledge. *Philosophy and Public Affairs*, 40(3):197–224.

Fisher, J. C. (2006). On Higher-Order and Free-Floating Chances. *British Journal for the Philosophy of Science*, 57(4):691–707.

Friedman, K. and Shimony, A. (1971). Jaynes' Maximum Entropy Prescription and Probability Theory. *Journal of Statistical Physics*, 4(3):381–384.

Frigg, R. and Hoefer, C. (2015). The Best Humean System for Statistical Mechanics. *Erkenntnis*, 80(3):551–574.

Gaifman, H. (1988). A Theory of Higher Order Probabilities. In Skyrms, B. and Harper, W., editors, *Causation, Chance, and Credence*, volume 1, pages 191–219. Dordrecht: Kluwer Academic Publishers.

Giere, R. N. (1973). Objective Single-Case Probabilities and the Foundations of Statistics. In Suppes, P., Henkin, L., Moisil, G., and Joja, A., editors, *Logic, Methodology, and Philosophy of Science IV: Proceedings of the Fourth International Congress for Logic, Methodology and Philosophy of Science, Bucharest, 1971*, volume 74, pages 467–483. Amsterdam: North Holland.

Giere, R. N. (1976). A Laplacean Formal Semantics for Single-Case Propensities. *Journal of Philosophical Logic*, 5(3):321–353.

Gillies, D. (2000). *Philosophical Theories of Probability.* London: Routledge.

Goldman, A. I. (1999). *Knowledge in a Social World.* Oxford: Clarendon Press.

Goldman, A. I. (2002). Quasi-Objective Bayesianism and Legal Evidence. *Jurimetrics*, 42(3):237–260.

Good, I. J. (1950). *Probability and the Weighing of Evidence.* London: Charles Griffin and Company.

Good, I. J. (1967). On the Principle of Total Evidence. *British Journal for the Philosophy of Science*, 17(4):319–321.

Graves, P. R. (1989). The Total Evidence Theorem for Probability Kinematics. *Philosophy of Science*, 56(2):317–324.

Greaves, H. and Wallace, D. (2006). Justifying Conditionalization: Conditionalization Maximizes Expected Epistemic Utility. *Mind*, 115(459):607–632.

Grünwald, P. (2000). Maximum Entropy and the Glasses You are Looking Through. In *Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence (UAI 2000)*, pages 238–246. San Fransisco: Morgan Kaufmann Publishers.

Grünwald, P. D. and Halpern, J. Y. (2003). Updating Probabilities. *Journal of Artificial Intelligence Research*, 19(1):243–278.

Hacking, I. (1975). *The Emergence of Probability*. Cambridge: Cambridge University Press.

Hájek, A. (1996). "Mises Redux"–Redux: Fifteen Arguments Against Finite Frequentism. *Erkenntnis*, 45(2/3):209–227.

Hájek, A. (2007). The Reference Class Problem is Your Problem Too. *Synthese*, 156(3):563–585.

Hájek, A. (2009). Fifteen Arguments Against Hypothetical Frequentism. *Erkenntnis*, 70(2):211–235.

Hall, N. (1994). Correcting the Guide to Objective Chance. *Mind*, 103(412):505–518.

Hall, N. (2004). Two Mistakes About Credence and Chance. *Australasian Journal of Philosophy*, 82(1):93–111.

Hamer, D. (1994). The Civil Standard of Proof Uncertainty: Probability, Belief and Justice. *The Sydney Law Review*, 16(4):506–536.

Hamer, D. (2012). Probability, Anti-Resilience, and the Weight of Expectation. *Law, Probability and Risk*, 11(2-3):135–158.

Ho, H. L. (2008). *A Philosophy of Evidence Law: Justice in the Search for Truth.* Oxford: Oxford University Press.

Hoefer, C. (2007). The Third Way on Objective Probability: A Sceptic's Guide to Objective Chance. *Mind*, 116(463):549–596.

Howson, C. and Urbach, P. (2006). *Scientific Reasoning. The Bayesian Approach.* La Salle, Illinois: Open Court.

Huttegger, S. M. (2014). Learning Experiences and the Value of Knowledge. *Philosophical Studies*, 171(2):279–288.

Ismael, J. (1996). What Chances Could Not Be. *British Journal for the Philosophy of Science*, 47(1):79–91.

Ismael, J. (2008). Raid! Dissolving the Big, Bad Bug. *Noûs*, 42(2):292–8211.

Jaynes, E. T. (1957). Information Theory and Statistical Mechanics. *The Physical Review*, 106(4):620–630.

Jeffrey, R. (1983). *The Logic of Decision.* Chicago: University of Chicago Press.

Jeffrey, R. (1988). Conditioning, Kinematics, and Exchangeability. In Skyrms, B. and Harper, W. L., editors, *Causation, Chance and Credence*, volume 1, pages 221–255. Dordrecht: Kluwer Academic Publishers.

Jeffreys, H. (1961). *Theory of Probability.* Oxford: Oxford University Press.

Joyce, J. M. (1998). A Nonpragmatic Vindication of Probabilism. *Philosophy of Science*, 65(4):575–603.

Joyce, J. M. (2009). Accuracy and Coherence: Prospects for an Alethic Epistemology of Partial Belief. In Huber, F. and Schmidt-Petri, C., editors, *Degrees of Belief*, pages 263–297. Dordrecht: Springer Netherlands.

Kaye, D. H. (1988). What is Bayesianism? A Guide for the Perplexed. *Jurimetrics*, 28(Winter 1988):161–177.

Keynes, J. M. (1921). *A Treatise on Probability.* New York: Dover Publications.

Koehler, J. and Shaviro, D. (1990). Veridical Verdicts: Increasing Verdict Accuracy Through the Use of Overtly Probabilistic Evidence and Methods. *Cornell Law Review*, 246(75):247–279.

Koehler, J. J. (1991). The Probity/Policy Distinction in the Statistical Evidence Debate. *Tulane Law Review*, 66(1):141–150.

Kyburg, H. E. (1974). Propensities and Probabilities. *British Journal for the Philosophy of Science*, 25(4):358–375.

Lange, M. (2006). Do Chances Receive Equal Treatment Under the Laws? Or: Must Chances Be Probabilities? *British Journal for the Philosophy of Science*, 57(2):383–403.

Laplace, P. ([1814]/1951). *A Philosophical Essay on Probabilities.* New York: Dover Publications.

Leitgeb, H. and Pettigrew, R. (2010a). An Objective Justification of Bayesianism I: Measuring Inaccuracy. *Philosophy of Science*, 77(2):201–235.

Leitgeb, H. and Pettigrew, R. (2010b). An Objective Justification of Bayesianism II: The Consequences of Minimizing Inaccuracy. *Philosophy of Science*, 77(2):236–272.

Lempert, R. (1977). Modeling Relevance. *Michigan Law Review*, 75(5/6):1021–1057.

Levins, R. (1966). The Strategy of Model Building in Population Biology. *American Scientist*, 54(4):421–431.

Lewis, D. (1970). How to Define Theoretical Terms. *Journal of Philosophy*, 67(13):427–446.

Lewis, D. (1983). New Work for a Theory of Universals. *Australasian Journal of Philosophy*, 61(4):343–377.

Lewis, D. (1986). A Subjectivist's Guide to Objective Chance. In *Philosophical Papers*, volume 2, pages 83–132. Oxford: Oxford University Press.

Lewis, D. (1994). Humean Supervenience Debugged. *Mind*, 103(412):473–490.

Lewis, D. (1999). Why Conditionalize? In *Papers in Metaphysics and Epistemology*, volume 2, pages 403–407. Cambridge: Cambridge University Press.

Loewer, B. (2001). Determinism and Chance. *Studies in History and Philosophy of Science Part B*, 32(4):609–620.

Loewer, B. (2004). David Lewis's Humean Theory of Objective Chance. *Philosophy of Science*, 71(5):1115–1125.

Lyon, A. (2010). Deterministic Probability: Neither Chance Nor Credence. *Synthese*, 182(3):413–432.

Maher, P. (1992). Diachronic Rationality. *Philosophy of Science*, 59(1):120–141.

Meacham, C. J. G. (2010). Two Mistakes Regarding the Principal Principle. *British Journal for the Philosophy of Science*, 61(2):407–431.

Mellor, D. H. (1982). Chance and Degrees of Belief. In McLaughlin, R., editor, *What? Where? When? Why?*, pages 49–68. Dordrecht: Reidel.

Miller, D. (1966). A Paradox of Information. *British Journal for the Philosophy of Science*, 17(1):59–61.

Milne, P. (1987). Physical Probabilities. *Synthese*, 73(2):329–359.

Mitchell, S. D. (2000). Dimensions of Scientific Law. *Philosophy of Science*, 67(2):242–265.

Morgan, E. M. (1948). Hearsay Dangers and the Application of the Hearsay Concept. *Harvard Law Review*, 62(2):pp. 177–219.

Murphy, A. H. (1973). A New Vector Partition of the Probability Score. *Journal of Applied Meteorology*, 12:595–600.

Nance, D. A. (2008). The Weights of Evidence. *Episteme*, 5(3):267–281.

Ornstein, D. (1974). *Ergodic Theory, Randomness, and Dynamical Systems*. New Haven and London: Yale University Press.

Peijnenburg, J. and Atkinson, D. (2012). An Endless Hierarchy of Probabilities. *American Philosophical Quarterly*, 49(3):267–276.

Pettigrew, R. (2012). Accuracy, Chance, and the Principal Principle. *Philosophical Review*, 121(2):241–275.

Pettigrew, R. (2013a). A New Epistemic Utility Argument for the Principal Principle. *Episteme*, 10(1):19–35.

Pettigrew, R. (2013b). Accuracy and Evidence. *Dialectica*, 67(4):579–596.

Pettigrew, R. (2013c). Epistemic Utility and Norms for Credences. *Philosophy Compass*, 8(10):897–908.

Pettigrew, R. (2016). *Accuracy and the Laws of Credence.* Oxford: Oxford University Press.

Pettigrew, R. and Titelbaum, M. G. (2014). Deference Done Right. *Philosophers' Imprint*, 14(35):1–19.

Popper, K. R. (1959). The Propensity Interpretation of Probability. *British Journal for the Philosophy of Science*, 10(37):25–42.

Predd, J., Seiringer, R., Lieb, E., Osherson, D., Poor, H. V., and Kulkarni, S. (2009). Probabilistic Coherence and Proper Scoring Rules. *IEEE Transactions on Information Theory*, 55(10):4786–4792.

Pundik, A. (2011). The Epistemology of Statistical Evidence. *The International Journal of Evidence & Proof*, 15(2):117–143.

Ramsey, F. P. (1931). Truth and Probability. In Braithwaite, R. B., editor, *Foundations of Mathematics and Other Logical Essays.* London: Routledge and P. Kegan.

Ramsey, F. P. (1990). Weight or the Value of Knowledge. *The British Journal for the Philosophy of Science*, 41(1):1–4.

Redmayne, M. (1998). Bayesianism and Proof. In Freeman, M. and Reece, H., editors, *Science in Court*, pages 61–81. Aldershot: Ashgate.

Redmayne, M. (2008). Exploring the Proof Paradoxes. *Legal Theory*, 14(4):281–309.

Reichenbach, H. (1949). *The Theory of Probability*. Berkeley: University of California Press.

Salmon, W. C. (1967). *The Foundations of Scientific Inference*. Pittsburgh: University of Pittsburgh Press.

Savage, L. J. (1954). *The Foundations of Statistics*. New York: John Wiley and Sons.

Schaffer, J. (2003). Principled Chances. *British Journal for the Philosophy of Science*, 54(1):27–41.

Schaffer, J. (2007). Deterministic Chance? *British Journal for the Philosophy of Science*, 58(2):113–140.

Schauer, F. (2003). *Profiles, Probabilities, and Stereotypes*. Cambridge, MA: Belknap Press.

Schoeman, F. (1987). Statistical vs. Direct Evidence. *Noûs*, 21(2):179–198.

Seidenfeld, T. (1986). Entropy and Uncertainty. *Philosophy of Science*, 53(4):467–491.

Selten, R. (1998). Axiomatic Characterization of the Quadratic Scoring Rule. *Experimental Economics*, 1(1):43–61.

Shimony, A. (1970). Scientific Inference. In Colodny, R., editor, *The Nature and Function of Scientific Theories*, pages 79–172. Pittsburgh: University of Pittsburgh Press.

Shimony, A. (1985). The Status of the Principle of Maximum Entropy. *Synthese*, 63(1):35–53.

Shimony, A. (1988). An Adamite Derivation of the Calculus of Probability. In Fetzer, J. H., editor, *Probability and Causality*, pages 79–89. Dordrecht: D. Reidel.

Shore, J. E. and Johnson, R. W. (1980). Axiomatic Derivation of the Principle of Maximum Entropy and the Principle of Minimum Cross-Entropy. *IEEE Transactions on Information Theory*, 26(1):26–37.

Shore, J. E. and Johnson, R. W. (1981). Properties of Cross-Entropy Minimization. *Information Theory, IEEE Transactions on*, 27(4):472–482.

Skyrms, B. (1977). Resiliency, Propensities, and Causal Necessity. *The Journal of Philosophy*, 74(11):704–713.

Skyrms, B. (1978). Statistical Laws and Personal Propensities. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 1978:551–562.

Skyrms, B. (1980a). *Causal Necessity: A Pragmatic Investigation of the Necessity of Laws.* New Haven and London: Yale University Press.

Skyrms, B. (1980b). Higher Order Degrees of Belief. In Mellor, D. H., editor, *Prospects for Pragmatism*, pages 109–137. Cambridge: Cambridge University Press.

Skyrms, B. (1984). *Pragmatism and Empiricism.* New Haven and London: Yale University Press.

Skyrms, B. (1985). Maximum Entropy Inference as a Special Case of Conditionalization. *Synthese*, 63(1):55–74.

Skyrms, B. (1987a). Dynamic Coherence and Probability Kinematics. *Philosophy of Science*, 54(1):1–20.

Skyrms, B. (1987b). Updating, Supposing, and Maxent. *Theory and Decision*, 22(3):225–246.

Skyrms, B. (1990). *The Dynamics of Rational Deliberation.* Cambridge: Harvard University Press.

Skyrms, B. (1991). Stability and Chance. In Skyrms, B., Spohn, W., and van Fraassen, B., editors, *Existence and Explanation*, pages 149–163. Dordrecht, Boston: Kluwer Academic Publishers.

Skyrms, B. (1997). The Structure of Radical Probabilism. *Erkenntnis*, 45(2-3):285–297.

Sober, E. (2010). Evolutionary Theory and the Reality of Macro Probabilities. In Eells, E. and Fetzer, J., editors, *The Place of Probability in Science*, volume 284, pages 133–60. Dordrecht: Springer Netherlands.

Stein, A. (2005). *Foundations of Evidence Law*. Oxford: Oxford University Press.

Strevens, M. (1995). A Closer Look at the 'New' Principle. *British Journal for the Philosophy of Science*, 46(4):545–561.

Strevens, M. (1999). Objective Probability as a Guide to the World. *Philosophical Studies*, 95(3):243–275.

Strevens, M. (2003). *Bigger Than Chaos: Understanding Complexity Through Probability*. Cambridge, MA: Harvard University Press.

Suppes, P. (1973). New Foundations of Objective Probability: Axioms for Propensities. In Suppes, P., Henkin, L., Moisil, G., and Joja, A., editors, *Logic, Methodology, and Philosophy of Science IV: Proceedings of the Fourth International Congress for Logic, Methodology and Philosophy of Science, Bucharest, 1971*, volume 74, pages 515–529. Amsterdam: North Holland.

Suppes, P. (1974). Popper's Analysis of Probability in Quantum Mechanics. In Schilpp, P. A., editor, *The Philosophy of Karl Popper*, volume 2, pages 760–774. La Salle: Open Court.

Teller, P. (1973). Conditionalization and Observation. *Synthese*, 26(2):218–258.

Thau, M. (1994). Undermining and Admissibility. *Mind*, 103(412):491–504.

Thomson, J. J. (1986). Liability and Individualized Evidence. *Law and Contemporary Problems*, 49(3):199–219.

Tversky, A. and Kahneman, D. (1977). Causal Thinking in Judgment under Uncertainty. In Butts, R. and Hintikka, J., editors, *The University of Western Ontario Series in Philosophy of Science: Basic Problems in Methodology and Linguistics*, volume 11, pages 167–190. Dordrecht-Holland: D. Reidel.

Tversky, A. and Kahneman, D. (1982). Evidential Impact of Base Rates. In Kahneman, D., Slovic, A., and Tversky, A., editors, *Judgment under Uncertainty: Heuristics and Biases*, pages 153–160. Cambridge: Cambridge University Press.

Uffink, J. (1996). The Constraint Rule of the Maximum Entropy Principle. *Studies in History and Philosophy of Science Part B*, 27(1):47–79.

Uffink, J. (2011). Subjective Probability and Statistical Physics. In Beisbart, C. and Hartmann, S., editors, *Probabilities in Physics*, pages 25–49. Oxford: Oxford University Press.

van Fraassen, B. (1979). Foundations of Probability: A Modal Frequency Interpretation. In di Francia, G. T., editor, *Problems in the Foundations of Physics*, pages 344–394. Amsterdam: North-Holland.

van Fraassen, B. (1981). A Problem for Relative Information Minimizers in Probability Kinematics. *British Journal for the Philosophy of Science*, 32(4):375–379.

van Fraassen, B. (1983). Calibration: A Frequency Justification for Personal Probability. In Cohen, R. and Laudan, L., editors, *Physics, Philosophy, and Psychoanalysis*, pages 295–319. Dordrecht: D. Reidel.

van Fraassen, B. (1984). Belief and the Will. *Journal of Philosophy*, 81(5):235–256.

van Fraassen, B. (1989). *Laws and Symmetry*. Oxford: Oxford University Press.

van Fraassen, B. (1995). Belief and the Problem of Ulysses and the Sirens. *Philosophical Studies*, 77(1):7–37.

van Fraassen, B. (1999). Conditionalization, a New Argument For. *Topoi*, 18(2):93–96.

van Fraassen, B., Hughes, R. I. G., and Harman, G. (1986). A Problem for Relative Information Minimizers, Continued. *British Journal for the Philosophy of Science*, 37(4):453–463.

Venn, J. (1866). *The Logic of Chance.* London and Cambridge: Macmillan.

von Mises, R. (1957). *Probability, Truth and Statistics.* New York: Macmillan.

Vranas, P. B. M. (2002). Who's Afraid of Undermining? *Erkenntnis*, 57(2):151–174.

Wasserman, D. (1992). The Morality of Statistical Proof and the Risk of Mistaken Liability. *Cardozo Law Review*, 13:935–976.

Weisberg, J. (2007). Conditionalization, Reflection, and Self-Knowledge. *Philosophical Studies*, 135(2):179–97.

Weisberg, J. (2011). Varieties of Bayesianism. In Gabbay, D., Hartmann, S., and Woods, J., editors, *Handbook of the History of Logic*, volume 10, pages 477–551. Amsterdam/New York: North Holland/Elsevier.

Williams, P. M. (1980). Bayesian Conditionalisation and the Principle of Minimum Information. *British Journal for the Philosophy of Science*, 31(2):131–144.

Williamson, J. (2010). *In Defence of Objective Bayesianism.* Oxford: Oxford University Press.

Woodward, J. (2003). *Making Things Happen.* Oxford: Oxford University Press.

Wright, R. W. (1988). Causation, Responsibility, Risk, Probability, Naked Statistics, and Proof: Pruning the Bramble Bush by Clarifying the Concepts. *Iowa Law Review*, 73:1001–1077.