Probabilistic Stability:

dynamics, nonmonotonic logics, and stable revision

MSc Thesis *(Afstudeerscriptie)*

written by

Krzysztof Mierzewski
(born May 12, 1989 in Gdańsk, Poland)

under the supervision of **Dr. Alexandru Baltag**, and submitted to the Board of Examiners in partial fulfillment of the requirements for the degree of

MSc in Logic

at the *Universiteit van Amsterdam*.

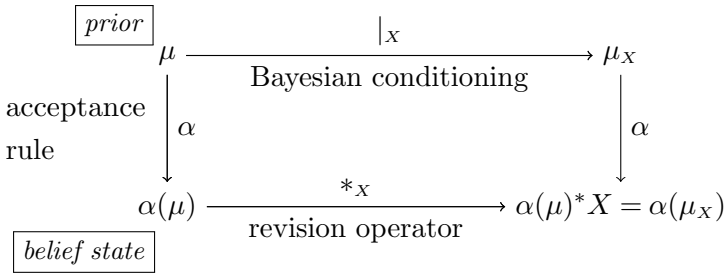
Date of the public defense:	**Members of the Thesis Committee:**
June 18th, 2018	Dr. Alexandru Baltag (*Supervisor*)
	Prof. Dr. Johan van Benthem
	Dr. Floris Roelofsen (*Chair*)
	Dr. Katrin Schulz

Institute for Logic, Language and Computation

Abstract

Leitgeb [30] proposes an acceptance rule based on the notion of *probabilistically stable* hypotheses. This *stability rule* offers a formal solution to the Lottery Paradox and suggests a promising account of the relationship between logical and probabilistic models of belief. In this thesis, we investigate the role of probabilistic stability in bridging logical information dynamics – modeled by revision operators – with probabilistic models of belief change, as captured by Bayesian conditioning.

Our first topic is the connection between Bayesian conditioning and AGM revision operators. The gold standard of dynamic compatibility between a logical revision operator and Bayesian conditioning is given by the *tracking* criterion, which amounts to the requirement that the revision operator commmute with Bayesian update modulo the acceptance rule.



A general impossibility theorem by Lin & Kelly [25] shows that no well-behaved acceptance rule allows AGM operators to track Bayesian update. We show that, even though Leitgeb's stability rule falls prey to Lin & Kelly's theorem, there is nonetheless a precise sense in which it allows to bridge AGM revision and conditioning. We establish this by appealing to notions from information theory: by an application of the principle of maximum entropy, we show that AGM revision operators can be generated, through Leitgeb's rule, by Bayesian conditioning. In situations of information loss, AGM revision is compatible with – and indeed emerges from – Bayesian conditioning.

Another approach to the tracking problem is to axiomatise the revision operators

generated by the stability rule: the study of these *probabilistically stable revision operators* constitutes our second topic. We show that the class of probabilistically stable revision operators can be captured using selection function models, as employed in non-monotonic logics. We first identify the key properties of the resulting non-monotonic logic. We then prove a probabilistic representation theorem for the selection function models in question. The theorem, which draws on the theory of comparative probability orders, yields a complete characterisation of probabilistically stable revision operators. Along the way, we prove a general result giving sufficient conditions for the joint representation of a pair of (respectively, strict and non-strict) comparative probability orders, and we point out an application of the representation theorem to simple voting games.

Babci Oldze, która wie, jak budować i jednoczyć, dzieląc się darem rozmowy i rozumienia,
a której ciepło i skromność podkreślają inne cnoty jej właściwe:
mądrość, wytrwałość,
spokój wszechogarniający,
życzliwość ludziom.

Contents

List of Figures

Acknowledgements

For a long time, this thesis has been on the back burner (and then on the back burner's back burner), while other projects came along and took center stage. Now that it is time to cross the $\top$'s and dot the ϕ_i's, I would like to thank Alexandru Baltag for his constant encouragement, even at times when there was only scant evidence of the stove still being on. Working with Alexandru has been a learning experience of the very finest kind. I consider myself unreasonably lucky to have been exposed to his creative energy in doing research: his drive, sharp curiosity, and infectious enthusiasm have brightened many a day by reviving in me the pleasure of finding things out. I owe him a debt of gratitude: my parting memory of Amsterdam might have been much different were it not for his kindness, patience and support in matters both academic and personal.

My warmest thanks also go to Sonja Smets and Johan van Benthem for inspiring lectures and discussions, connecting the world of logic to its wider mathematical and philosophical horizons. It was a joy to survey the depth and breadth of logical methods through Sonja's guided excursions into proofs and paradoxes, probability problems and quantum quandaries, and all manner of questions formal and foundational. Johan, ever a master at recursive clauses, taught me much logic, and taught me how to teach logic; I hope one day to be able to teach to teach to teach as well as he does. Communication is indeed all about not getting lost in translation (standard and non-), and it is always a privilege to witness the ease with which Johan finds not only the right word, but also the *bon mot*. This is as it should be: who but logicians to cultivate *le sens de la formule*?

Infinite th$\aleph_1$nks to Tanja Kassenaar, who stepped in uncountably many times to help with all sorts of matters. Thanks to Floris Roelofsen and Katrin Schulz for kindly joining my thesis committee and taking the time to read through these pages. Thanks to Thomas Icard for prodigiously productive discussions, invariably insightful comments, and unfailingly helpful advice.

Thanks to friends in Amsterdam who made life good and dinners delicious: Andrea, Michele, Margherita, Giovanni, and Mathias. Thanks to London friends who came to Amsterdam to visit – Teddy and Maria, Mikołaj and Tom – and to those – Rhiannon – who

then came to stay. Thanks also to those who intended to visit but didn't (though meeting in Ghent was a good compromise, Aaron).

My time in Amsterdam was one of intellectual excitement and discovery, but was also marked by family illness: in the midst of all this, my parents and my sister Natalia have been a beacon of light. I am constantly impressed by their unwavering compassion, their courage, and their salutary ability to take a step back and look at matters from an ironic distance – the healthy skill of taking even the most bitter pills to swallow with a pinch of salt. Their resolve demonstrates how to stay resilient in the face of what is not stable. I am grateful, more than this page or any margin can contain, for their caring love and support (*hanc marginis exiguitas non caperet*). Dziękuję Wam, ptaszki starsze; dziękuję, Robalu.

Fafa, thank you for keeping the things most precious invariant under all transformations.

1

Introduction

When a Bayesian agent reasons about the world, she formulates a probability model within which she evaluates the probabilities of various hypotheses. Such a model involves a probability measure over an algebra of events which is taken to reflect the agent's credences, or degrees of belief. This quantitative notion of belief plays an important role in Bayesian statistics, game and decision theory, and in probabilistic approaches to artificial intelligence. On the other hand, in models of belief employed in applied logic, logic-based artificial intelligence, and traditional epistemology, the central notion is that of *qualitative* belief, generally taken to be a coarser, all-or-nothing attitude: a proposition is either believed or not.

It is natural to think of 'all-or-nothing' belief as being a coarse-grained analogue of the quantitative representation of belief, as provided by the Bayesian account. In spite of this intuitive connection, the formal details of the relationship between qualitative and quantitative belief have proven to be rather elusive. How exactly does the probabilistic information encoding a rational agent's credences get translated into the categorical information representing her qualitative beliefs? Can this be done in a way that satisfies some very elementary desiderata for beliefs of rational agents – e.g., consistency, or closure under logical consequence?

An analogous question arises in the context of statistical reasoning. Much of statistical theory – particularly so in areas concerned with statistical hypothesis testing – is aimed at determining which hypotheses should be believed, or accepted, and which hypotheses should be rejected on the basis of the available probabilistic information. Acceptance and rejection of hypotheses are qualitative, categorical notions. Is there a general recipe for rationally extracting such qualitative content from probabilistic information? Or, in slightly different terms, can we provide some lossy – yet reasonably well-behaved – qualitative description of statistical information in terms of all-or-nothing commitment to hypotheses?

One prominent way to approach these questions is through the study of *acceptance rules*. An acceptance rule is a map which assigns to each probability model a collection of propositions that the agent accepts. It can be seen as a systematic method to extract the essential qualitative content of uncertain information, or as a rigorous model of the functional dependence between a rational agent's credences and her propositional belief state. Acceptance rules thus provide a simple and natural mathematical framework to elucidate the logic(s) of uncertain acceptance.

Although providing a reasonable acceptance rule has been fraught with difficulties – as evidenced by the ever-growing literature on Kyburg's Lottery paradox [29] – several recent proposals have opened up some promising avenues [26, 30]. Notable among these is Leitgeb's *stability rule* [30, 32, 33], which is based on the notion of probabilistic stability (itself adapted from Skyrms' notion of probabilistic resiliency [53]). The key idea is that accepted hypotheses ought to be resilient, or stable, under new information. The rule is promising in that it succeeds in preserving some intuitions behind the Lockean rule (which recommends the acceptance of all and only propositions with probability above a fixed threshold) while avoiding the Lottery paradox: it also allows to preserve the closure of accepted propositions under logical consequence.

The search for well-behaved acceptance rules raises several methodological questions. What are reasonable desiderata for acceptance rules? Which transformations on the underlying probability models should acceptance rules be sensitive to, and which ones should they be invariant under? How should acceptance rules relate to our policies for updating our beliefs in the face of new information? Can acceptance rules reconcile the differences between probabilistic and logical models of uncertainty and information dynamics? How can we guarantee that an acceptance rule allows successful inductive learning?

These methodological concerns raise several logical and mathematical questions about acceptance rules, and the stability rule in particular. In this thesis we explore some of these aspects of probabilistic stability and the stability rule for acceptance: we will appeal to tools and perspectives from various areas (logic, belief revision theory, as well as probability and information theory) to solve certain formal questions that naturally arise in investigating the behaviour of acceptance rules. We will address some of the underlying methodological and philosophical concerns along the way.

From a more general perspective, part of the motivation for the present work is a fundamental interest in the connections between logic and probability theory. Probability and logic interact in a variety of fascinating ways that continue to stimulate much mathematical and philosophical research. This thesis investigates the relationship between logic and probability in the context of studying the dynamics of informational states: the particular focus is on formal models of probabilistic learning and logical accounts of belief dynamics.

Our results are situated along two main lines of research, both of which concern distinct aspects of the relationship between probabilistic and qualitative accounts of belief dynamics. The first one is the connection between Bayesian conditioning and AGM revision operators. We appeal to elementary notions from information theory to bridge the two: by an application of the principle of maximum entropy, we show that we can see AGM revision as emerging from Bayesian conditioning.

The second line of inquiry concerns studying the logical revision operators and the non-monotonic logic (or conditional doxastic logic) *generated by* the stability rule. The stability rule generates a qualitative revision operator that automatically commutes with Bayesian conditioning. Here we capture the resulting revision operation using selection function models, as employed in non-monotonic logics. We identify the key properties of the resulting logic. We draw on the theory of comparative probability orders to give a probabilistic representation theorem for selection function models. This gives a complete characterisation of probabilistically stable revision operators.

The specific content of each chapter is outlined below in more detail.

1.1 Thesis outline

In Chapter 2, we begin our investigation into the relationship between qualitative belief revision operators and Bayesian conditioning. In particular, we study how AGM belief revision operators can be related to Bayesian conditioning via Leitgeb's acceptance rule, in order to flesh out some (in)compatibilities between Bayesian and AGM-compliant models of rational belief dynamics. Our starting point is Lin and Kelly's No-Go Theorem [25] which entails that, in a precise sense, AGM revision operators do not agree with Bayesian conditioning under any acceptance rule which satisfies some modest requirements. Leitgeb's rule in particular falls prey to this No-Go Theorem, but it has nonetheless been argued to offer hope for a reconciliation between Bayesian and AGM dynamics [30]. We consider some ways in which one may circumvent the No-Go Theorem so as to approximate agreement between AGM and Bayesian conditioning, using Leitgeb's rule. We show that threshold raising, a very natural idea in this context, fails; as we argue, this failure raises further difficulties for the "peace project" between Bayesian and AGM-compliant revision operators. However, we also show how an information-theoretic perspective allows to derive a close connection between them: there is a precise sense in which AGM revision can be seen as deriving from (1) Leitgeb's rule, (2) Bayesian conditioning, and (3) a version of the maximum entropy principle. This suggests that one could study qualitative revision operators as special cases of Bayesian reasoning which naturally arise in situations of information loss or incomplete probabilistic specification of the agent's credal state.

In Chapter 3, we approach the problem of bridging qualitative and probabilistic dynamics from a different perspective: instead of trying to harmonise AGM revision operators with Bayesian conditioning via the notion of probabilistic stability, we consider the "dual" problem of characterising the qualitative revision operators generated *from* Bayesian conditioning by Leitgeb's rule. This automatically yields a revision operation – *probabilistically stable revision* – which commutes with conditioning. We identify certain key properties of the revision generated by Leitgeb's rule and briefly compare its behaviour to that of AGM operators. We then investigate the problem of giving a purely qualitative description of stability-based revision. Firstly, we formulate the problem in the framework of non-monotonic logic: we want to identify the non-monotonic consequence relations corresponding to this new kind of revision (or, alternatively, the *conditional doxastic logic of probabilistically stable belief*). We ask how to characterise qualitatively the corresponding class of models: we show this can be done via models based on selection functions which emulate probabilistic *strongest-stable-set*-operators. We appeal to some notions from the theory of comparative probability orders and prove a probabilistic representation theorem for these selection function models, thus obtaining a purely 'qualitative' (non-probabilistic) characterisation of strongest-stable-set operators. We briefly point out an interesting connection between our representation theorem and the theory of *simple voting games* [57]. Lastly, we discuss the problem of giving a complete axiomatisation for the logic of probabilistic stability.

1.2 Main results

The main results of this thesis are the following:

- We show that one cannot in general regain commutativity between AGM revision and Bayesian conditioning by raising the stability threshold in Leitgeb's acceptance rule (§2.3.2).
- We prove that AGM revision operators can be generated from Bayesian conditioning and the maximum entropy principle, using Leitgeb's rule (§2.4, Proposition 2.4.1). We give an explicit formula for computing the maximum entropy distribution generating a given plausibility ranking (Proposition 2.4.4).
- We provide sufficient conditions for the joint probabilistic representation of two partial comparative probability orders, one of which extends the other (Proposition 3.3.2).
- We provide a qualitative characterisation of strongest-stable-set operators for finite probability spaces, using selection functions (Propositions 3.3.2 and Theorem 3.3.8). This gives a probabilistic representation theorem for models of the non-monotonic logic of probabilistic stability.

2

Bridging Bayesian conditioning and AGM revision

The Bayesian account of rational belief comprises both a static and a dynamic component. The static component consists in representing the agent's credal states as probability measures; the dynamic one is embodied in the requirement that the revision of a credal state be carried out by Bayesian conditioning. By contrast, the more common notion of belief encountered in traditional epistemology is qualitative[1]: similarly, in applied logic and artificial intelligence, doxastic states are often represented 'qualitatively' as logical propositions, sometimes endowed with some extra structure (e.g., plausibility orderings). In this setting, the most prominent logic-based account of rational belief change is given by AGM belief revision theory [2], in which revisions triggered by new information are modelled by AGM belief revision operators – functions taking one propositional belief state to another.

Thus we have two rather intuitive representations of doxastic states – one probabilistic, and one qualitative – and two corresponding accounts of rational revision of a doxastic state. The question of how Bayesian belief dynamics differ from those of AGM revision is very natural, albeit not so straightforward, as it first requires a well-behaved translation between the two representations. Such a comparison leaves us with a challenge both formal and methodological, motivated by two philosophical questions: firstly, can one reduce the all-or nothing notion of belief to an intrinsically quantitative one? Secondly, what do those two accounts of belief dynamics have in common? Are they competing or compatible? Is there some common notion of dynamic rationality underlying both accounts?

An immediate idea for translating between the two representations of doxastic states consists in defining an *acceptance rule* [25], which maps each probability measure to a

[1]The word *qualitative* suffers from a certain ambiguity. Here, by 'qualitative' we mean little more than not explicitly involving a (real-valued) measure on the relevant space of propositions.

propositional belief state[2]. However, providing a reasonable acceptance rule has been, for a long time, fraught with difficulties. Those can be traced back to Kyburg's notorious Lottery paradox [29], which shows that a most intuitive such rule dubbed the *Lockean rule* – i.e., given a subjective probability measure μ, accept all and only propositions X with $\mu(X) \geq t$, where t is some threshold in (0.5,1] – forces the agent into inconsistent belief states, unless she accepts exactly propositions with probability 1. In turn, this 'probability-1' rule has been criticised for requiring rational acceptance to be *too* cautious [30]: it seems very plausible that a rational agent should believe at least some propositions whose subjective probability falls below 1. Further, the probability-1 proposal renders any formal system for qualitative reasoning under uncertainty essentially trivial, given that it makes it inapplicable to cases of genuine uncertainty. It also makes the comparison with AGM revision quite simple: no (non-trivial) revision of beliefs is possible under the probability-1 rule, since it would require the agent to condition on an event of measure 0.

In recent years, new proposals have appeared to replace the probability-1 rule, due to Leitgeb [30], Lin and Kelly [25], and Delgrande [12]. The shift in perspective comes from the fact that those rules are motivated by the *dynamics* of doxastic states; their behaviour under Bayesian conditioning and/or AGM operators played a role in selecting them as desirable. Interestingly, they succeed in preserving some intuitions behind the Lockean rule, while avoiding the Lottery paradox, and without collapsing into the probability-1 solution. In particular, they allow Bayesian conditioning to generate non-trivial revisions. Thus, they constitute interesting bridges between the probabilistic and qualitative frameworks.

In this chapter, we focus on one such rule due to Leitgeb [30], based on the notion of *stably high probability*. Our central question concerns what Lin and Kelly [25] have called the *tracking* problem (represented schematically in Figure 2.1). Tracking is a simple commutativity condition which gives an obvious criterion of dynamic compatibility between probabilistic and qualitative revision: roughly, a qualitative belief revision method tracks Bayesian conditioning *modulo* an acceptance rule if, starting from any probabilistic credal state, translation through the acceptance rule followed by qualitative revision results in the same belief state as using Bayesian conditioning followed by translation. We study the tracking problem for AGM revision in the light of Leitgeb's rule, and derive two lessons on the (in)compatibilities between Bayesian and AGM-inspired belief dynamics.

After introducing the framework and notation (Section 2.1), we sketch Leitgeb's theory and set the stage for our investigation (Section 2.2). Its starting point is Lin and Kelly's No-Go Theorem [25], which shows that AGM revision operators cannot in general track

[2]In the literature, a distinction is sometimes made between the notions of *acceptance* and *belief* [61]. For the purposes of this thesis, we shall treat the two as synonymous. It is, in fact, an interesting issue whether one could characterise this distinction formally in terms of the behaviour of acceptance rules; but we shall not be concerned with this matter here.

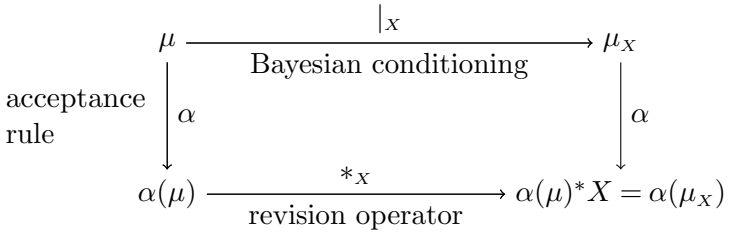


Figure 2.1: The general tracking problem.

Bayesian conditioning modulo an acceptance rule, provided the acceptance rule satisfies some modest requirements. In particular, the No-Go Theorem entails that Leitgeb's rule cannot yield perfect commutativity between Bayesian conditioning and AGM revision. Nonetheless, the principles behind Leitgeb's rule point to an interesting connection between the two operations. We thus consider some ways in which one may circumvent the No-Go Theorem so as to approximate commutativity between AGM and Bayesian conditioning, using Leitgeb's notion of stability (Section 2.3). We show that threshold raising, a very natural idea in this context, fails; as we argue, this failure raises further difficulties for the 'peace project' between Bayesian and AGM-compliant operators. This constitutes our first lesson. However, we also show (in Section 2.4) how an information-theoretic perspective allows to derive a close connection between the Bayesian and AGM accounts: there is a sense in which AGM revision can be seen as deriving from (1) Leitgeb's rule, (2) Bayesian conditioning, and (3) a version of the *maximum entropy principle*. This is our second lesson: it suggests that one could study qualitative revision operators as special cases of Bayesian reasoning which naturally arise in situations of information loss or incomplete probabilistic specification of the agent's doxastic state.

2.1 Preliminaries

We work with probability spaces $(\Omega, \mathfrak{A}, \mu)$, with $\mathfrak{A}$ a set algebra over a sample space Ω, and μ a probability measure on $\mathfrak{A}$. We represent propositions X, Y, Z as elements of the set algebra $\mathfrak{A}$. In the infinite case, we sometimes require $\mathfrak{A}$ to be a σ-algebra, in which case we mention it explicitly. We let $\Delta_{\mathfrak{A}}$ denote the set of all probability distributions on $\mathfrak{A}$. An *acceptance rule* α maps a probability distribution μ in $\Delta_{\mathfrak{A}}$ to the strongest accepted proposition $\alpha(\mu) \in \mathfrak{A}$; we then say an agent accepts (or 'believes') a proposition $X \in \mathfrak{A}$ if and only if $\alpha(\mu) \subseteq X$. By a slight abuse of terminology, the strongest accepted proposition will also be called the 'belief set'.

In this framework, a qualitative (or *propositional*) belief revision operator is a function $* : \mathfrak{A} \times \mathfrak{A} \to \mathfrak{A}$: it is understood that the first variable represents the current strongest accepted

proposition, and the second the new revision input. We will often write a revision by $X \in \mathfrak{A}$ as a (projection) function $(\cdot)^*X : \mathfrak{A} \to \mathfrak{A}$, parametrised by X.

We assume some acquaintance with the basics of AGM theory, as presented in [2, 4, 20]. Since the AGM revision postulates are usually formulated in terms of operators acting on sets of logical formulae, it is worth noting that we can adapt them to our context as follows: for any belief state $K \in \mathfrak{A}$ (where K is the strongest accepted proposition) and propositions X, Y, the revision $*$ is AGM-compliant (or simply AGM) if we have the following:

- $K^*X \subseteq X$
- $K \cap X \subseteq K^*X$ (Inclusion)
- If $K \cap X \neq \emptyset$, then $K^*X \subseteq K \cap X$ (Preservation)
- If $K^*X = \emptyset$ then $K = \emptyset$ or $X = \emptyset$
- $(K^*X) \cap Y \subseteq K^*(X \cap Y)$
- If $(K^*X) \cap Y \neq \emptyset$, then $K^*(X \cap Y) \subseteq (K^*X) \cap Y$

As usual, a probability measure on $\mathfrak{A}$ is a function $\mu : \mathfrak{A} \to [0, 1]$ which is additive (namely, $X \cap Y = \emptyset$ entails $\mu(X \cup Y) = \mu(X) + \mu(Y)$) and satisfies $\mu(\Omega) = 1$. When Ω is infinite, we sometimes require countable additivity. Instead of $\mu(\{\omega\})$ we will write $\mu(\omega)$ for simplicity. For finite powerset algebras, a probability distribution on Ω is a function $\mu : \Omega \to [0, 1]$ such that $\sum_{\omega \in \Omega} \mu(\omega) = 1$. Such a function extends uniquely to a probability measure on $\mathfrak{A}$. Similarly to the above, we will often denote Bayesian conditioning on $X \in \mathfrak{A}$ in parametric form as $|_X$: as usual we have $\mu_X(Y) := \mu(Y|X) = \frac{\mu(Y \cap X)}{\mu(X)}$.

For finite probability spaces with $\Omega = \{\omega_1, .., \omega_n\}$, we will identify probability measures μ with vectors $(\mu(\omega_1), ..., \mu(\omega_n)) \in \mathbb{R}^n$, in which case $\Delta_{\mathfrak{A}}$ is a regular $(n-1)$-simplex Δ^{n-1}. In the last section we will make use of the notion of *Shannon entropy* for probability distributions on finite spaces. When $\mathfrak{A}$ is a finite powerset algebra, the Shannon entropy $\mathcal{H}(\mu)$ of a distribution $\mu \in \Delta_{\mathfrak{A}}$ is defined as $\mathcal{H}(\mu) = \sum_{\omega \in \Omega} -\mu(\omega) \log \mu(\omega)$. When $S \in \mathfrak{A}$ is some finite set, we write $\mathcal{H}(\mu \restriction S) := \sum_{\omega \in S} -\mu(\omega) \log \mu(\omega)$. Sometimes we may wish to distinguish $\mathcal{H}$ as a function of n arguments (e.g. seeing the argument μ as $(\mu(\omega_1), ..., \mu(\omega_n))$), in which case we denote it as $H_n(x_1, ..., x_n) = \sum_{i=1}^n -x_i \log x_i$. A motivation for the notion of entropy (and its use in uncertain reasoning) can be found in [46, 44, 21]; see [46] for basic properties of entropy measures. A useful fact is the following *grouping* property: whenever we have a finite partition of Ω so that $\Omega = \biguplus_{i \leq m} B_i$, we have $\mathcal{H}(\mu) = \mathcal{H}_m(\mu(B_1), ..., \mu(B_m)) + \sum_{i=1}^m \mu(B_i)\mathcal{H}(\frac{1}{\mu(B_i)}\mu \restriction B_i)$, where the notation $k\mu$ denotes the measure defined as $(k\mu)(\omega) = k \cdot \mu(\omega)$.

2.2 Stability principles, AGM revision and the tracking problem

Stability-based acceptance principles were introduced by Leitgeb [30] to provide a bridge between probabilistic and qualitative representations of doxastic states, whilst avoiding the difficulties caused by the Lockean thesis. Those acceptance principles come in two forms: one is based on a fixed threshold parameter determined in advance, and is an acceptance *rule* in the strict sense above. The other, of a less reductionist flavour, allows the threshold to be co-dependent on the probability measure under consideration. Here, we will focus on the first variant, which we dub the τ-rule (we shall briefly discuss the other one in section 2.3). To set the stage for our investigation, we will first recall the essentials of Leitgeb's results, define the τ-rule and explain how it can be seen as deriving from two plausible requirements. We will then clarify why it provides an elegant bridge from probabilistic reasoning to AGM revision. This will lead us to characterise the *tracking problem*. We will then present Lin and Kelly's No-Go Theorem [25] and explain how it applies to Leitgeb's τ-rule.

2.2.1 Stability

As is well-known, Lockean acceptance can be understood as the conjunction of two principles. Suppose the agent's credal state is represented by a probability measure μ on some fixed space. Let then $\lambda_t(\mu)$ denote the strongest accepted proposition under the *Lockean rule* (λ) with threshold $t \in (0.5, 1]$. Consider the following:

$$(\rightarrow) \quad \text{If } \lambda_t(\mu) \subseteq X \text{ then } \mu(X) \geq t$$
$$(\leftarrow) \quad \text{If } \mu(X) \geq t \text{ then } \lambda_t(\mu) \subseteq X$$

The conjunction of those principles constitutes what is known the *Lockean thesis* [18]. Both of them are highly intuitive but, as shown by the Lottery paradox, easily lead to accepting contradictions: for let $\frac{1}{2} \leq t < \frac{n-1}{n}$ for some natural $n > 2$, and suppose μ is a uniform distribution on some finite space $\Omega = \{\omega_1, ..., \omega_n\}$ (e.g., a lottery with n tickets, which the agent believes to be fair, where each $\{\omega_i\}$ represents the proposition 'ticket i will win', and it is assumed only one ticket can win). Then we have, for any $i \leq n$: $\mu(\omega_i) = 1/n$, so $\mu(\Omega \setminus \{\omega_i\}) = \frac{n-1}{n} > t$, and so $\lambda_t(\mu) \subseteq \Omega \setminus \{\omega_i\}$. As this holds for any $i \leq n$, it means that the agent believes of each ticket that it will not win. But it was an elementary assumption (encoded in the sample space Ω) that one ticket will win. Formally, we see that $\lambda_t(\mu) \subseteq \bigcap_{i \leq n} \Omega \setminus \{\omega_i\} = \emptyset$. So $\lambda_t(\mu) = \emptyset$: i.e., the agent believes a contradiction.

To avoid the shortcomings of the Lockean rule, Leitgeb introduces in [30] an interesting new acceptance rule, based on the notion of *stably high probability*. Leitgeb's rule follows the

basic intuition behind the Lockean rule, but it avoids Lottery-like paradoxes: the idea is to preserve the ($\rightarrow$)-direction of the Lockean thesis, while modifying the ($\leftarrow$)-direction so that the agent is never led to accept a contradiction. Instead of believing *all* propositions with probability above the threshold, one restricts acceptance to only some of them. To explain this restriction, we need the following:

Definition 2.2.1 (**Stability**)
Let $(\Omega, \mathfrak{A}, \mu)$ *a probability space and* $t \in (0.5, 1]$. *A set* $X \in \mathfrak{A}$ *is* (μ, t)-stable *if and only if* $\forall Y \in \mathfrak{A}$ *such that* $X \cap Y \neq \emptyset$ *and* $\mu(Y) > 0$, $\mu_Y(X) \geq t$.

Stability captures a notion of *robustness under new information*: a proposition X is (μ, t)-stable if *only* learning a proposition *inconsistent* with X can bring the probability of X below the threshold[3]. In this sense, X has no *defeaters* – propositions consistent with X which lower its probability below t. Leitgeb [33] advocates the requirement that, given a probability measure μ and threshold t, the strongest accepted proposition be (μ, t)-stable. This is the first requirement for acceptance:

> **The Stability Principle (SP):** *given a threshold* t *and* $\mu \in \Delta_{\mathfrak{A}}$, *the strongest accepted proposition must be a* (μ, t)*-stable set in* $\mathfrak{A}$.

(SP) demands that, whenever K is the agent's strongest accepted proposition, no proposition that can be consistently learnt lowers the probability of K below the threshold: only *disbelieved* propositions (inconsistent with K) can affect the probability of K in this way. In other words, K cannot have any defeaters, understood as above.

According to (SP), a necessary condition for accepting some proposition $X \in \mathfrak{A}$ is that X be entailed by a chosen (μ, t)-stable proposition (it is important to note here that (SP) requires only entailment by a stable set, *not* that *every* accepted proposition be (μ, t)-stable). Two remarks are in order: firstly, any measure-1 set is always (μ, t)-stable, as all its potential defeaters have measure 0 and cannot be conditioned upon: so there is always *some* (μ, t)-stable set in $\mathfrak{A}$ which can be chosen as strongest accepted proposition. Secondly, any (μ, t)-stable set K also has probability above the threshold (consider conditioning on the tautological proposition Ω). We can then say that the probability of K is *stably high*. This guarantees that the ($\rightarrow$)-direction of the Lockean thesis is always satisfied: supposing K is (μ, t)-stable, we have that $K \subseteq X$ entails $t \leq \mu(K) \leq \mu(X)$. It also guarantees that no contradiction is ever accepted: and in this setup, logical closure of accepted propositions is immediate, as we take all and only consequences of the strongest accepted proposition. So (SP) suffices to avoid Lottery-like paradoxes, as we shall soon explain more in detail.

[3]For standard probability measures such as here, (μ, t)-stable sets are analogous to *high-probability cores*, a concept which was arrived at independently by Arló-Costa and Pedersen in the context of dyadic probability functions [3].

In the above, the Lockean thesis specified exactly one rule for each value of $t \in (0.5, 1]$, in the sense of a uniquely defined map $\lambda_t : \Delta_{\mathfrak{A}} \to \mathfrak{A}$. (SP), however, is too weak to uniquely specify what the belief set should be[4]: typically, there will be many (μ, t)-stable sets for a given μ and t.

There is, however, another natural constraint that one may impose to guide the choice of a (μ, t)-stable set. Recall that the $(\leftarrow)$-direction of the Lockean thesis requires the agent to accept *all* propositions with measure greater or equal to t. We know, from the Lottery paradox, that this is too strong a requirement. Nonetheless, in order to remain as close as we can to the Lockean thesis, we can weaken it somewhat and opt for the following:

> **Relativised Lockean Principle (RLP):** *accept as many propositions X with $\mu(X) \geq t$ as is possible without violating* (SP).

Leitgeb's rule follows from this strengthening of (SP). In short, it recommends that the strongest accepted proposition be the *logically strongest* (μ, t)-stable proposition, or equivalently, the $\subseteq$-least (μ, t)-stable set. For suppose that $K \subset K'$ and both K, K' are stable: then both have probability above t, and so does any proposition either one entails. It is immediate than choosing K' as strongest accepted proposition is a more severe departure from the $(\leftarrow)$-direction of the Lockean thesis than selecting K: here K entails anything that K' does, but not vice-versa. So, in choosing K', there are more propositions X with $\mu(X) \geq t$ that the agent fails to accept (and K is one of them).

Of course, for this definition to generate a well-defined acceptance rule, we need to make sure that a unique $\subseteq$-minimal stable set always exists. This is guaranteed by Leitgeb's main results from [30], which we now briefly recapitulate. In the reminder of this section we assume, as Leitgeb does, that $\mathfrak{A}$ is a σ-algebra, and that measures μ on it are σ-additive [30]. For our purposes, the most significant is the following:

Proposition 2.2.2 (Leitgeb [30])
Let $\mu \in \Delta_{\mathfrak{A}}$ a σ-additive measure, $t \in (0.5, 1]$.
Then the set $\mathfrak{S}^t_{<1}(\mu) := \{X \in \mathfrak{A} \,|\, \mu(X) < 1$ and X is (μ, t)-stable$\}$ is well-ordered by set inclusion, and has order type at most ω.

Thus, the collection of all (μ, t)-stable sets with probability less than 1 is well-ordered (and at most countable): as a consequence, whenever there is at least one such set for a given μ and t, a $\subseteq$-least one exists. Then (RLP) designates this least (μ, t)-stable set as the strongest accepted proposition.

[4] In fact, this is in tune with the non-reductionistic variant of Leitgeb's theory: it relies on using (SP) to give a 'coherence' restriction for *pairs* (μ, K), which tells us when the probabilistic and propositional representations of a doxastic state are in harmony, while reducing neither of the two representations to the other. We will come back to this idea in Section 2.3.

However, if the collection $\mathfrak{S}^t_{<1}(\mu)$ above is empty – i.e., all (μ, t)-stable sets have measure 1 – the well-order property is not guaranteed. In order to avoid this difficulty here, we follow Leitgeb in restricting our attention to those probability spaces which admit a $\subseteq$-least set among all sets with measure 1. More formally, we work with probability spaces which satisfy the following *Least Certain Set* property (LCS): $\exists X \in \mathfrak{A}$ s.t. $\mu(X) = 1$ and for any Y, if $\mu(Y) = 1$ then $X \subseteq Y$. This trivially ensures the following:

Proposition 2.2.3 (Leitgeb [30])
Let $(\Omega, \mathfrak{A}, \mu)$ a (σ-additive) probability space satisfying (LCS)*, and $t \in (0.5, 1]$. Let S_∞ the least measure-1 set in $\mathfrak{A}$. Then the set $\mathfrak{S}^t(\mu) := \mathfrak{S}^t_{<1}(\mu) \cup \{S_\infty\}$ is well-ordered by set-inclusion.*

We call $\mathfrak{S}^t(\mu)$ the *system of spheres* generated by the measure μ, in reference to Grove's well-known construction [20], to which the link will be made shortly. When the μ and t are implicit, we can simply refer to the (μ, t)-stable sets in $\mathfrak{S}^t(\mu)$ as *spheres*.

Now we are all set to define Leitgeb's acceptance rule: selecting the minimal sphere as the strongest believed proposition clearly satisfies (SP) (as it is (μ, t)-stable) but also (RLP) (as any proposition X with $\mu(X) \geq t$ which is entailed by some other sphere already follows from the least one). We can define:

Definition 2.2.4 (**The τ-rule**)
For any probability measure μ on $\mathfrak{A}$ which satisfies (LCS), and any $t \in (0.5, 1]$, let $\mathfrak{S}^t(\mu)$ the system of spheres generated by μ. Then we define the map $\tau_t : \Delta_{\mathfrak{A}} \to \mathfrak{A}$ as

$$\tau_t(\mu) := \min_{\subseteq} \mathfrak{S}^t(\mu)$$

Since, under (LCS), the system of spheres $\mathfrak{S}^t(\mu)$ is always well-ordered by $\subseteq$, the expression $\tau_t(\mu)$ is clearly well-defined: the strongest accepted proposition under Leitgeb's τ-rule is the least (strongest) stable set for the given μ and t. Whenever the threshold is fixed and implicit in the discussion, we drop the subscript and denote this map as τ.

How restrictive is the (LCS) assumption? Among the measures satisfying (LCS), we can find all probability measures on finite algebras, all countably additive measures on full powersets of countable sets, and all countably additive measures on regular spaces (s.t. $\mu(X) = 0$ iff $X = \emptyset$) [30]. Many probability spaces that are typically of interest in artificial intelligence or used as examples in formal epistemology satisfy (LCS)[5].

Next, how does the τ-rule avoid the Lottery paradox? One can easily see that, for any uniform distribution μ on a finite algebra $\mathfrak{A} = \mathcal{P}(\Omega)$, the only (μ, t)-stable set is Ω: take any

[5] In the last chapter, we shall briefly discuss the case of richer probability spaces which violate (LCS), such as probability models typically occurring in Bayesian statistical inference.

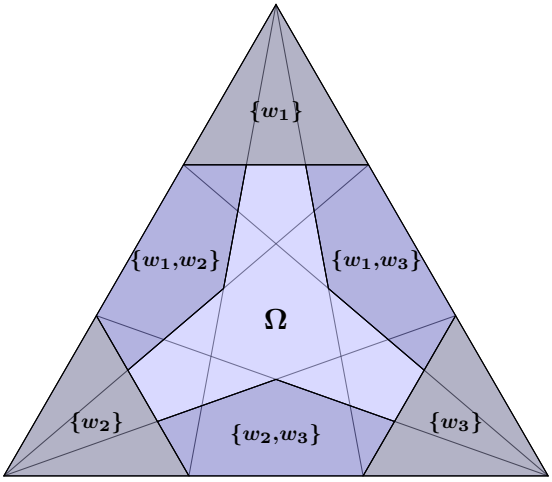


Figure 2.2: Acceptance zones for Leitgeb's τ-rule with $t = 2/3$ and $|\Omega| = 3$.

$X \subset \Omega$. Take $\omega \in X$, $\omega' \notin X$ and let $Y := \{\omega, \omega'\}$. Then $\mu(Y) > 0$ and $\mu_Y(X) = \frac{\mu(X \cap Y)}{\mu(Y)} = \frac{\mu(\omega)}{\mu(\omega)+\mu(\omega')} = 1/2 < t$. So no such X is stable: we have $\tau_t(\mu) = \min_{\subseteq}\{\Omega\} = \Omega$, so as expected, the agent accepts only the tautological proposition. Figure 2.2 gives an example of how the τ-rule partitions a simplex $\Delta_{\mathfrak{A}}$ into *acceptance zones* (an acceptance zone for a proposition $X \in \mathfrak{A}$ is defined as $\tau^{-1}(X) = \{\mu \in \Delta_{\mathfrak{A}} \mid \tau(\mu) = X\}$).

2.2.2 Stability and AGM revision

We have seen that, on the static side, the τ-rule avoids the Lottery paradox, and is motivated by two plausible principles: (SP) and a weakened version of the ($\leftarrow$)-direction of the Lockean thesis. It thus remains very close to the original Lockean proposal, without leading to contradiction. Now, what makes the rule so interesting for our purpose is that, on the dynamic side, it is very closely connected to AGM revision operators. This connection is made by noticing that each system $\mathfrak{S}^t(\mu)$ of (μ, t)-stable sets can be seen as a *system of spheres centered on* $\tau(\mu)$, in the sense of Grove [20]: i.e., (i) it is totally ordered by $\subseteq$ with minimum $\tau(\mu)$, (ii) we have $S_\infty \in \mathfrak{S}^t(\mu)$, and (iii) for every proposition $X \in \mathfrak{A}$, if X intersects some $S \in \mathfrak{S}^t(\mu)$, then there is a $\subseteq$-minimal $S_X \in \mathfrak{S}^t(\mu)$ which intersects X (this follows form the well-ordering of $\mathfrak{S}^t(\mu)$). Grove's well-known representation theorem (see [20, 4]) states that a revision operator $K^*(\cdot) : \mathfrak{A} \to \mathfrak{A}$ acting on a belief set K is AGM if and only if there is a system of spheres $\mathfrak{S}$ centered on K such that, for any X, we have $K^*X = S_X \cap X$, where $S_X = \min_{\subseteq}\{S \in \mathfrak{S} \mid S \cap X \neq \emptyset\}$. A consequence of this is that, given a system of the form $\mathfrak{S}^t(\mu)$ – which we can now call 'sphere system' with full legitimacy – we can define a revision operator on $\tau(\mu)$ as $\tau(\mu)^*X := S_X \cap X$, where $S_X := \min_{\subseteq}\{S \in \mathfrak{S}^t(\mu) \mid S \cap X \neq \emptyset\}$. Grove's theorem then entails that this revision is AGM. We will simply refer to it as the revision

operator *generated by* $\mathfrak{S}^t(\mu)$, or equivalently, generated by τ (for some fixed threshold).

Observe that the sphere system $\mathfrak{S}^t(\mu)$ generates a ranking (total preorder) on Ω. First note that we can index all spheres in $\mathfrak{S}^t(\mu)$ by natural numbers (with the possible exclusion of S_∞) so that $S_i \subseteq S_j$ if and only if $i \leq j$: this is possible as the order type of $\mathfrak{S}^t_{<1}(\mu)$ is at most the ordinal ω (under the assumption of countable additivity). We can then define *ranks* generated by $\mathfrak{S}^t(\mu)$ as follows: $R_0 := S_0$, and $R_{n+1} := S_{n+1} \setminus S_n$. If $\mathfrak{S}^t_{<1}(\mu)$ is infinite, we have $S_\infty = \bigcup_{i\in\mathbb{N}} S_i = \biguplus_{i\in\mathbb{N}} R_i$: then the sphere S_∞ does not define a rank. We say that, for $\omega, \omega' \in \Omega$, $\omega \preceq^\tau_\mu \omega'$ if and only if $\min\{i \mid \omega \in R_i\} \leq \min\{i \mid \omega' \in R_i\}$. We read $\omega \preceq^\tau_\mu \omega'$ as saying that ω is *at least as plausible as* ω'. For finite powerset spaces, we can also treat all states in Ω with measure 0 as being less plausible than all other states. In this way, the τ-rule effectively translates a probability measure into a qualitative plausibility ordering, using the notion of stability as a bridge between the quantitative and propositional representations.

Thus we see how the τ-rule generates a qualitative revision which is AGM. In this sense, it provides a qualitative revision *policy*: to each probability measure, it assigns not only a qualitative belief state, but also an (AGM-complying) qualitative revision operator for revising the latter. We can now turn to dynamics, and ask to what extent the resulting AGM revisions can be said to agree with Bayesian conditioning.

2.2.3 Tracking and the No-Go Theorem

In general, a qualitative revision policy A maps each $\mu \in \Delta_{\mathfrak{A}}$ to a proposition $\alpha(\mu)$ and a revision operator $*$ applicable to that proposition[6], and dependent only on μ. We then say the policy A is *based on* the underlying acceptance rule α. It is *AGM* whenever all revision operators it generates are.

Let us begin by giving a formal definition of tracking.

Definition 2.2.5 (**Tracking**)
A qualitative belief revision policy based on the acceptance rule α tracks *Bayesian conditioning if we have the following commutativity property:*

$$\forall \mu \in \Delta_{\mathfrak{A}},\ \forall X \in \mathfrak{A} \text{ with } \mu(X) > 0,\ \alpha(\mu)^* X = \alpha(\mu_X),$$

where $*$ *is the associated revision operator.*

This notion is illustrated in Figure 2.1. We say that AGM revision can track Bayesian conditioning modulo α if there is some AGM-complying revision policy that is based on α and tracks Bayesian conditioning. This corresponds to a straightforward requirement of agreement between the probabilistic and qualitative doxastic states under translation by α,

[6] We allow operators defined only for a restricted set of revision inputs, such as sets with positive measure under μ.

which must persist under updating by new information. The problem of tracking Bayesian conditioning with qualitative revision operators seems to have been first explicitly addressed by Lin and Kelly in [25]: there, they severely constrain the hope for a harmonious link between Bayesian kinematics and AGM operators, by proving the following:

Theorem 2.2.6 (**The No-Go Theorem**, Lin&Kelly [25])
Let $|\Omega| > 2$, $\mathfrak{A}$ a field of sets over Ω, and let $\alpha : \Delta_{\mathfrak{A}} \to \mathfrak{A}$ be any sensible *acceptance rule. Then no AGM revision policy based on α tracks Bayesian conditioning.*

What is a *sensible* rule? Sensibility amounts to a list of four properties (from Lin and Kelly [25]) which are intended to give minimal conditions for acceptance rules to count as well-behaved. The most important of those conditions is that the acceptance rule never leads to accept the contradictory proposition $\emptyset$; the other three give fairly natural constraints on the geometry of acceptance zones in $\Delta_{\mathfrak{A}}$. We omit the exact definition, as the general case for arbitrary rules is not at the center of our attention here: for our purposes, suffice it to say that Leitgeb's τ-rule can be easily checked to be sensible. Nonetheless, the No-Go Theorem deserves to be stated in its general form, as it indicates that the problem of reconciling AGM revision with Bayesian kinematics goes beyond the difficulties encountered by the τ-rule[7]: simply put, under relatively weak constraints on the acceptance rule, AGM revision cannot track Bayesian conditioning.

Let us take a closer look at those difficulties, to understand how the No-Go Theorem applies in our case. It entails that, once we fix a threshold, a sample space Ω and algebra $\mathfrak{A}$ (with $|\Omega| > 2$: we assume this henceforth), there always will be some $\mu \in \Delta_{\mathfrak{A}}$ and $X \in \mathfrak{A}$ s.t. $\mu(X) \neq 0$ and $\tau(\mu)^* X \neq \tau(\mu_X)$, where $*$ is the τ-generated revision operator. Note the following:

Observation 2.2.7
Let $\mu \in \Delta_{\mathfrak{A}}$, $t \in (0.5, 1]$, and $*$ the AGM revision generated by τ_t. Then $\forall X \in \mathfrak{A}$ with $\mu(X) > 0$, the set $\tau(\mu)^* X$ is (μ_X, t)-stable.

Proof. We show $S_X \cap X$ is (μ_X, t)-stable (where S_X is the least stable set intersecting X, as defined above). Let $Y \in \mathfrak{A}$ such that $S_X \cap X \cap Y \neq \emptyset$ and $\mu_X(Y) > 0$. As $\mu_X(Y) = \frac{\mu(Y \cap X)}{\mu(X)}$, this entails $\mu(X \cap Y) > 0$. We have $S_X \cap (X \cap Y) \neq \emptyset$, so since S_X is (μ, t)-stable, we can write $\mu_{X \cap Y}(S_X) \geq t$. But $\mu_{X \cap Y}(S_X) = \frac{\mu((S_X \cap X) \cap Y)}{\mu(X \cap Y)} = \mu_{X \cap Y}(S_X \cap X)$; we can write $\mu_{X \cap Y}(S_X \cap X) = \mu_X(S_X \cap X \,|\, Y) \geq t$, as required. □

Since $\tau(\mu_X)$ is the minimal (μ_X, t)-stable set, this observation entails that all revision cases yield $\tau(\mu_X) \subseteq \tau(\mu)^* X$: thus, the belief state obtained by Bayesian conditioning followed

[7] In fact, the No-Go Theorem as originally proven in [25] is even more general, where the impossibility result is extended not only to AGM revision, but to any revision operators satisfying Inclusion and Preservation.

by translation through the τ-rule is in general logically stronger than the one obtained by translation followed by the associated AGM revision. As a consequence, whenever tracking fails for the τ-rule (and the associated revision), we must have the *strict* entailment $\tau(\mu_X) \subset \tau(\mu)^* X$. The converse holds trivially: this gives us a clear characterisation of all revision cases for which τ-generated revision fails to commute with Bayesian conditioning (modulo τ). Consider the following:

Example 2.2.8
Let $\Omega := \{\omega_1, ..., \omega_4\}$ and $\mathfrak{A}$ the full power set algebra over Ω. Set $t = 0.7$. Consider the distribution $\mu = (0.5, 0.12, 0.05, 0.33) \in \Delta_{\mathfrak{A}}$. We have $\tau(\mu) = \{\omega_1, \omega_2, \omega_4\}$. Let $X := \{\omega_1, \omega_2, \omega_3\}$. We have $\tau(\mu)^* X = \{\omega_1, \omega_2\} = \tau(\mu) \cap X$ (this is in accordance with the Inclusion *and* Preservation *postulates). But conditioning on X gives $\mu_X \approx (0.746, 0.179, 0.075, 0)$, and we get $\tau(\mu_X) = \{\omega_1\}$: conditioning raises the probability of ω_1 just enough to make it (μ_X, t)-stable. So $\tau(\mu_X) \subset \tau(\mu)^* X$, and tracking fails.*

As this example illustrates, we cannot always guarantee that $\tau(\mu)^* X$ will be the *minimal* (μ_X, t)-stable set. Thus the τ-generated revision operator $*$ cannot track Bayesian conditioning modulo τ. This shows how the No-Go Theorem affects the specific class of revision operators generated by τ (for any threshold t). Note, however, that the theorem extends to *all* AGM revision operators: no matter what AGM operator we begin with, a counterexample to commutativity will exists, effectively preventing the τ-rule itself – and not only the duo τ-rule + τ-generated revision – to establish the desired harmony between AGM revision and Bayesian conditioning.

Our example also illustrates one reason why this happens. Here, consider the qualitative revision $\tau(\mu) \mapsto \tau(\mu_X)$, generated by Bayesian conditioning and the τ-rule. It was bad enough that this "Bayesian" revision did not coincide with the τ-generated revision. This does not, in itself, prevent the possibility of representing the former through an AGM-complying operation. But, to make things worse, it is clear that this cannot be done, as the revision $\tau(\mu) \mapsto \tau(\mu_X)$ simply fails to satisfy the AGM postulates. In particular, we have $\tau(\mu) \cap X \not\subseteq \tau(\mu_X)$; so the Inclusion postulate fails.

One interpretation of this situation is as follows: in the above, consider the events $\{\omega_i\}$ as mutually exclusive and exhaustive hypotheses H_i under consideration. Given μ and $\tau_{0.7}$, the agent accepts the disjunction $H_1 \vee H_2 \vee H_4$. But – so the interpretation goes – a look at the distribution μ reveals that the "main reason" the agent believes $H_1 \vee H_2 \vee H_4$ is because $H_1 \vee H_4$ appears very plausible to her: H_2 is retained as possible only because $H_1 \vee H_4$ is not quite (μ, t)-stable without H_2 – the disjunction $H_1 \vee H_4$ by itself fails to be stable (by a small margin) since $\mu(\{\omega_1, \omega_4\} \mid \{\omega_2, \omega_3, \omega_4\}) = 0.66 < 0.7 = t$. But once the possibility H_4 is eliminated through conditioning, H_1 is so much more plausible that $H_2 \vee H_3$ than no doubt remains: the agent believes H_1, while the more conservative AGM revision recommends to

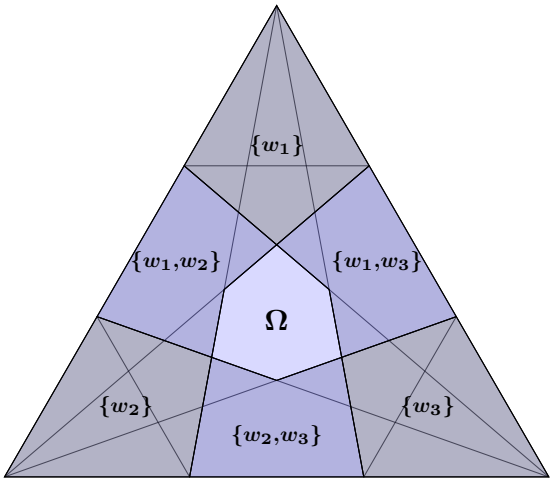
be more cautious and accept only $H_1 \vee H_2$.

This phenomenon is typical of counterexamples to tracking. The idea is that probabilistic representations provide enough information to identify some propositions φ, ψ which were the "main" reason for accepting a disjunction $\varphi \vee \psi \vee \chi$, in situations where (1) $\varphi \vee \psi$ was not plausible (stable) enough to be accepted, and (2) φ is the most plausible proposition and χ the least, as given by the probability measure. Then, after conditioning on $\neg\psi$, comparing the probabilities of φ and χ leads to φ being believed in view of how much more probable it is than the alternatives. In the qualitative AGM setting, however, once a disjunction is accepted, all of its disjuncts stand on equal footing; no distinctions based on relative plausibility can be made. One is then not warranted in arbitrarily excluding one of them, unless it is contradicted by new information. This analysis gives more substance to the intuition that Bayesian conditioning is too fine-grained to agree with the more coarse-grained, and conservative, AGM revision.

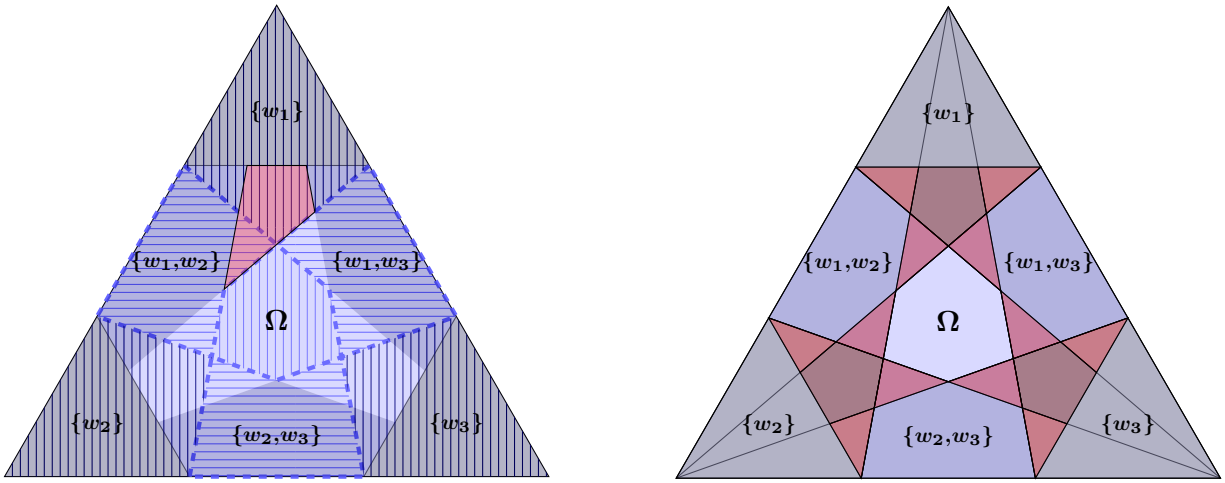
Of course, a Bayesian convinced by the Lockean and stability principles behind the τ-rule could very well use this line of reasoning as an argument against Inclusion. Arguing in this manner against Inclusion should be seen as complementing an argument of Lin and Kelly against Preservation given in [25] and [26]. This argument is based on the following kind of examples: working in a space with at least three mutually exclusive propositions φ, ψ, χ, one accepts $\varphi \vee \psi$ 'mostly because' of φ. The idea is that one's probability distribution roughly measures the subjective strength of the *rationale* for accepting compound propositions. The proposition φ has a much higher probability than either ψ or χ – so much so that the revision by the very 'surprising' proposition $\neg\varphi$ leads one to reconsider the previous acceptance entirely and accept $\psi \vee \chi$, including χ again as a possibility. It is then argued that this revision is reasonable: it encodes the intuition that revising by $\neg\varphi$ amounts to invalidating the main rationale for believing $\varphi \vee \psi$ in the first place, and thus for having excluded χ.

In fact, it turns out that, as soon as the Bayesian reasoner tries to track AGM revision, she is led to a trade-off between Inclusion and Preservation (when provided with a sensible acceptance principle). This, in broad terms, is the reasoning found in Lin and Kelly's proof of the No-Go Theorem. In [25], Lin and Kelly not only give a Bayesian-style argument against the Preservation postulate[8], but they also offer an alternative acceptance rule – the so called Shoham-driven rule – which leads the Bayesian reasoner to fail Preservation while satisfying Inclusion [25, 26]. We leave a more careful comparison between the Shoham and τ rules for another time; for an idea of how the Shoham-driven rule behaves, see Figure 3. By contrast, under Leitgeb's τ-rule, Bayesian reasoning will not necessarily satisfy Inclusion, while always complying with Preservation. To see this, suppose $\tau(\mu) \cap X \neq \emptyset$: by Observation 2.2.7, we get $\tau(\mu_X) \subseteq \tau(\mu)^* X$. As the τ-generated revision $*$ satisfies Inclusion and Preservation, we

[8] In [26], they also use it to argue against the closely related *Rational Monotonicity* rule for nonmonotonic logics.



(a) Acceptance zones for Lin's and Kelly's Shoham-driven rule with $t = 2/3$.



(b) Disagreement zones between the τ and Shoham-driven rules with $t = 2/3$. On the left-hand side, we have coloured in red the zone where Inclusion fails for the τ-rule under conditioning by $\Omega \setminus \{\omega_2\}$. On the right-hand side, all disagreement zones are coloured.

Figure 2.3: Comparing Leitgeb's τ and Lin and Kelly's Shoham-driven rules.

have $\tau^* X = \tau(\mu) \cap X$, and we are done.

Here is a more interesting example where tracking fails:

Example 2.2.9

The agent is given an urn. She knows that it is either of the type **A** *– containing 30% black marbles and 70% white marbles, or* **B** *– containing 70% black and 30% white marbles. She believes option* **A** *and option* **B** *are equally plausible. Suppose she draws (with replacement) 10 marbles form the urn. How many black marbles would she have to draw to be convinced the urn is of type* **A***?*

Our sample space contains the 22 propositions in $\{\mathbf{A} \cap D_i, \mathbf{B} \cap D_i \mid 0 \leq i \leq 10\}$*, where* **A, B** *indicate which urn was given, while* D_i *means that* i *black marbles have been drawn in our 10-draw trial. Here we assume a 50-50 prior distribution* μ *for urns* **A, B** *and we use a binomial distribution to compute conditional probabilities. We obtain the following joint distribution:*

	D_0	D_1	D_2	D_3	D_4	D_5	D_6	D_7	D_8	D_9	D_{10}
A	*.0141*	*.0605*	*.1167*	*.1334*	*.1000*	*.0514*	*.0183*	*.0045*	*.0007*	*.00...*	*.00...*
B	*.00...*	*.00...*	*.0007*	*.0045*	*.0183*	*.0514*	*.1000*	*.1334*	*.1167*	*.0605*	*.0141*

For a threshold of 1/2, the system of spheres $\mathfrak{S}^{1/2}$ *generates the following ranking (we have two ranks 0 and 1):*

	D_0	D_1	D_2	D_3	D_4	D_5	D_6	D_7	D_8	D_9	D_{10}
A	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*1*	*1*	*1*	*1*
B	*1*	*1*	*1*	*1*	*0*	*0*	*0*	*0*	*0*	*0*	*0*

This means that, using the τ*-generated revision policy, drawing 0, 1, 2, or 3 black marbles convinces the agent that she was given urn* **A***: e.g., learning* D_2 *leaves only the propositions* $\mathbf{A} \cap D_2$ *(with rank 0) and* $\mathbf{B} \cap D_2$ *(with rank 1), so the agent believes* $\mathbf{A} \cap D_2$*. However, drawing 4 marbles yields disagreement between conditioning and revision: on the AGM side, the agent is undecided between the two urns, as she remains with the propositions* $\mathbf{A} \cap D_4$ *and* $\mathbf{B} \cap D_4$ *of equal rank. On the Bayesian side, however, we get* $\mu(\mathbf{A} \mid D_4) \approx 0.845$*, while* $\mu(\mathbf{B} \mid D_4) \approx 0.155$*. The proposition* $\mathbf{A} \cap D_4$ *is then the least* μ_{D_4}*-stable set, and so gets the least rank: the agent believes the urn is of type* **A***.*

This down-to-earth example is a good illustration of how the cautiousness of AGM revision may prevent agreement with Bayesian conditioning.

While the No-Go Theorem precludes the possibility of perfect commutativity with AGM modulo the τ-rule, it is natural to ask if there is any way one could still make the case for a certain harmony between AGM and Bayesian reasoning. Despite its failures, the τ rule imposes itself as a natural tool for this purpose. Our discussion so far reveals two reasons why this is so: first, it is a well-motivated acceptance principle, which avoids Lottery-like paradoxes and remains close to Lockean intuitions. Secondly, it provides a very elegant (though imperfect) bridge to qualitative revision dynamics through its connection with sphere-based models. We will consider the following question: is there any sense in which one can *approximate* commutativity through τ-generated revisions? The idea is to weaken somewhat the tracking requirement, and try to show that the τ-rule allows AGM operators to 'approximate' Bayesian reasoning in this weaker sense. Now that we know *how* the No-Go Theorem affects the τ-rule and its associated revisions, we have a clearer idea of where the problem lies; let us consider some ways to address it.

2.3 Approximating Agreement

We now consider two ways in which one may attempt to show the (approximate) compatibility of Bayesian and AGM reasoning by weakening the commutativity criterion. Both involve modifying the notion of acceptance rule in a way that renders τ-generated revisions and Bayesian updates dynamically compatible under the resulting (weakened) criterion. The first one consists in adopting a highly non-reductionist notion of acceptance, which yields a criterion for dynamic compatibility so weak that even the most trivial qualitative revisions pass the test. The second attempt relies on allowing an acceptance principle the output of which depends not only on the current probability measure, but also on the history of past updates. While this proposal appears more promising, it cannot be carried out without violating Lockean intuitions. A closer analysis shows that this violation reveals a deeper conflict between Lockean intuitions and AGM revision, which arises in the light of the stability principles.

2.3.1 Non-reductionism: lowering the bar

As we mentioned above, Leitgeb's stability theory of belief admits a variant of a strongly non-reductionistic flavour. The theory amounts to keeping the Stability Principle (SP) and abandoning the Relativised Lockean Principle (RLP) which gives rise to the τ-rule. Thus, given μ and t, selecting *any* (μ, t)-stable set constitutes a reasonable translation from the probabilistic credal state: in other words, the pair (μ, K) – with $K \in \mathfrak{A}$ as strongest accepted proposition – is considered coherent, as long as K is (μ, t)-stable. In this way, qualitative belief states are not seen as emerging deterministically from subjective probabilities. Accordingly, if one accepts only (SP) as a criterion of acceptance, this also yields a weaker criterion of dynamic coherence between conditioning and a qualitative revision $*$: one requires only that, given a pair (μ, K) that is coherent in the sense above, the pair (μ_X, K^*X) be coherent for any X with $\mu(X) > 0$. In this sense, we achieve coherence between AGM and Bayesian conditioning by taking $*$ to be the τ-generated revision: as we have seen, K^*X is then (μ_X, t)-stable and coherence is preserved at every revision step.

This is, however, a very weak criterion of dynamic coherence. It is simply due to the fact that (SP) is a rather weak constraint. For example, the following revision process would pass the coherence test: begin by accepting only the least probability-1 proposition S_∞ and, for any Bayesian update by X, accept as strongest $S_\infty \cap X = X$, iterating this process as needed by taking intersections. This makes acceptance trivial: at every step, the strongest accepted proposition simply represents the weakest proposition that has not been yet ruled out: the plausibility ordering given by the corresponding system of spheres forgets all other information provided by the probability distribution. One can hardly say

that such a procedure 'approximates' tracking: it simply consists in ignoring virtually any detail provided by the probabilistic description of the agent's doxastic state. Nothing in this weakened coherence criterion allows to rule out such cases as degenerate.

Can we do any better? Here is another idea: use the τ-rule as long as tracking works, and 'force' commutativity when it does not. This amounts to setting another acceptance map τ' as follows: starting at some prior measure μ and some fixed threshold, we set $\tau'(\mu) = \tau(\mu)$. For the next revision step by $X \in \mathfrak{A}$, simply *define* the map τ' as picking $\tau(\mu)^*X$ as strongest accepted proposition. Thus, if the revision $\tau(\mu) \mapsto \tau(\mu_X)$ coincides with the τ-generated revision, we have $\tau'(\mu_X) = \tau(\mu_X)$; but whenever tracking fails and we have $\tau(\mu_X) \subset \tau(\mu)^*X$, the τ' map selects $\tau(\mu)^*X$. So we force τ' to coincide with τ in all revision cases except the problematic ones, by setting it to respect the AGM revision generated by the sphere system of (μ, t)-stable sets. It is important to note that τ' is *not* an acceptance rule in the technical sense: it is clearly not a function from $\Delta_{\mathfrak{A}}$ to $\mathfrak{A}$, as the value of $\tau'(\mu)$ for some measures μ – indeed, the problematic ones, of the form μ_X, for which $\tau(\mu_X) \subset \tau(\mu)^*X$ – will depend on the particular revision history that has taken place so far[9].

Leitgeb himself has suggested a solution similar to the above [31]. An acceptance principle behaving like the one just described yields a much closer connection between AGM revision and Bayesian dynamics. It respects (SP) at all times, and imitates the behaviour of the τ-rule, except when the τ-rule fails to respect commutativity. In this sense, it can be said to approximate tracking.

However, on one level, this connection remains unsatisfactory. To begin with, the acceptance principle above lacks an independent motivation: it is simply designed to satisfy AGM, and is defined directly in terms of the AGM operator generated by the τ-rule. That one can use this principle to approximate tracking without violating (SP) is, of course, a step towards reconciliating AGM with Bayesian models of reasoning – particularly so since (SP) can be motivated in a manner entirely independent of the tracking problem for AGM revision. But an obvious difficulty is that this solution is in direct conflict with (RLP). Consider any case where tracking fails: we have some rule τ, threshold $t \in (0.5, 1]$, and $X \in \mathfrak{A}$ such that $\tau(\mu_X) \subset \tau(\mu)^*X$. After conditioning on X, the probability of $\tau(\mu_X)$ is above the threshold t: thus, Lockean intuitions – as expressed in (RLP) – dictate that $\tau(\mu_X)$ should be believed. So picking $\tau(\mu)^*X$ as strongest accepted proposition goes against (RLP).

This violation of (RLP) calls for a justification. After all, an important motivation for

[9]For instance, to identify the strongest accepted proposition $\tau'(\mu_X)$ given a distribution μ_X, the τ' map needs to be provided at least enough information to identify μ and X, in order to know what $\tau(\mu)^*X$ is. We may well have $\rho = \mu_X = \nu_X$ for different measures μ and ν such that commutativity holds when updating μ by X (in which case we set $\tau'(\rho) = \tau(\rho)$), but fails when updating ν by X (in which case we want $\tau'(\rho) = \tau(\nu)^*X$). The τ' map must be able to differentiate between those cases. One way to ensure this is to provide it with (1) the initial prior distribution μ and (2) the proposition $\bigcap_{i \leq n} X_i$, where $\langle X_i \mid i \leq n \rangle$ is the sequence of all revision inputs provided so far.

the stability-based account – and one of its strengths – was to remain as close as possible to the Lockean thesis whilst avoiding paradoxes of the Lottery kind. Indeed, the τ-rule, which selects the *minimal* (μ, t)-stable set, owes its specificity (among all rules satisfying (SP)) to the fact that it satisfies the greatest number of instances of the ($\leftarrow$)-direction of the Lockean thesis. But we see that selecting $\tau(\mu)^*X$ as strongest accepted proposition, in cases where commutativity fails, cannot be done for free. Not only does a Bayesian reasoner have no positive incentives to obey the AGM requirements in those cases, she can do so only *at the cost of violating the Lockean principle* (RLP). Thus, she is confronted with a true trade-off between AGM and Lockean intuitions.

One ideal of compatibility could be described as follows: under an independently established acceptance principle[10], feasible AGM revisions can be probabilistically represented by a Bayesian update, and any Bayesian update translates into an AGM revision. This corresponds to a two-way agreement between the Bayesian and AGM-compliant reasoner, mediated by an acceptance principle that is acceptable to both. But under this proposal, only the latter side can be content: the Bayesian, on the other hand, may naturally still demand a rationale for privileging AGM dynamics over the ($\leftarrow$)-direction of the Lockean thesis.

So if, in Leitgeb's words, a "peace project" [30, p.70] between AGM and Bayesian models of reasoning is to be carried out successfully, it would be reasonable to require a more *principled* reason to 'force' commutativity: justifying the commutative choice by a simple desire to preserve AGM revision gets us only half of the way there.

2.3.2 Raising the bar

As the τ-rule depends on the ("contextually determined" [30, p.14]) threshold parameter t, a very simple idea suggests itself for justifying the commutative map $\tau'(\mu_X) := \tau(\mu)^*X$. Whenever commutativity fails for the rule τ_t and revision of μ by X, one could try and show that $\tau'(\mu_X) = \tau_q(\mu_X)$, where q is some new threshold, so that the revision $\tau_t(\mu) \mapsto \tau_q(\mu_X)$ is AGM. In other words, the idea is to show that using the τ_t-generated revision on $\tau_t(\mu)$ can be represented as Bayesian conditioning followed by τ_q. Then, we could claim that selecting the AGM-compliant proposition as strongest does not really violate (RLP), but instead is a perfectly legitimate instance of τ-based acceptance, only with a different threshold. It is immediate that we should set $q > t$ for this to work: it follows directly form the definition of stability that (μ, t)-stability entails (μ, q)-stability for any $q \leq t$. So picking any $q \leq t$ would not do, as it would preserve the stability of $\tau_t(\mu_X)$.

Recall Example 2.2.8: we have a distribution $\mu = (0.5, 0.12, 0.05, 0.33)$, and a threshold

[10] Of course, as the No-Go theorem indicates, this acceptance principle cannot be a (sensible) acceptance *rule* in the technical sense.

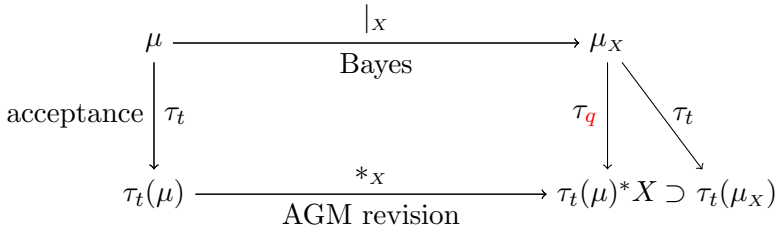


Figure 2.4: Rationalising AGM revision by raising the threshold: here we pick a new threshold $q > t$.

$t = 0.7$, hence $\tau(\mu) = \{\omega_1, \omega_2, \omega_4\}$. Let $Y := \{\omega_1, \omega_2\}$. Then $\mu_Y \approx (0.806, 0.194, 0, 0)$ and we get $\tau(\mu_Y) = \{\omega_1\}$. Then tracking fails, since $\tau(\mu)^*Y = \{\omega_1, \omega_2\} = Y$. But now we can lift the threshold to a new value $q = 0.807$ (any value above 0.806 will do). Then clearly $\tau(\mu_Y)$ is not (μ_Y, q)-stable (here, $\mu_Y(\tau(\mu_Y))$ is not even above the threshold). However $\tau(\mu)^*Y$ *is* – indeed it has probability 1 – and is also the least (μ_Y, q)-stable set. We have thus 'corrected' the threshold as required: we approximate commutativity as we have $\tau_t(\mu)^*Y = \tau_q(\mu_Y)$, with $*$ the AGM revision operator generated by τ_t.

However, this threshold-raising method does not work in general. To see why, first consider how the method works when it does: we start with a measure $\mu \in \Delta_{\mathfrak{A}}$ and $X \in \mathfrak{A}$ such that $\tau_t(\mu_X) \subset \tau_t(\mu)^*X$, with both of those sets (μ_X, t)-stable. For the method to work, we need at least to raise the threshold enough so that $\tau_t(\mu_X)$ is no longer stable, but $\tau_t(\mu)^*X$ is. So it works if and only if $\exists q \in (0.5, 1]$, $\tau_t(\mu)^*X$ is the least (μ_X, q)-stable set, while $\tau_t(\mu_X)$ is not stable for q. Now, to determine whether such a q exists, we can first check, for each of these sets, what is the *maximal* value of q which would make them (μ_X, q)-stable. This is a set's *degree of stability*[11]:

Definition 2.3.1 (**Degree of stability**)
The degree of stability of $X \in \mathfrak{A}$ with respect to a measure $\mu \in \Delta_{\mathfrak{A}}$, denoted $\mathcal{S}(\mu, X)$ (or simply $\mathcal{S}(X)$ when μ is implicit) is defined as:

$$\mathcal{S}(\mu, X) := \sup\{q \in [0, 1] \mid X \text{ is } (\mu, q)\text{-stable}\}.$$

Note that this is defined only when $\mu(X) > 0$. Before giving a counterexample to threshold-raising, it is useful to have the following:

Proposition 2.3.2
Let $\mu \in \Delta_{\mathfrak{A}}$ and $X \in \mathfrak{A}$, with $\mu(X) > 0$. Then, $\mathcal{S}(\mu, X) = \inf\{\mu_Y(X) \mid \mu(Y) > 0,\ X \cap Y \neq \emptyset\}$

[11] Here we extend the notion of stability to cases with $t < 1/2$.

and, as a consequence, we have that

$$X \text{ is } (\mu, t)\text{-stable if and only if } \mathcal{S}(\mu, X) \geq t.$$

Proof. Let us write $T_X := \{q \in [0,1] \mid X \text{ is } (\mu, q)\text{-stable}\}$ (for the set of good Thresholds of X) and $D_X := \{\mu_Y(X) \mid \mu(Y) > 0,\ X \cap Y \neq \emptyset\}$, for the set of probabilities given by potential Defeaters of X. We show $\sup T_X = \inf D_X$. First suppose that $\sup T_X > \inf D_X$. Then $\exists q \in T_X$ with $\inf D_X < q$. But this means $\exists Y$ with $\mu(Y) > 0$, $X \cap Y \neq \emptyset$ s.t. $\mu_Y(X) < q$. But then X is not (μ, q)-stable, so $q \notin T_X$; contradiction. So $\sup T_X \leq \inf D_X$ after all. For the other direction, it is enough to note the following. Let Y such that $\mu(Y) > 0$, $X \cap Y \neq \emptyset$. Then $\mu_Y(X) \in D_X$, so $\mu_Y(X) \geq \inf D_X$. As this holds for any such Y, this means that X is $(\mu, \inf D_X)$-stable. This means $\inf D_X \in T_X$, an so $\sup T_X \geq \inf D_X$. We can conclude $\mathcal{S}(\mu, X) = \sup T_X = \inf D_X$.

Now we show X is (μ, t)-stable if and only if $\inf D_X \geq t$. First assume $\inf D_X \geq t$. Then no potential defeater Y of X brings the probability of X strictly below t: more precisely, $\forall Y$ s.t. $\mu(Y) > 0$, and $X \cap Y \neq \emptyset$, we have $\mu_Y(X) \geq t$. This says exactly that X is (μ, t)-stable. For the other direction, suppose X is (μ, t)-stable, and assume for reductio that $\inf D_X < t$. Then there is a Y with $\mu(Y) > 0$, and $X \cap Y \neq \emptyset$, and such that $\mu_Y(X) < t$. So X has a defeater Y, hence is not (μ, t)-stable, which is contrary to our assumption. So $\inf D_X \geq t$ holds and we are done. □

Both of those simple characterisations of $\mathcal{S}(\mu, X)$ can be useful, depending on the context. The result above is rather intuitive. It says that, by computing $\mathcal{S}(\mu, X)$, we in fact check how low the probability of X can be brought down by conditioning on potential 'defeaters' in the sense of Section 3. If $\inf\{\mu_Y(X) \mid \mu(Y) > 0,\ X \cap Y \neq \emptyset\} \geq t$, this means that no such potential defeater Y can lower the probability of X strictly below t. So X does not actually have any defeaters with respect to t: it is (μ, t)-stable. Note that the proof of the proposition above also allows us to write $\mathcal{S}(\mu, X) = \max T_X$; the set $T_X = \{q \in [0,1] \mid X \text{ is } (\mu, q)\text{-stable}\}$ always has a maximum. This is because, as shown in the proof, $\sup T_X = \inf D_X \in T_X$, so $\mathcal{S}(\mu, X) = \sup T_X = \max T_X$.

One last remark before proceeding. Here is a simple method for computing $\mathcal{S}(\mu, X)$, particularly useful for finite probability spaces. Suppose you know that A is a (nonempty) minimal-measure subset of X and that $\mu(X) < 1$. Then it is straightforward that we have $\mathcal{S}(\mu, X) = \mu(X \mid X^c \cup A) = \frac{\mu(A)}{\mu(A)+\mu(X^c)}$: i.e., then $X^c \cup A$ is the 'strongest' defeater: we have $(X^c \cup A) \cap X = A \neq \emptyset$ and $\mu(X^c \cup A) > \mu(X^c) > 0$, and no other such set can bring the probability of X any lower. In fact, we then have $\mu_{X^c \cup A}(X) = \min D_X$.

Let us go back to the threshold-raising method. By the above, we now know that if B is not (μ, t)-stable but A is, clearly we have $\mathcal{S}(\mu, B) < t \leq \mathcal{S}(\mu, A)$. If follows that, whenever

tracking fails and we have $\tau(\mu_X) \subset \tau(\mu)^*X$, one can raise the threshold to "correct" the revision process *only if* $\mathcal{S}(\mu_X, \tau(\mu_X)) < \mathcal{S}(\mu_X, \tau(\mu)^*X)$. But not all such cases are correctible in this way. We illustrate this with a simple counterexample:

Example 2.3.3

Consider the same setting as in Example 2.2.8. We have $t = 0.7$*, a distribution* $\mu = (0.5, 0.12, 0.05, 0.33)$*, and* $X = \{\omega_1, \omega_2, \omega_3\}$*. Here* $\tau_t(\mu) = \{\omega_1, \omega_2, \omega_4\}$*. Then* $\mu_X \approx (0.746, 0.179, 0.075, 0)$*, and tracking fails since* $\tau_t(\mu_X) = \{\omega_1\}$ *and* $\tau_t(\mu)^*X = \{\omega_1, \omega_2\}$*. We want to find* $q \in (0.5, 1]$ *such that* $\tau_t(\mu)^*X = \tau_q(\mu_X)$*: for such a* q*, we must have* $\tau_t(\mu)^*X$ *as (the least)* (μ_X, q)*-stable set, while* $\tau_t(\mu_X)$ *is not stable. But we have the following degrees of stability with respect to* μ_X*:*

$$\mathcal{S}(\tau_t(\mu_X)) = \frac{\mu_X(\omega_1)}{\mu_X(\omega_1) + \mu_X(\Omega \setminus \{\omega_1\})} = \frac{0.746}{1} = 0.746$$

And

$$\mathcal{S}(\tau_t(\mu)^*X) = \frac{\mu_X(\omega_2)}{\mu_X(\omega_2) + \mu_X(\Omega \setminus \{\omega_1, \omega_2\})} = \frac{0.179}{0.179 + 0.075} \approx 0.705$$

This means that the maximal q *such that* $\tau_t(\mu)^*X$ *is* (μ_X, q)*-stable is* $q = 0.705$*. But for* $\tau_t(\mu_X)$*, this maximal threshold it at* $q = 0.746$*. So any threshold* q *for which* $\tau_t(\mu)^*X$ *is* (μ_X, q)*-stable also makes* $\tau_t(\mu_X)$ *stable. So we cannot raise the threshold so as to approximate commutativity in our sense.*

So, one cannot in general justify the choice of the AGM-compliant belief state by raising the threshold, as there exist such 'non-correctible' cases. In fact, it can be shown such cases exist whenever the algebra $\mathfrak{A}$ is generated by more than 3 elements[12], and, for full powerset algebras at least, they form an open neighbourhood in $\Delta_{\mathfrak{A}}$ of significant size (their geometric Lebesgue measure is > 0). The failure of the threshold-raising method simply follows form the fact that degrees of stability do not necessarily respect the $\subseteq$-order on $\mathfrak{A}$.

The fact that such counterexamples exist also seems to undermine another attempt to legitimise the method of forcing commutativity. This argument, which Leitgeb sketches in [31], relies on a specific instance of threshold-raising and applies to cases where $\mu_X(\tau(\mu_X)) < 1$. Here the idea is again to move to a new threshold q, by setting $q = \mu_X(\tau(\mu)^*X)$ when commutativity fails. Then – the argument goes – selecting $\tau(\mu)^*X$ as strongest accepted proposition is in fact in accordance with Lockean principles, since one gets the *full* Lockean thesis w.r.t the measure μ_X and threshold q. In other words, we get $\forall Y \in \mathfrak{A}$, $\tau(\mu)^*X \subseteq Y$ if and only if $\mu_X(Y) \geq q$, as shown in [30, p. 34]. On this basis, one may be tempted to claim that threshold-raising is a reasonable solution after all[13]. But our remarks above should cast

[12] For $\mathfrak{A} = \mathcal{P}(\Omega)$ with $|\Omega| = 3$, each revision which fails tracking is in fact correctible: this is because, in each such case we have $\tau(\mu_X)$ is a singleton, in which case it suffices to raise the threshold above $\mu_X(\tau(\mu_X))$.

[13] This argument cannot apply to cases where $\mu_X(\tau(\mu_X)) = 1$, for we cannot get the Lockean thesis with $\tau(\mu)^*X$ as strongest accepted proposition. This is immediate, since we then have $q = 1$ and so $\mu_X(\tau(\mu_X)) \geq q$ but $\tau(\mu_X) \subset \tau(\mu)^*X$.

some doubt on this claim, as they yield the following:

Observation 2.3.4
Let $\mu \in \Delta_{\mathfrak{A}}$, $X \in \mathfrak{A}$, and let τ have a fixed threshold. Suppose the following hold:

- $\tau(\mu_X) \subset \tau(\mu)^*X$ *(tracking fails)*
- $\mathcal{S}(\mu_X, \tau(\mu_X)) \geq \mathcal{S}(\mu_X, \tau(\mu)^*X)$ *(the case is non-correctible)*
- $\mu_X(\tau(\mu_X)) < 1$

Set $q := \mu_X(\tau(\mu)^*X)$. Then $\tau(\mu)^*X$ is not (μ_X, q)-stable.

Proof. For some fixed $t \in (0.5, 1]$, take any 'non-correctible' case as above. Suppose $\tau(\mu)^*X$ is (μ_X, q)-stable. Then, since we know $\mathcal{S}(\mu_X, \tau(\mu_X)) \geq \mathcal{S}(\mu_X, \tau(\mu)^*X)$, the set $\tau(\mu_X)$ is also (μ_X, q)-stable. This entails $\mu_X(\tau(\mu_X)) \geq q$. But we also know $\tau(\mu_X) \subset \tau(\mu)^*X$. This entails[14] $\mu_X(\tau(\mu_X)) < \mu_X(\tau(\mu)^*X)$, which means simply $\mu_X(\tau(\mu_X)) < q$. This is a contradiction.

Alternatively, we can derive the contradiction using Leitgeb's result about the full Lockean thesis for q. We know $\mu_X(\tau(\mu_X)) \geq q$. Now Leitgeb's result entails that for any Y we have $\tau(\mu)^*X \subseteq Y$ if and only if $\mu_X(Y) \geq q$. In particular, this means that $\mu_X(\tau(\mu_X)) \geq q$ entails $\tau(\mu)^*X \subseteq \tau(\mu_X)$. But this last inclusion cannot hold, since $\tau(\mu_X) \subset \tau(\mu)^*X$. □

What does this say about Leitgeb's variant of the threshold-raising method? In a word, when we apply the method to non-correctible cases by raising the threshold to q as above, we necessarily violate the Stability Principle (SP) with respect to q. For example:

Example 2.3.5
Consider again Example 2.2.8. There, set $\tau(\mu)^*X$ as strongest accepted proposition, and let $q := \mu_X(\tau(\mu)^*X) = 0.925$. It is easy to see that we obtain the full Lockean thesis for q, but $\tau(\mu)^*X$ itself is not (μ_X, q)-stable. For take $Y = \{\omega_2, \omega_3\}$. Then $\mu_X(Y) = 0.254 > 0$ and $(\tau(\mu)^*X) \cap Y = \{\omega_2\} \neq \emptyset$; but by conditioning on Y we get $\mu_X(\tau(\mu)^*X \mid Y) = \frac{\mu_X(\omega_2)}{\mu_X(Y)} = \frac{0.179}{0.254} = 0.705 < q$. Thus Y is a defeater for $\tau(\mu)^*X$.

Observation 2.3.4 entails that this happens in any case to which Leitgeb's argument applies. Thus, when we force commutativity by selecting $\tau(\mu)^*X$ as the strongest belief, we have another dilemma: either we keep our initial threshold t – in which case Lockean principles are violated – or we switch to this new threshold q, thus getting the full Lockean thesis with respect to q, but failing the stability requirement (SP) (with respect to q). This, of course, follows from the very existence of non-correctible cases, for which no threshold satisfies both stability and Lockean principles, as required by (RLP).

[14] To see why this is strict: suppose we have equality here. This means $(\tau(\mu)^*X) \setminus \tau(\mu_X)$ has measure 0 under μ_X. But we know, by assumption, that $\mu_X(\tau(\mu)^*X) < 1$. So $\tau(\mu)^*X$ is a (μ_X, t)-stable set with probability < 1, and with a measure 0 subset. This cannot happen in general: for suppose S is (μ, t)-stable, $\mu(S) < 1$ and $A \subset S$ with $\mu(A) = 0$. Then $S^c \cup A$ is a defeater for S, contradicting the fact that S is stable.

One way to avoid those negative conclusions would be to see the τ' acceptance map as generated by *two* thresholds: in fact, in the cases above, we can see that $\tau(\mu)^*X$ is stable with respect to the old threshold t, while it yields the Lockean thesis with respect to the new threshold q. So we could always require stability with respect to the parameter t, and the Lockean thesis with respect to q. To understand this proposal, it is useful to sketch Leitgeb's more recent presentation of the non-reductionistic version of the stability rule, which explicitly follows this path [35].

The non-reductionistic stability rule can be derived from what Leitgeb calls the *Humean Thesis*. Let the belief set of an agent be represented by a set $\mathbb{K} \subseteq \mathfrak{A}$ of accepted propositions. The Humean thesis is a constraint on accepted beliefs according to which a rational agent's belief set $\mathbb{K}$ must satisfy

$$A \in \mathbb{K} \text{ if and only if } \forall X \in Poss(\mathbb{K}) \text{ such that } \mu(X) > 0,\ \mu(A \mid X) > t \tag{2.1}$$

where $Poss(\mathbb{K}) := \{X \in \mathfrak{A} \mid X^c \notin \mathbb{K}\}$ is the collection of *doxastically possible* hypotheses. A probabilistic reasoner who follows the Humean thesis believes all and only those hypotheses that are stable under conditioning on any proposition that is not *dis*believed. Leitgeb shows that, on any finite probability space, the sets $\mathbb{K}$ satisfying (2.1) correspond exactly to sets of the form $\{X \in \mathfrak{A} \mid S \subseteq X\}$ where S is a (μ, t)-stable set. So the Humean reasoner selects a stable set and closes under deduction. Further, we have the following

Theorem 2.3.6 (Leitgeb [35])
Representation theorem for the stability rule on finite spaces.
Let $(\Omega, \mathfrak{A}, \mu)$ a finite probability space and $\mathbb{K} \subseteq \mathfrak{A}$. Fix $t \in (0.5, 1)$. The following are equivalent.

(i) $\mathbb{K}$ *satisfies the Humean thesis* (2.1).

(ii) $\mathbb{K} = \{X \in \mathfrak{A} \mid S \subseteq X\}$ *where* S *is a* (μ, t)*-stable set.*

(iii) $\mathbb{K} = \{X \in \mathfrak{A} \mid \mu(X) \geq q\}$ *where* $q = \mu(S)$ *for a* (μ, t)*-stable set* S.

In (ii) and (iii), if $\mu(S) = 1$, then S is the least set of measure 1.

The Humean Thesis (and the stability rule) derive much of their strength from this representation theorem. The theorem shows that, on finite spaces, the non-deterministic stability rule coincides with the Humean thesis; and that both coincide with a version of the Lockean thesis in which the choice of threshold is constrained by the measure of stable sets[15]. In other words, it shows that three distinct ways of thinking about acceptance yield one and the same acceptance rule:

[15] Leitgeb [35, Thm 7, p.121] shows something slightly stronger: namely, that (i) and (ii) above are equivalent to the statement that $\mathbb{K}$ is consistent, closed under consequence and conjunctions, and is generated by some unique strongest proposition S such that it satisfies the right-to-left direction of the Lockean Thesis with threshold equal to $\mu(S)$.

- acceptance as (Humean) stability under doxastically possible propositions,
- acceptance as entailment by a chosen probabilistically stable hypothesis,
- acceptance as selecting a consistent, logically closed belief given by the Lockean thesis (with the choice of the Lockean threshold set to the measure of some stable set)[16].

This account of acceptance evidently does not yield a functional acceptance rule in the reductionistic sense: a probability measure does not uniquely determine a set of accepted hypotheses. Given a probability measure, there exist as many possible belief sets for the agent as there are probabilistically stable hypotheses: one generates a belief set by selecting a stable hypothesis and closing the belief set under logical consequence. This deterministic stability rule τ corresponds to selecting the logically strongest belief set satisfying either one of (i), (ii) and (iii); this canonical choice can additionally be justified as the collection of propositions that admit a *probabilistically stable justification*. That is, the τ-rule amounts to accepting all and only those propositions X for which there is some *stable* hypothesis entailing X.

Now, one can argue, as Leitgeb does [35], that the choice of a generating stable set corresponds to the choice of a Lockean threshold, which tracks the agent's degree of 'cautiousness'. In this way, this extra degree of freedom in applying the acceptance rule is explained by the existence of a Lockean threshold for the agent. Crucially, as our observations above show, the Lockean threshold must be independent of the threshold for stability, if we want revisions to preserve both the Lockean thesis and the principle that the strongest accepted proposition be stable. Bayesian agents can follow AGM revisions by selecting an appropriate Lockean threshold q: the fact that the resulting strongest accepted proposition is not stable with respect to the chosen Lockean threshold does not matter, since it is only required to be stable with respect the distinct *stability* threshold t.

In the context of dynamics of credal states, such a separation of the stability threshold from the Lockean threshold may strike one as being somewhat inelegant and, at worse, entirely ad-hoc. Setting aside the cost of introducing an additional parameter, it is far from clear what the relationship between these two distinct parameters is, and what exact role they play in the acceptance process. For one thing, in order to justify AGM revisions one would need to explain why the Bayesian agent should feel compelled to select a Lockean threshold that generates the AGM-compliant revision. It is not clear how one would justify the introduction of this extra parameter in a manner independent enough of AGM revision to be acceptable for the Bayesian. One would then also lose the advantage of the stability theory as initially introduced, in which the use of a single threshold for stability and Lockean principles constitutes a rather simple and plausible refinement of the original Lockean thesis.

[16]It can be shown that the probabilities of stable sets are Lockean thresholds that yield a conjunctive, deductively closed, and consistent belief set.

Finally, looking at degrees of stability may give rise to one more worry, albeit a minor one, when 'forcing' the relevant AGM-compliant belief state. In non-correctible cases, inducing AGM revision in this way means not only departing from the Lockean principle, but also choosing a belief state $\tau(\mu)^*X$ that is *less stable* than the one given by Bayesian Conditioning and the τ rule. Presumably, if the stability requirement has any plausibility at the outset – at least enough plausibility to convince the Bayesian reasoner to opt for stable sets as qualitative representations of belief – then perhaps it would be even more natural to advocate the choice of $\tau(\mu_X)$ as belief state, given that it is not only logically stronger, but also *stabler* than $\tau(\mu)^*X$.

We can sum up what we have seen so far:

- The threshold-raising method does not always work, and for many probability spaces there exist counterexamples to tracking that cannot be captured by the application of the stability rule with a higher threshold;
- Non-correctible cases highlight an incompatibility between the Lockean and Stability principles and the τ-generated AGM revision, even if one allows thresholds to vary;
- Leitgeb's variant of the threshold-raising method, which relies on the non-reductionistic variant of the stability rule, requires the parameters (the Lockean and stability thresholds) to be distinct; only a very limited choice of Lockean thresholds leads the Bayesian reasoner to comply with AGM revision.

Thus, we do not appear to be any closer to a satisfactory justification for the method of 'forcing' commutativity. The threshold-raising argument brings more difficulties than it solves. And if the same threshold is kept throughout, making Bayesian conditioning and AGM revision compatible entails giving up on the Lockean principles underlying (RLP). This is a significant price to pay. In the absence of an independent rationale for this, it is difficult to see what incentive the Bayesian reasoner would have to recognise those AGM-compliant representations as legitimate. But if the hope is to allow Bayesian and AGM reasoning to 'make peace', one must convince *both* parties – hence the Bayesian also – that peace is desirable in the first place. Unless some justification is found for prioritising AGM-based intuitions over Lockean ones, the AGM reasoner and the Bayesian will have to agree to disagree.

2.4 Recovering revision operators via Maximal Entropy

In this section, we show that the conclusions drawn so far – largely negative for Leitgeb's peace project – can be mitigated by considering another approach to the comparison of probabilistic and qualitative belief dynamics. We employ elementary information theory to show

that an interesting connection emerges between AGM revision and Bayesian conditioning, in cases where the probabilistic information about the agent's doxastic state is incomplete. In such situations, as we will prove, Bayesian conditioning generates AGM revisions purely from the stability rule and a version of the maximum entropy principle. This suggests that some qualitative revision operators can be studied as special cases of probabilistic reasoning under the constraint of information loss.

Bayes with no measure. It would appear thus far that the dynamic behaviour of AGM operators cannot be easily reconciled with that of Bayesian conditioning, despite the significant step in this direction made by Leitgeb's stability principles. The cautiousness of AGM revision does not mix well with the fine-grained nature of subjective probabilities. But what if the probabilistic representation of the agent's doxastic state is not fully specified? Suppose, for example, that only a qualitative representation is available to the agent (say, a single proposition X, or a sphere system $\mathfrak{S}$), but the agent is strictly committed to Bayesian conditioning as an update method. What is then needed is a principled way to obtain a probabilistic representation μ of her doxastic state (suitably 'matching' X or $\mathfrak{S}$) in order to perform conditioning on μ and translate the result back into a qualitative representation, by means of a selected acceptance rule α.

As we will now see, it turns out that for $\mathfrak{A}$ a finite powerset algebra, and when the acceptance rule in question is Leitgeb's τ-rule, this can be done in such a way that the resulting revision operation is always AGM. In fact, we show how AGM revision emerges in a natural way from two purely probabilistic principles together with the τ-rule, in situations where the information about the agent's probabilistic credal state is incomplete. This will yield a way of approximating commutativity in cases where translating a probability measure into a plausibility ordering results in some loss of information.

The first of the probabilistic principles in question is simply Bayesian conditioning. The other principle is the following version of the *Maximum Entropy Principle*.

Maximum Entropy Principle (MEP):
If all that is known to the agent is that a probability distribution lies within some zone $\mathcal{N} \subseteq \Delta_{\mathfrak{A}}$, the agent selects a distribution with maximal entropy among those in $\mathcal{N}$, if such exist.

Versions of the maximum entropy principle abound in the information theory and artificial intelligence literature (see [21]) and numerous justifications have been offered for its application, both in the context of statistical inference in particular sciences (e.g. in statistical mechanics, as famously advocated by Jaynes [24]) as well as in the general

mathematical study of uncertain reasoning (for instance, see [44] or [21]). The gist of the argument is that the Shannon entropy $\mathcal{H}(\mu)$ of a distribution μ measures the uncertainty about which state in Ω is the true one (in our case, the one that should be believed); so the higher the entropy of the distribution, the less biased it is towards any particular element of the sample space. In this way, if the only available constraint on the distribution is that it lies within some zone $\mathcal{N} \subseteq \Delta_{\mathfrak{A}}$, selecting a maximal entropy distribution in $\mathcal{N}$ amounts to choosing a *least biased* distribution, given the available information. Such a distribution is naturally thought to best represent the available information.

In the problem at hand, how does this principle come into play? Starting from a qualitative representation $\mathcal{Q}$ of the agent's doxastic state ($\mathcal{Q}$ could be a proposition in $\mathfrak{A}$, or a sphere system $\mathfrak{S}$), MEP guides the selection of a probabilistic representation of $\mathcal{Q}$. First of all, it would be of course reasonable to expect the chosen acceptance rule τ to help specify *what* counts as an acceptable probabilistic representation of $\mathcal{Q}$. For instance, for a probability measure μ to represent a belief set K we must at least require that μ generates the belief set K via our acceptance rule. So, more generally, for a measure μ to count as a representation we must at least require $\tau(\mu) = \mathcal{Q}$ if $\mathcal{Q} \in \mathfrak{A}$, and $\mathfrak{S}^t(\mu) = \mathcal{Q}$ if $\mathcal{Q}$ is a system of spheres (recall there that $\mathfrak{S}^t(\mu)$ is the system of spheres given by the (μ, t)-stable sets, as defined at the start of this chapter). Secondly, there will in general be many suitable representations of $\mathcal{Q}$, forcing a subset $\mathcal{N} \subseteq \Delta_{\mathfrak{A}}$: the MEP comes in handy in the selection of a minimally biased representation.

From here on we assume $\mathfrak{A} = \mathcal{P}(\Omega)$ with Ω finite (of size > 2). We can now give mathematical substance to our claim and prove the following:

Proposition 2.4.1
Let $X \neq \emptyset$ a proposition in $\mathfrak{A}$, and $\tau = \tau_t$ for some $t \in [0.5, 1)$. Then there is a unique maximal entropy distribution $\mu \in \Delta_{\mathfrak{A}}$ such that $\tau(\mu) = X$. Moreover, for any $Y \in \mathfrak{A}$, we have $\tau(\mu_Y) = X^*Y$, where $*$ is the AGM revision operator generated by $\mathfrak{S}^t(\mu)$.

First let us introduce the following useful notions: given a measure $\mu \in \Delta_{\mathfrak{A}}$, we say that μ is *rank-uniform* if it is uniform on all ranks in the sphere system $\mathfrak{S}^t(\mu)$ associated with μ (see subsection 2.2.2 for the definition of ranks). We say that two measures ρ and μ are *rank-equivalent* if for they generate the same ranks, and for any rank R in their associated systems of spheres we have $\rho(R) = \mu(R)$. A measure ρ *entropy-dominates* μ if $\mathcal{H}(\rho) \geq \mathcal{H}(\mu)$. Lastly, we say that μ *has m ranks* if $\mathfrak{S}^t(\mu)$ does (equivalently, when $|\mathfrak{S}^t(\mu)| = m$). We will make use of the following observation:

Observation 2.4.2
Any $\mu \in \Delta_{\mathfrak{A}}$ is entropy-dominated by some rank-equivalent, rank-uniform probability measure $\rho \in \Delta_{\mathfrak{A}}$.

Proof. We adapt a standard argument for the entropy-maximality of uniform distributions. Let $\mu \in \Delta_{\mathfrak{A}}$. Write all μ-ranks as $R_1, ..., R_m$ (generated by the τ rule for some fixed threshold t). We have

$$\begin{aligned}\mathcal{H}(\mu) = \sum_{\omega\in\Omega} -\mu(\omega)\log\mu(\omega) &= \sum_{i=1}^{m}\left(\sum_{\omega\in R_i} -\mu(\omega)\log\mu(\omega)\right)\\ &= \sum_{i=1}^{m}\mathcal{H}(\mu\restriction R_i) \qquad (2.2)\end{aligned}$$

Now assume μ is not rank-uniform. Take the rank-equivalent, rank-uniform measure ρ, defined as $\rho(\omega) = \frac{\mu(R_i)}{|R_i|}$ for $\omega \in R_i$. This generates the exact same system of spheres as μ: (i) clearly ρ cannot create any new stable sets, as all states in the same rank have equal measure, and (ii) ρ preserves the stability of any (μ, t)-stable set. Suppose S is (μ, t)-stable: this entails that $\frac{\mu(\omega_m)}{\mu(\omega_m)+\mu(S^c)} \geq t$, where ω_m is the state of minimal measure in S. But the definition of ρ entails $\rho(\omega_m) \geq \mu(\omega_m)$ and $\rho(S^c) = \mu(S^c)$, so we have $\frac{\rho(\omega_m)}{\rho(\omega_m)+\rho(S^c)} \geq t$: S is also (ρ, t)-stable. So μ and ρ generate the same ranks. We also have, for each rank R_i, $\rho(R_i) = \mu(R_i)$.

For all ranks R' on which μ is uniform, we have $\mu\restriction R' = \rho\restriction R'$, and so $\mathcal{H}(\mu\restriction R') = \mathcal{H}(\rho\restriction R')$. Now let R be a rank on which μ is *not* uniform. We show $\mathcal{H}(\rho\restriction R) > \mathcal{H}(\mu\restriction R)$: then, equation (2.2) gives us the desired result. We have $\mathcal{H}(\rho\restriction R) = \sum_{\omega\in R}\frac{\rho(R)}{|R|}\log\left(\frac{|R|}{\rho(R)}\right) = \rho(R)\log\left(\frac{|R|}{\rho(R)}\right)$. Now consider the function $\theta : x \mapsto x\log x$ defined on $\mathbb{R}^+$. This is strictly convex on $(0, \infty)$. Using Jensen's inequality[17], we can write

$$\theta\left(\sum_{\omega\in R}\frac{\mu(\omega)}{|R|}\right) < \sum_{\omega\in R}\frac{1}{|R|}\theta(\mu(\omega)) = \frac{1}{|R|}\sum_{\omega\in R}\theta(\mu(\omega))$$

As $\sum_{\omega\in R}\mu(\omega) = \mu(R) = \rho(R)$, we write

$$\begin{aligned}|R|\cdot\theta\Big(\frac{\rho(R)}{|R|}\Big) &< \sum_{\omega\in R}\theta(\mu(\omega))\\ \text{so } -|R|\cdot\theta\Big(\frac{\rho(R)}{|R|}\Big) &> -\sum_{\omega\in R}\theta(\mu(\omega)) = -\sum_{\omega\in R}\mu(\omega)\log\mu(\omega)\end{aligned}$$

[17] Jensen's inequality states that whenever we have a strictly convex map $f : \mathbb{R}\to\mathbb{R}$, then $\forall x_i \in Dom(f)$ (with $i \leq n$) and $\alpha_i \in [0, 1]$ with $\sum_i \alpha_i = 1$, we have that

$$f\left(\sum_{i=1}^{n}\alpha_i x_i\right) < \sum_{i=1}^{n}\alpha_i f(x_i)$$

unless all x_i's are equal and all $\alpha_i > 0$, in which case we have equality. Here we can use the strict inequality, as the x_i correspond to the $\mu(\omega)$'s, which, by assumption, are not all equal.

Now the left-hand side equals $-|R| \cdot (\frac{-\rho(R)}{|R|}) \log(\frac{|R|}{\rho(R)}) = \rho(R) \log\left(\frac{|R|}{\rho(R)}\right)$, while the right-hand side equals $\mathcal{H}(\mu \restriction R)$. So we have

$$\rho(R) \log\left(\frac{|R|}{\rho(R)}\right) = \mathcal{H}(\rho \restriction R) > \mathcal{H}(\mu \restriction R)$$

So ρ entropy-dominates μ, as required. □

Note that we have shown something slightly stronger: that any non-rank uniform μ is *strictly* dominated by a rank-equivalent, rank-uniform measure. Another key observation is:

Observation 2.4.3
If μ is rank-uniform, then for any $X \in \mathfrak{A}$, the revision $\tau(\mu) \mapsto \tau(\mu_X)$ is the AGM revision generated by $\mathfrak{S}^t(\mu)$.

Proof. Suppose μ is rank-uniform. We show that $\mathfrak{S}^t(\mu_X) = \mathfrak{S}^t(\mu) \restriction X$. By an argument similar to the one in Observation 2.2.7, we get $\mathfrak{S}^t(\mu) \restriction X \subseteq \mathfrak{S}^t(\mu_X)$. We only need the other inclusion. Suppose, for reductio, that $\mathfrak{S}^t(\mu_X)$ strictly refines $\mathfrak{S}^t(\mu) \restriction X$. Then there exist at least two states $\omega_1, \omega_2 \in X$ which both belong to the same μ-rank R, but are separated into different μ_X-ranks: say $\omega_1 \in R_1$ and $\omega_2 \in R_2$. As μ is rank-uniform, we must have $\mu(\omega_1) = \mu(\omega_2)$ and hence $\mu_X(\omega_1) = \frac{\mu(\omega_1)}{\mu(X)}$ clearly equals $\mu_X(\omega_2)$. But states with equal measure cannot have different ranks. Now $\mathfrak{S}^t(\mu_X) = \mathfrak{S}^t(\mu) \restriction X$ entails $\tau(\mu_X) = \min \mathfrak{S}^t(\mu_X) = \tau(\mu)^* X$: first recall that $\tau(\mu)^* X = S_X \cap X$, where S_X is the smallest sphere in $\mathfrak{S}^t(\mu)$ intersecting X. Suppose towards a contradiction that $\exists Y \in \mathfrak{S}^t(\mu) \restriction X$ such that $Y \subset S_X \cap X$: then we have $Y = S' \cap X$ for some $S' \in \mathfrak{S}^t$ with $S' \subset S_X$, contradicting the minimality of S_X. □

We now prove Proposition 2.4.1.

Proof of Proposition 2.4.1. Let $\emptyset \neq X \in \mathfrak{A}$. We can assume $X \neq \Omega$ (if $X = \Omega$ we simply take μ uniform on Ω and we are done). Let $\tau^{-1}(X) := \{\mu \in \Delta_{\mathfrak{A}} \mid X = \tau(\mu)\}$. We want to find a maximum entropy measure μ in $\tau^{-1}(X)$.

We define a distribution μ such that (1) μ is uniform on X, (2) μ is uniform on $\Omega \setminus X$, and (3) X is (μ, t)-stable. The requirements (1) and (2) directly imply

$$\mu(\omega) = \begin{cases} \frac{\mu(X)}{|X|} & \text{if } \omega \in X \\ \frac{1-\mu(X)}{|\Omega|-|X|} & \text{if } \omega \notin X \end{cases} \qquad (\star)$$

(3) in turn requires that $\mu(\omega_X) \geq t(1-\mu(X))/1-t$, where ω_X is a state in X with minimal μ-measure; as we want μ uniform on X, this condition is equivalent to

$$\frac{\mu(X)}{|X|} \geq \frac{t}{1-t}(1-\mu(X))$$

Here we set μ to satisfy $\frac{\mu(X)}{|X|} = \frac{t}{1-t}(1-\mu(X))$. In other words, among all distributions satisfying (1), (2) and (3), we pick the one that assigns a minimal measure to X. Rewrite the last equation as

$$\mu(X) = \frac{t \cdot |X|}{1 + t(|X| - 1)}$$

Together with ($\star$) this gives us the following definition for μ:

$$\mu(\omega) = \begin{cases} \frac{t}{1+t(|X|-1)} & \text{if } \omega \in X \\ \frac{1}{(|\Omega|-|X|)} \cdot \frac{1-t}{(1+t(|X|-1))} & \text{if } \omega \notin X \end{cases}$$

It should be clear that $\mathfrak{S}^t(\mu)$ contains exactly two spheres (hence two ranks, X and $\Omega \setminus X$) and μ is uniform on both ranks. We claim μ is the unique distribution with maximum entropy in $\tau^{-1}(X)$, i.e., we show $\forall \rho \neq \mu$ in $\tau^{-1}(X)$, $\mathcal{H}(\rho) < \mathcal{H}(\mu)$. Let $\rho \in \tau^{-1}(X)$, $\rho \neq \mu$. By Observation 2.4.2 above, we know that each measure is entropy-dominated by a rank-equivalent measure which is uniform on all ranks. This means we can assume, without loss of generality, that ρ is rank-uniform and show that μ *strictly* entropy-dominates ρ.

(i) First, consider the case where $|\mathfrak{S}^t(\rho)| > 2$, hence ρ has > 2 ranks. Take the measure ρ' s.t. $\rho' \restriction X = \rho \restriction X$, and ρ' uniform on $\Omega \setminus X$. Then ρ' is a measure with 2 ranks, with minimal rank X: the fact that $\rho' \restriction X = \rho \restriction X$ clearly guarantees that X is the least (ρ', t)-stable set, and since ρ' is uniform on $\Omega \setminus X$, the only other (ρ', t)-stable set is Ω itself (as states with equal measure cannot belong to distinct spheres). Now we can write

$$\begin{aligned} \mathcal{H}(\rho) = &\sum_{\omega \in X} -\rho(\omega) \log(\rho(\omega)) + \sum_{\omega \in \Omega \setminus X} -\rho(\omega) \log(\rho(\omega)) \\ = &\mathcal{H}(\rho \restriction X) + \mathcal{H}(\rho \restriction X^c) \end{aligned}$$

And similarly for $\mathcal{H}(\rho')$. This means we have $\mathcal{H}(\rho') - \mathcal{H}(\rho) = \mathcal{H}(\rho' \restriction X^c) - \mathcal{H}(\rho \restriction X^c)$. Note that, since ρ has strictly more than 2 ranks, it cannot be uniform on X^c, while ρ' is. Moreover we have $\rho'(X) = \rho(X)$ so $\rho'(X^c) = \rho(X^c)$. We have shown in the proof of Observation 2.4.2 that this entails $\mathcal{H}(\rho' \restriction X^c) > \mathcal{H}(\rho \restriction X^c)$, which gives us $\mathcal{H}(\rho') > \mathcal{H}(\rho)$. So ρ' is a (rank-uniform) measure with 2 ranks which strictly entropy-dominates ρ. Thus it only remains to show that μ entropy-dominates all other rank-uniform measures with two ranks (and with minimal rank X). We take care of this in the next case.

(ii) We consider the case of (rank-uniform) measures in $\tau^{-1}(X)$ with 2 ranks. We denote the set of all such measures $\mathcal{U}$. Note that any $\rho \in \mathcal{U}$ is entirely specified by the value of $\rho(X)$(also note that, by choice of μ, we cannot have $\rho(X) < \mu(X)$ since this contradicts the (ρ, t)-stability of X). We show that for $\rho \in \mathcal{U}$, $\mathcal{H}(\rho)$ is maximised exactly when $\rho(X) = \mu(X)$; i.e., for $\rho = \mu$.

Notice that for any measure $\rho \in \Delta_{\mathfrak{A}}$, we can write

$$\mathcal{H}(\rho) = \mathcal{H}_2(\rho(X), \rho(X^c)) + \rho(X)\mathcal{H}\left(\frac{1}{\rho(X)}\rho \restriction X\right) + (1-\rho(X))\mathcal{H}\left(\frac{1}{(1-\rho(X))}\rho \restriction X^c\right) \quad (\circledast)$$

This equality can be derived directly from the definition of $\mathcal{H}$, or alternatively it can be seen as an immediate consequence of the "grouping" property of entropy[18]. Now if ρ is uniform on X we have $\forall \omega \in X$, $\rho(\omega) = \rho(X)/|X|$, so $\frac{1}{\rho(X)}\rho(\omega) = \frac{1}{|X|}$. This also means

$$\begin{aligned}\mathcal{H}\left(\frac{1}{\rho(X)}\rho \restriction X\right) &= \sum_{\omega \in X} \frac{1}{\rho(X)}\rho(\omega) \log\left(\frac{1}{\frac{1}{\rho(X)}\rho(\omega)}\right) \\ &= \sum_{\omega \in X} \frac{1}{|X|} \log(|X|) \\ &= \log(|X|)\end{aligned}$$

By the same reasoning, when ρ is uniform on X^c we get $\mathcal{H}\left(\frac{1}{(1-\rho(X))}\rho \restriction X^c\right) = \log(|\Omega \setminus X|)$. Now for $\rho \in \mathcal{U}$ we can rewrite $(\circledast)$ as

$$\begin{aligned}\mathcal{H}(\rho) &= \mathcal{H}_2(\rho(X), \rho(\Omega \setminus X)) + \rho(X)\log(|X|) + (1-\rho(X))\log(|\Omega \setminus X|) \\ &= \mathcal{H}_2(\rho(X), 1-\rho(X)) + \rho(X)\log\left(\frac{|X|}{|\Omega \setminus X|}\right) + \log(|\Omega \setminus X|)\end{aligned}$$

It remains to show that this expression is maximised on $\mathcal{U}$ exactly when $\rho(X) = \mu(X)$. For convenience we write $\rho(X) = a$ and so $\mathcal{H}(\rho) = \mathcal{H}_2(a, 1-a) + a\log\left(\frac{|X|}{|\Omega \setminus X|}\right) + \log(|\Omega \setminus X|)$. We have

$$\begin{aligned}\mathcal{H}_2(a, 1-a) &= a\log\left(\frac{1}{a}\right) + (1-a)\log\left(\frac{1}{1-a}\right) \\ &= a\log\left(\frac{1-a}{a}\right) - \log(1-a)\end{aligned}$$

Differentiating $\mathcal{H}(\rho)$ w.r.t a, we obtain

$$\frac{d\mathcal{H}}{da} = \log\left(\frac{1-a}{a}\right) + \log\left(\frac{|X|}{|\Omega \setminus X|}\right)$$

By choice of μ, for any $\rho \in \mathcal{U}$, we must have $a = \rho(X) \geq \mu(X) = \frac{t \cdot |X|}{1+t(|X|-1)}$: so we are only concerned with the value of the derivative $d\mathcal{H}/da$ for $a \in [\mu(X), 1]$. First note that $t > 1/2$ entails[19] $\frac{t \cdot |X|}{1+t(|X|-1)} > \frac{1/2 \cdot |X|}{1+1/2(|X|-1)} = \frac{|X|}{|X|+1}$. So $a \in [\mu(X), 1]$ entails $a > \frac{|X|}{|X|+1}$. This yields

[18] See the Preliminaries at the beginning of the chapter.

[19] The expression for $\mu(X)$ increases in t.

$\frac{1-a}{a} < \frac{1}{|X|}$, so $\log\left(\frac{1}{a}-1\right) < \log\left(\frac{1}{|X|}\right)$, and we get

$$\log\left(\frac{1-a}{a}\right) + \log\left(\frac{|X|}{|\Omega \setminus X|}\right) < \log\left(\frac{1}{|X|}\right) + \log\left(\frac{|X|}{|\Omega \setminus X|}\right) = \log\left(\frac{1}{|\Omega \setminus X|}\right)$$

But it is clear that $\log\left(\frac{1}{|\Omega\setminus X|}\right) \leq 0$ (recall $X \neq \Omega$). So we have $\log\left(\frac{1-a}{a}\right) + \log\left(\frac{|X|}{|\Omega\setminus X|}\right) < 0$. This means that $d\mathcal{H}/da$ is strictly negative for $a \in [\mu(X), 1]$. Thus $\mathcal{H}(\rho)$ strictly decreases on $\mathcal{U}$ as $\rho(X)$ increases, with $\mathcal{H}(\rho)$ seen as a function of $\rho(X)$. Since entropy is strictly decreasing for $\rho(X) \in [\mu(X), 1]$, it is maximised exactly when $\rho(X) = \mu(X)$, or equivalently for $\rho = \mu$. So μ, as we have defined it, is the required maximum entropy distribution, and it is unique.

Finally, μ is rank-uniform, so Observation 4 guarantees that any feasible revision $\tau(\mu) \mapsto \tau(\mu_Y)$ is the AGM revision $X \mapsto X^*Y$ generated by $\mathfrak{S}^t(\mu)$. □

The upshot is that one can uniquely recover AGM revision in the case where the agent's doxastic state is represented only by her strongest accepted proposition. Suppose the doxastic state of the agent is represented by a proposition $X \in \mathfrak{A}$, with no information about her numerical credences – say, the full probability distribution is too costly to remember, or information has been lost after some quantitative-qualitative translation (or perhaps, there never was any probabilistic representation). The agent selects, in accordance with MEP, the distribution which best represents her state of knowledge – namely, the maximum entropy distribution lying in the acceptance zone for X under τ. Then we have commutativity: using Bayesian conditioning and an application of the τ-rule, we get the same result as applying to X the AGM revision generated by τ and the corresponding maximum entropy distribution.

Further, it is not necessary to 'forget' this much information for this maximum-entropy method to work: a similar result still holds if more information is available about the doxastic state. Suppose the agent begins with a full plausibility ranking (total preorder) of the various hypotheses $\omega \in \Omega$ – or, equivalently, suppose we begin with a system of spheres. This corresponds to a case where we do not retain complete information about the agent's probability measure, but we have preserved more information than merely the agent's raw propositional belief set: we store an intermediate description of her doxastic state consisting of a qualitative plausibility ranking. Note, in particular, that the system of spheres can be seen as encoding *conditional* beliefs, or belief-revision strategies. Then there is still a unique maximum entropy distribution generating this system of spheres: from this distribution, Bayesian conditioning always generates the AGM revision corresponding to the ranking[20]:

[20]The same holds if one remembers not only the plausibility ordering, but also the measures of each rank.

Proposition 2.4.4

Let $\mathfrak{S}$ be a system of spheres on $\mathfrak{A}$, with $\mathfrak{A}$ finite. Then there is a unique maximum entropy μ on $\mathfrak{A}$ such that $\mathfrak{S}^t(\mu) = \mathfrak{S}$. Moreover, for any $X \in \mathfrak{A}$ with $\mu(X) > 0$, we have $\mathfrak{S}^t(\mu_X) = \mathfrak{S}^t(\mu) \restriction X$, and so the associated revision $\tau(\mu) \mapsto \tau(\mu_X)$ is AGM.

Proof. We maximise $\mathcal{H}$ over $[\mathfrak{S}] := \{\mu \in \Delta \mid \mathfrak{S}^t(\mu) = \mathfrak{S}\}$. By Observation 2.4.2, each non-rank uniform distribution is strictly dominated in entropy by a rank-uniform one: so if a maximum entropy distribution exists on this domain, it is rank-uniform. Thus it is enough to maximise entropy on the set $\mathcal{U}$ of all *rank-uniform* measures μ such that $\mathfrak{S}^t(\mu) = \mathfrak{S}$. We have

$$\arg\max_{\mu \in \mathcal{U}} \mathcal{H}(\mu) = \arg\max_{\mu \in [\mathfrak{S}]} \mathcal{H}(\mu)$$

For any $\mu \in \mathcal{U}$, we express entropy as a function of the measure of the n ranks $R_1, ..., R_n$ generated by $\mathfrak{S}$ (we assume $n > 1$: if there is only one rank, we simply take the uniform distribution). Writing $x_i = \mu(R_i)$, we obtain $\mathcal{H}(\mu) = h(x_1, ..., x_n) := \sum_{i=1}^n x_i \log(\frac{|R_i|}{x_i})$. We thus have a convex optimisation problem with linear inequality constraints. To see this, note that the stability constraints – that each $\bigcup_{j \leq i}^{i} R_j$ be (μ, t)-stable – are of the form:

$$\forall i \text{ with } 1 \leq i < n, \quad \frac{x_i}{|R_i|} \geq \frac{t}{1-t} \sum_{j=i+1}^{n} x_j$$

since $\frac{x_i}{|R_i|} = \mu(\omega)$ for each $\omega \in R_i$. So we are maximising the (strictly concave) function $h(\mathbf{x}) = \sum_{i=1}^n x_i \log(\frac{|R_i|}{x_i})$ under one equality constraint $\sum_{i=1}^n x_i = 1$ and linear inequality constraints given by $x_i \geq 0$ $(i \leq n)$ and $g_i(\mathbf{x}) \geq 0$ for each $i < n$, where $g_i(\mathbf{x}) = x_i - |R_i| \cdot \frac{t}{1-t} \sum_{j=i+1}^n x_j$. We want to maximise h on $\mathcal{D} := \{\mathbf{x} \in \Delta^{n-1} \mid g_i(\mathbf{x}) \geq 0 \text{ for all } i < n\}$. Clearly we have a one-to-one correspondence between vectors in $\mathcal{D}$ and distributions in $\mathcal{U}$: for every $\mu \in \mathcal{U}$, we have $(\mu(R_1), ..., \mu(R_n)) \in \mathcal{D}$ with $h(\mu(R_1), ..., \mu(R_n)) = \mathcal{H}(\mu)$, and for each $\mathbf{x} \in \mathcal{D}$ we have a unique measure $\mu \in \mathcal{U}$ with $h(\mathbf{x}) = \mathcal{H}(\mu)$ defined by $\mu(\omega) = x_i/|R_i|$ where $\omega \in R_i$. So $\arg\max_{\mathbf{x} \in \mathcal{D}} h(\mathbf{x})$ uniquely determines $\arg\max_{\mu \in \mathcal{U}} \mathcal{H}(\mu)$.

The optimization region $\mathcal{D}$ is an intersection of closed half-spaces, and so it is a closed convex set. As a subset of the simplex Δ^{n-1}, $\mathcal{D}$ is also bounded. This means that $\mathcal{D}$ is compact: and since the function h is continuous, it admits a maximum on $\mathcal{D}$. Because the function is strictly concave, and we maximise it over a convex set, the maximum is unique[21]. Since

$$\max_{\mathbf{x} \in \mathcal{D}} h(\mathbf{x}) = \max_{\mu \in \mathcal{U}} \mathcal{H}(\mu),$$

such a maximum gives us the desired maximal entropy distribution. Lastly, this distribution is rank-uniform, and so by Observation 2.4.3 it generates an AGM revision. This suffices

[21] When f is strictly concave and $\mathcal{D}$ convex, if $\arg\max_{x \in \mathcal{D}} f(x)$ exists, it is unique [56, Thm 7.14, p. 186].

to establish the theorem. We now show an explicit computation of the maximum entropy distribution.

Explicit form. An explicit solution for the maximum entropy distribution can be obtained as follows. The maximum is reached exactly for the unique $\mathbf{x} \in \Delta^{n-1}$ which satisfies all equalities $g_i(\mathbf{x}) = 0$, where all the g_i constraints are active. Solving this system of equations, an expression for this $\mathbf{x}$ is obtained. Writing $r_i := |R_i|$ and $k := \frac{t}{1-t}$, the solution for $\mathbf{x} = (x_1, ..., x_n)$ is

$$x_n = \frac{1}{\prod_{j=1}^{n-1}(1 + r_j k)} \tag{2.3}$$

$$x_i = \frac{r_i k}{\prod_{j=1}^{i}(1 + r_j k)} \text{ for } i \leq n-1 \tag{2.4}$$

This solution can be checked via Lagrange multipliers by appealing to the Karush-Kuhn-Tucker conditions for convex optimisation [9, p. 244] (checking that this is indeed a feasible solution satisfying the KKT conditions is sufficient, since h is concave, h and the g_i's are all differentiable, and the constraints are all linear). Here is a more elementary (and less cumbersome) argument.

Take any $\mathbf{x} \in \mathcal{D}$ for which the first $i-1$ constraints are active ($g_j(\mathbf{x}) = 0$ for $j < i$), but the i-th one is not, i.e., $g_i(\mathbf{x}) > 0$: this last inequality means we have

$$x_i > \frac{r_i k}{\prod_{j=1}^{i}(1 + r_j k)}$$

We first show that we can then find another $\mathbf{y} \in \mathcal{D}$ with higher entropy – that is, $h(\mathbf{y}) > h(\mathbf{x})$ – and for which $g_j(\mathbf{y}) = 0$ for *all* of $\{1, .., i\}$. Given $\epsilon > 0$, define an (ϵ, i)-*improvement* of $\mathbf{x}$ as

$$\mathbf{x}_i^\epsilon = \big(x_1, ..., x_i - \epsilon, x_{i+1} + \epsilon, ..., x_n\big).$$

Now consider the following lemma (which we prove below):

Lemma 2.4.5
Let $\mathbf{x} \in \mathcal{D}$ with $g_j(\mathbf{x}) = 0$ for all $j < i$ but $g_i(\mathbf{x}) > 0$ $(i < 1)$. We have

$$\forall \epsilon \in \left(0, x_i - \frac{r_i k}{\prod_{j=1}^{i}(1 + r_j k)}\right], \; h(\mathbf{x}_i^\epsilon) > h(\mathbf{x}).$$

Given the lemma, the result easily follows: take any $\mathbf{x} \in \mathcal{D}$ for which some constraint is

inactive. Pick the least such i for which $g_i(\mathbf{x}) > 0$. Take its (ϵ_i, i)- improvement with

$$\epsilon_i = x_i - \frac{r_i k}{\prod_{j=1}^{i}(1 + r_j k)}$$

Proceeding in this way for each subsequent $j > i$, we have

$$h\big(\mathbf{x}\big) < h\big(\mathbf{x}_i^{\epsilon_i}\big) < h\big((\mathbf{x}_i^{\epsilon_i})_{i+1}^{\epsilon_{i+1}}\big) < ... < h\big((...(\mathbf{x}_i^{\epsilon_i})...)_{n-1}^{\epsilon_{n-1}}\big).$$

The last element $(...(\mathbf{x}_i^{\epsilon_i})...)_{n-1}^{\epsilon_{n-1}}$ in this improvement sequence is always equal to the optimal solution $\mathbf{x}^*$ given by equations (2.3) and (2.4), since it is the unique vector $\mathbf{x}^*$ satisfying all $g_i(\mathbf{x}^*) = 0$: at each step $i \leq n-1$, the i-th coordinate is replaced by the term $r_i k / \prod_{j=1}^{i}(1+r_j k)$. Now every point $\mathbf{y} \in \mathcal{D}$ different from $\mathbf{x}^*$ has some inactive constraints, and so we can construct an improvement sequence culminating in $\mathbf{x}^*$, which witnesses that $h(\mathbf{y}) < h(\mathbf{x}^*)$. This improvement procedure corresponds to moving along the gradient of h along the boundary of the polytope $\mathcal{D}$, at each step reaching the intersection of the first i hyperplanes $g_i(\mathbf{x}) = 0$. The procedure terminates at the maximum entropy point $\mathbf{x}^*$, as given by the intersection of all hyperplanes.

Proof of the Lemma.
We first prove the lemma for $(\epsilon, 1)$-improvements: that is, we show:

$$\text{For any } \mathbf{x} \in \mathcal{D} \text{ with } g_1(\mathbf{x}) > 0, \text{ we have } h(\mathbf{x}_1^{\epsilon}) > h(\mathbf{x}) \text{ for all } \epsilon \in \Big(0, x_1 - \frac{r_1 k}{1 + r_1 k}\Big]. \quad (2.5)$$

. Take any $\mathbf{x} \in \mathcal{D}$ for which the first constraint is not active, i.e., $g_1(\mathbf{x}) > 0$ – this means we have

$$x_1 > \frac{r_1 k}{1 + r_1 k}.$$

(Recall that here $r_i := |R_i| \in \mathbb{N} \setminus \{0\}$ and $k := \frac{t}{1-t} > 1$). Now consider an $(\epsilon, 1)$-improvement $\mathbf{x}_1^{\epsilon} = \big(x_1 - \epsilon, x_2 + \epsilon, x_3, ..., x_n\big)$. Their difference in entropy is equal to

$$h(\mathbf{x}_1^{\epsilon}) - h(\mathbf{x}) = x_1 \log\left(\frac{x_1}{x_1 - \epsilon}\right) + x_2 \log\left(\frac{x_2}{x_2 + \epsilon}\right) - \epsilon \log\left(\frac{r_1}{x_1 - \epsilon}\right) + \epsilon \log\left(\frac{r_2}{x_2 - \epsilon}\right) \quad (2.6)$$

$$= x_1 \log\left(\frac{x_1}{x_1 - \epsilon}\right) + x_2 \log\left(\frac{x_2}{x_2 + \epsilon}\right) + \epsilon \log\left(\frac{r_2}{r_1} \cdot \frac{x_1 - \epsilon}{x_2 + \epsilon}\right) \quad (2.7)$$

We show that this is strictly positive for $\epsilon \in (0, x_1 - \frac{r_1 k}{1+r_1 k}]$. First note that

$$\log\left(\frac{r_2}{r_1} \cdot \frac{x_1 - \epsilon}{x_2 + \epsilon}\right) \geq \log(r_2 k) > 0 \quad (2.8)$$

This is so because $x_1 - \epsilon \geq \frac{r_1 k}{1+r_1 k}$, and it follows from and $x_1 + x_2 \leq 1$ that $x_2 + \epsilon \leq 1 - \frac{r_1 k}{1+r_1 k}$:

but then we have that

$$\frac{x_1-\epsilon}{x_2+\epsilon} \geq r_1 k, \text{ and so } \frac{r_2}{r_1}\cdot\frac{x_1-\epsilon}{x_2+\epsilon} \geq r_2 k,$$

establishing (2.8). Now we can write

$$h(\mathbf{x}_1^\epsilon) - h(\mathbf{x}) \geq g(\epsilon) + \epsilon\log(r_2 k), \tag{2.9}$$

with $g(\epsilon) := x_1 \log\left(\frac{x_1}{x_1-\epsilon}\right) + x_2 \log\left(\frac{x_2}{x_2+\epsilon}\right)$. Observe that $\frac{dg}{d\epsilon} = \frac{\epsilon(x_1+x_2)}{(x_1-\epsilon)(x_2+\epsilon)}$ is strictly positive when $\epsilon \in (0, x_1 - \frac{r_1 k}{1+r_1 k}]$. Now we have $g(0)=0$ and $g(\epsilon)$ strictly increases on this interval: it follows from (2.9) that $h(\mathbf{x}^\epsilon) - h(\mathbf{x}) > 0$ for $0 < \epsilon \leq \frac{r_1 k}{1+r_1 k}$, and statement (2.5) is established. This case suffices to prove the Lemma. To see why, consider some $\mathbf{x} \in \mathcal{D}$ with $g_j(\mathbf{x}) = 0$ for all $j < i$ but $g_i(\mathbf{x}) > 0$ $(i < n)$. This means that we have

$$\mathbf{x} = \Big(\frac{r_1 k}{1+r_1 k}, \ldots, \frac{r_{i-1}k}{\prod_{j=1}^{i-1}(1+r_j k)}, x_i, \ldots x_n\Big)$$

with $x_i > \frac{r_i k}{\prod_{j=1}^{i}(1+r_j k)}$. Consider the restricted distribution $\mathbf{y} \in \mathbb{R}^{n-i+1}$ defined by re-normalising:

$$(y_i, ..., y_n) := \frac{1}{1-\sum_{j=1}^{i-1} x_j}(x_i, \ldots, x_n)$$

Note that we have

$$\big(1 - \sum_{j=1}^{i-1} x_j\big) \;=\; 1 - \sum_{j=1}^{i-1}\left[\frac{r_j k}{\prod_{\ell=1}^{j}(1+r_\ell k)}\right] \;=\; \frac{1}{\prod_{j=1}^{i-1}(1+r_j k)}$$

so that

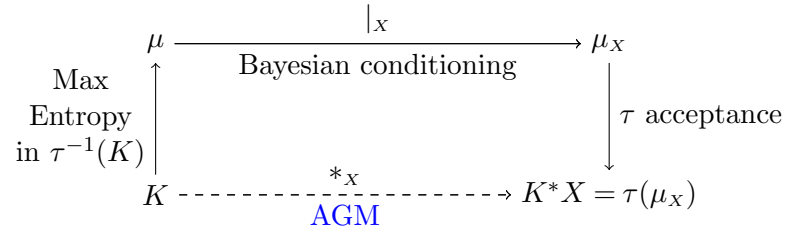
$$y_m = x_m \cdot \Big(\prod_{j=1}^{i-1}(1+r_j k)\Big) \text{ for all } m \in \{i, \ldots, n\}.$$

Since $x_i > \frac{r_i k}{\prod_{j=1}^{i}(1+r_j k)}$, this entails that $y_i > \frac{r_i k}{1+r_i k}$. Now, consider the restricted function $\tilde{h} = \sum_{j=i}^{n} x_j \log(r_j/x_j)$ (h restricted to the last $i-1$ coordinates). We apply the proof of (2.5) to the vector $(y_i, ..., y_n)$, by taking the improvement with $0 < \epsilon \leq y_i - \frac{r_i k}{1+r_i k}$, obtaining

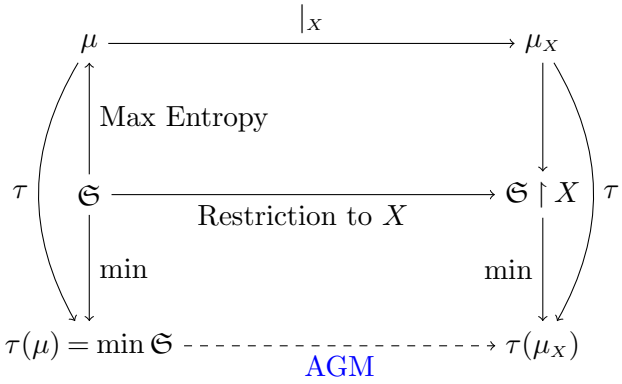
$$\mathbf{y}^* = \Big(y_i - \epsilon,\, y_{i+1} + \epsilon, \ldots, y_n\Big)$$

with $\tilde{h}(\mathbf{y}^*) > \tilde{h}(\mathbf{y})$. Now consider

$$\mathbf{x}^* = \Big(\frac{r_1 k}{1+r_1 k}, \ldots, \frac{r_{i-1}k}{\prod_{j=1}^{i-1}(1+r_j k)}, \alpha y_i^*, \ldots, \alpha y_n^*\Big) \quad \text{with } \alpha = \frac{1}{\prod_{j=1}^{i-1}(1+r_j k)}$$



(a) Recovering AGM from the strongest accepted proposition.



(b) Recovering AGM revision from a plausibility ordering.

Figure 2.5: AGM revision emerges from Bayesian conditioning via the maximum entropy principle.

and observe that this corresponds to the improvement $\mathbf{x}^* = \mathbf{x}_i^{\alpha\epsilon}$ with $\alpha\epsilon \in \left(0, x_i - \frac{r_i k}{\prod_{j=1}^{i}(1+r_j k)}\right]$. Since $(x_i, \ldots, x_n) = (\alpha y_i, \ldots, \alpha y_n)$, we get

$$h(\mathbf{x}^*) - h(\mathbf{x}) = \tilde{h}(\alpha \mathbf{y}^*) - \tilde{h}(\alpha \mathbf{y}) = \alpha[\tilde{h}(\mathbf{y}^*) - \tilde{h}(\mathbf{y})] > 0$$

This concludes the proof of Lemma 2.4.5. □

Thus we can appeal to the maximum entropy principle to pick a unique distribution which generates an AGM revision, even when some more information about the credal state is preserved – i.e., we have the full plausibility ordering encoded in a system of spheres. Assume the agent's doxastic state is specified by a given system of spheres $\mathfrak{S}$ (or, equivalently, the corresponding plausibility ordering); then, following MEP, the agent chooses the relevant maximum entropy measure as a probabilistic representation of the doxastic state. From there, Bayesian conditioning followed by the τ-rule yields the same result as using AGM revision on the initial sphere system directly (i.e., taking the appropriate restriction of the

system of spheres). Figure 2.5 illustrates both ways of recovering AGM revision through the maximum entropy principle.

More explicitly, Proposition 2.4.4 and equations (2.3) and (2.4) give us an immediate solution for the maximum entropy distribution that generates a given ranking: given a system of spheres $\mathfrak{S}$ with ranks $R_1, \ldots, R_n$, the maximum entropy distribution such that $\mathfrak{S}^t(\mu) = \mathfrak{S}$ is the rank-uniform μ determined by:

$$\mu(R_n) = \frac{1}{\prod_{j=1}^{n-1}(1 + |R_j| \cdot \frac{t}{1-t})} \tag{2.10}$$

$$\mu(R_i) = \frac{|R_i| \cdot \frac{t}{1-t}}{\prod_{j=1}^{i}(1 + |R_j| \cdot \frac{t}{1-t})} \text{ for } i \leq n-1 \tag{2.11}$$

so that the probability of each state $\omega \in \Omega$ is given by

$$\mu(\omega) = \frac{\mu(R_i)}{|R_i|} = \frac{\frac{t}{1-t}}{\prod_{j=1}^{i}(1 + |R_j| \cdot \frac{t}{1-t})},$$

where R_i is the rank containing ω. This depends only on the threshold and the size of the ranks. Example 2.4.6 below illustrates a simple application of this result.

Example 2.4.6

Let $t = 3/4$, so that $k := \frac{t}{1-t} = 3$. Suppose we have a system of spheres $\mathfrak{S}$ given by the following ranking:

R_4	ω_{10} ω_{11}
R_3	ω_6 ω_7 ω_8 ω_9
R_2	ω_4 ω_5
R_1	ω_1 ω_2 ω_3

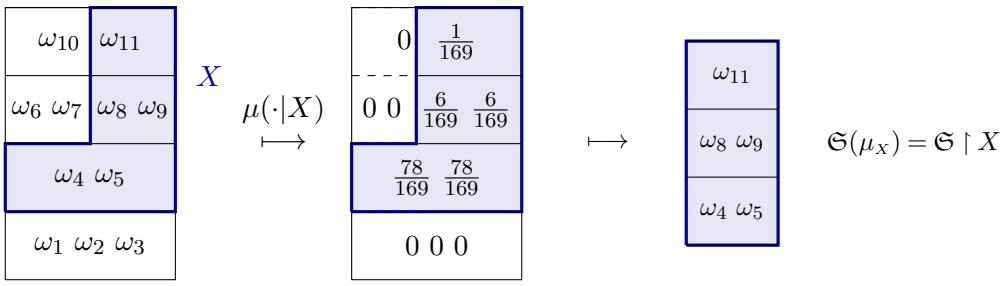
We have $(r_1, r_2, r_3, r_4) = (3, 2, 4, 2)$. Then equations (2.3) and (2.4) give us

$$\Big(\mu(R_1), \mu(R_2), \mu(R_3), \mu(R_4)\Big) = \Big(\frac{819}{910}, \frac{78}{910}, \frac{12}{910}, \frac{1}{910}\Big)$$

and so the maximum entropy distribution that generates this ranking is given by

$$\arg\max_{\mu \in [\mathfrak{S}]} \mathcal{H}(\mu) = \frac{1}{910}(273, 273, 273, 39, 39, 3, 3, 3, 3, 0.5, 0.5)$$

Now consider updating on new information $X = \{\omega_4, \omega_5, \omega_8, \omega_9, \omega_{11}\}$.



We see that $\mathfrak{S}(\mu_X) = \mathfrak{S} \restriction X$*, so that* $\tau(\mu_X) = \tau(\mu)^*X = \{\omega_4, \omega_5\}$*. Note that the choice of the maximum entropy distribution is significant here, as there also exist distributions that generate* $\mathfrak{S}$ *but do not commute with the corresponding AGM revision. Take for instance* $\mu = \frac{1}{4060}\big(1218, 1218, 1218, 300, 80, 6, 6, 6, 6, 1, 1\big)$*. This also generates* $\mathfrak{S}$*, although it is not rank-uniform since* $\mu(\omega_4) > \mu(\omega_5)$*. Here* $\mathfrak{S}(\mu_X)$ *gives the ranking*

$$\{\omega_4\} < \{\omega_5\} < \{\omega_8, \omega_9\} < \{\omega_{11}\}$$

and so $\tau(\mu_X) = \{\omega_4\} \neq \tau(\mu)^*X$*.*

What can we get out of these results? On the one hand, they can be seen as complementing Lin and Kelly's No-Go Theorem by further precisifying the sense in which AGM is too coarse-grained to fully track Bayesian conditioning: it cannot deal with retaining too much information about the probability measure. On the other hand, Propositions 2.4.1 and 2.4.4 do justify the slogan "AGM = τ-rule + Maximum Entropy + Bayesian Conditioning" for situations involving information loss, or an incomplete probabilistic specification of the credal state. An stability-driven agent who complies with the probabilistic principles of maximum entropy and Bayesian conditioning, but who stores her information in a qualitative form – e.g. a belief set together with a plausibility ranking – will automatically comply with AGM revision. Thus AGM can be seen, in this sense, as actually *resulting from Bayesian conditioning*. And if, for instance, one believes that such information loss is inevitable, or if one is generally wary of the sharpness of subjective probabilities, but favourable to the dynamics of Bayesian conditioning – then the above could indicate that AGM revision is a very natural option, even for the staunch Bayesian[22].

This approach invites more research into the information-theoretic aspects of passing from the quantitative to the qualitative framework. For example, the No-Go theorem shows that if *no* information is lost in the specification of the probabilistic credal state,

[22]In particular, this is to be contrasted with Lin and Kelly's [25] theory. Their preferred acceptance/revision duo – the Shoham-driven rule with the so-called *Shoham revision* – does not exhibit the same kind of regularity under information loss: one can show that one cannot, given only the strongest accepted proposition, uniquely recover a Shoham revision operation from the Shoham-driven rule and the MEP.

commutativity fails. Our result shows that, when some of the probabilistic information is forgotten, we can recover a kind of commutativity. It is then natural to wonder *how much* information must be lost in order for AGM to emerge through an acceptance rule, MEP, and Bayesian conditioning: how to measure this information loss? Further, given some reasonably minimal structural constraints on acceptance rules, can we characterise the class of rules that allow for AGM recovery via maximum entropy[23]? Is this class of rules *definable* in a convenient logical framework (and if so, what is its definitional complexity)? Can we provide a useful characterisation of the class of Leitgeb rules (with varying thresholds) as defined by elementary information-theoretic and/or geometric constraints? We leave these questions for another occasion.

We have argued earlier that, as it stands, we have not yet found a truly principled way to temper the conflict between AGM revision and Bayesian conditioning; the prospects for this look dim, at least as long as we insist on preserving our intuitions behind stability and Lockean principles. Now we see that forgetting some information on the way provides an interesting (and arguably less ad-hoc) bridge between the two – one built from Leitgeb's rule and purely probabilistic principles. There may be more to be learnt from it. Perhaps at the root of it all lies another simple moral; agreements cannot always be forced and conflicts cannot always be resolved by negotiation. To make peace, sometimes all it takes is to forget.

2.5 Summary

We introduced Leitgeb's τ-rule and sketched the motivations behind it; in particular, we showed how it derives from the Stability principle and a weakened version of the Lockean thesis. We considered the tracking problem for Bayesian conditioning and AGM in the light of the τ-rule, and saw the way in which Lin and Kelly's No-Go theorem poses a problem when one wants to render AGM revision compatible with Bayesian conditioning. Given (i) the inherent plausibility of the principles behind the τ-rule and (ii) the rule's close connection to AGM revision, we considered some simple ways to approximate tracking using the τ-rule; we found them wanting. Nonetheless, in the light of the τ-rule, we showed that AGM revision may emerge plausibly from probabilistic principles: this happens when only an incomplete probabilistic representation of doxastic states is available.

[23]Natural candidates include the geometrical constraints imposed by Lin and Kelly [25, 26], or convexity requirements for acceptance zones, as advocated by Levi [37].

3

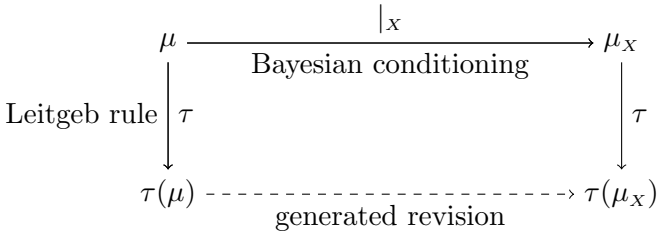
Probabilistically stable revision operators

Our investigations so far highlight several points of discord between AGM-based belief revision theory and the Bayesian model of belief dynamics. While Leitgeb's stability rule is partially successful in bridging the gap between AGM operators and Bayesian conditioning, the harmony between the two is certainly less than perfect.

In this chapter, we will approach the problem from a different angle: instead of asking which acceptance rules can bridge a chosen class of belief revision operators with Bayesian reasoning, we will study the qualitative revisions that emerge from Bayesian dynamics as a result of adopting a given acceptance principle. As we explained in Chapter 2, an interesting aspect of Leitgeb's τ-rule is that it can be justified, as an acceptance principle, by appealing to plausible reasons (stability and Lockean principles) that do not depend on the prior adoption of any particular qualitative revision operator. This suggests the study of a qualitative counterpart of Bayesian reasoning generated by an independently plausible acceptance rule. Following this idea, we shall drop the requirement that our qualitative revision obey the AGM postulates. Instead, we will consider the case where we keep an acceptance rule fixed, and employ it to obtain a qualitative revision *from* Bayesian conditioning.

Bayesian conditioning naturally generates a qualitative revision operator via Leitgeb's rule (see Figure 3.1). In what follows, we will keep Leitgeb's stability rule fixed and ask: what is the qualitative revision generated by Leitgeb's rule and Bayesian conditioning? In this way, we will obtain a tracking result for Bayesian conditioning which relies on Leitgeb's stability-based acceptance.

A particularly useful approach to describe the resulting revision operators is to capture their essential properties as structural rules for nonmonotonic consequence relations. There is a well-known correspondence between AGM revision and the system R of nonmonotonic logic due to Kraus, Lehmann and Magidor [28], whereby a revision operator generates a



Resulting consequence relation: $X \mathrel{|\!\sim_\tau} Y$ iff $\tau(\mu_X) \vdash Y$

Figure 3.1: Probabilistically stable revision generated by Leitgeb's rule.

canonical class of nonmonotonic consequence relations through a translation based on the *Ramsey test* [25]. In a similar way, it is natural to ask what the nonmonotonic logic generated by Leitgeb's rule is, and how it is related to well-known systems for nonmonotonic reasoning. This perspective offers a helpful starting point towards characterising probabilistically stable revision operators.

After defining the notion of *Ramsey-test probabilistic acceptance models* (introduced in [26]) (§3.1), we will discuss the non-monotonic logic of probabilistic stability generated by Leitgeb's rule (§3.2). We will first observe certain distinctive properties of the consequence relations generated by the stability rule: notably, the fact that they satisfy the rule of Rational Monotonicity for non-monotonic logics, while violating the (Or) rule. We will then explain how probabilistically stable revision operators can be described via selection function models from nonmonotonic logic. The problem of characterising this class of models amounts to finding an axiomatic description of *strongest-stable-set operators*, which send every event X to the strongest stable set given X (that is, they map the update input X to the logically strongest event in the probability space that is probabilistically stable once the prior has been updated by X).

In order to achieve this, we first discuss the geometry of the stability rule: a simple geometric analysis highlights some structural properties of the stability rule that play an important role in the axiomatic description of strongest-stable-set operators. We then prove the main result of this chapter: a representation theorem for strongest-stable-set operators (§3.3). This result, which builds on the theory of comparative probability orders, gives necessary and sufficient conditions for a selection function to capture the behaviour of a strongest-stable-set operator on a finite probability space. These selection functions capture exactly the revision operators generated by the stability rule. This solves the characterisation problem for representable selection structures, and we thus obtain an interesting revision operation which serves as a qualitative representation of Bayesian reasoning.

We then connect this result to work in game theory and combinatorics on the numerical

representations of *simple voting games* [57]: our representation theorem gives a solution to a problem concerning the weighted representation of simultaneous voting games.

Lastly, we discuss the problem of giving a complete axiomatisation for the logic of probabilistically stable belief (§3.4).

3.1 The Ramsey test and τ-models

As explained above, we can employ Leitgeb's rule to obtain qualitative revisions *from* Bayesian conditioning. Starting from a probability distribution μ and new update input X (with $\mu(X) > 0$), consider the (restricted) revision operator $*_\tau$ that takes as input the proposition X and the current belief state $\tau(\mu)$, and outputs the revised belief state $\tau(\mu)^{*_\tau} X := \tau(\mu_X)$, as illustrated in Figure 3.1. We call these operators *probabilistically stable revision operators*.

The general form of our problem is as follows: given an acceptance rule α, what is the class of revision operators $*_\alpha : X \mapsto \alpha(\mu_X)$ generated by it? What is its resulting nonmonotonic logic – i.e., the logic of Bayesian conditional belief generated by α? We can describe such revision operators by studying their associated nonmonotonic consequence relations $\mid\!\sim_{\mathfrak{M}}$ understood via a Ramsey-test semantics. That is, we say that $\varphi \mid\!\sim_{\mathfrak{M}} \psi$ holds if and only if, given the agent's (probabilistic) credal state given by a probabilistic structure $\mathfrak{M}$, conditioning on φ leads, through the acceptance rule α, to a new doxastic state where the agent believes ψ. Equivalently, this means that applying the generated revision $*_\alpha$ on input φ leads the reasoner to accept ψ. More precisely, the models of such a logic are given by the following structures [26]:

Definition 3.1.1 (**Probabilistic α-models**)
Fix an acceptance rule α. Define an α-model *as a structure $\mathfrak{M} := (\Omega, \mathfrak{A}, \mu, [\![\cdot]\!], \alpha)$, where $(\Omega, \mathfrak{A}, \mu)$ is a probability space and $[\![\cdot]\!] : \mathcal{L} \to \mathfrak{A}$ is a valuation*[24]*. Set*

$$\varphi \mid\!\sim_{\mathfrak{M}} \psi \text{ if and only if } \alpha(\mu_{[\![\varphi]\!]}) \subseteq [\![\psi]\!] \text{ or } \mu([\![\varphi]\!]) = 0$$

Here $\mathcal{L}$ is a classical propositional language. Given an acceptance rule α, the consequence relations of the form $\mid\!\sim_{\mathfrak{M}}$ (where $\mathfrak{M}$ is an α-model) provide a description of α-generated revision in the sense above: they characterise conditional belief statements of the form $\varphi \mid\!\sim \psi$, expressing that ψ is believed after applying α-generated revision by φ.

At this point, it is useful to recall some key elements of the nonmonotonic logic framework, as introduced by Kraus, Lehmann and Magidor [28]. Here, 'logics' are identified with a class of consequence relations, and the idea is to classify consequence relations $\mid\!\sim$ via the collection of inference rules under which $\mid\!\sim$ is closed.

[24] Here we take a valuation to be a homomorphism from the boolean language $\mathcal{L}$ to the set-algebra $\mathfrak{A}$: the exact class of models of interest to us will be defined more precisely in section 3.4.

Definition 3.1.2 (**KLM-style nonmonotonic logics**)
System C *consists of the rules* (Ref), (Left Equivalence), (Right Weakening), (Cut) *and* (CM) *below. System* P *consists of all the rules of inference below* except *for Rational Montonicity* (RM). *System* R *is obtained from* P *by adding* (RM). *(Note that* $\vdash$ *denotes classical entailment here).*

$$\frac{}{\varphi \mid\sim \varphi}\ (\mathsf{Ref})$$

$$\frac{\varphi \dashv\vdash \psi \quad \varphi \mid\sim \gamma}{\psi \mid\sim \gamma}\ (\mathsf{Left\ Equivalence}) \qquad \frac{\varphi \mid\sim \psi \quad \psi \vdash \gamma}{\varphi \mid\sim \gamma}\ (\mathsf{Right\ Weakening})$$

$$\frac{\varphi \wedge \beta \mid\sim \gamma \quad \varphi \mid\sim \beta}{\varphi \mid\sim \gamma}\ (\mathsf{Cut})$$

$$\frac{\varphi \mid\sim \psi \quad \varphi \mid\sim \gamma}{\varphi \mid\sim \psi \wedge \gamma}\ (\mathsf{And}) \qquad \frac{\varphi \mid\sim \gamma \quad \psi \mid\sim \gamma}{\varphi \vee \psi \mid\sim \gamma}\ (\mathsf{Or})$$

$$\frac{\varphi \mid\sim \psi \quad \varphi \mid\sim \gamma}{\varphi \wedge \psi \mid\sim \gamma}\ (\mathsf{CM}) \qquad \frac{\varphi \mid\sim \gamma \quad \varphi \not\mid\sim \neg\psi}{\varphi \wedge \psi \mid\sim \gamma}\ (\mathsf{RM})$$

In this setting, we say that an acceptance rule α *validates* an inference rule if and only if, for any α-model $\mathfrak{M}$, the consequence relation $\mid\sim_{\mathfrak{M}}$ is closed under the inference rule. We have a basic notion of derivability: let $\Gamma \cup \{\Phi\}$ a finite set of flat conditionals of the form $\varphi \mid\sim \psi$. Given a system of inference rules S, the notation $\Gamma \vdash_{\mathsf{S}} \Phi$ means that the conditional Φ is derivable from Γ using (finitely many applications of) the rules from the system S. Similarly, we say that a class of probabilistic models $\mathcal{M}$ validates this inference (written $\Gamma \models_{\mathcal{M}} \Phi$) whenever it is the case that, for any model $\mathfrak{M}$ in $\mathcal{M}$, if $\mid\sim_{\mathfrak{M}}$ validates all conditionals in Γ, it also validates the conditional Φ.

In this framework, it is natural to ask what the logic of a fixed acceptance rule is. The non-monotonic logic generated by an acceptance rule α can be characterised through a completeness result: the completeness problem amounts to characterising the consequence relation $\models_{\mathcal{M}}$, where $\mathcal{M}$ is the class of all α-models. In other words, it consists in finding a system of inference rules S such that $\vdash_{\mathsf{S}}$ and $\models_{\mathcal{M}}$ coincide.

For an example of such a completeness theorem, it is worth reminding a result by Lin [38] and Lin and Kelly [25, 26], who provide a completeness theorem for their preferred acceptance rule – *Shoham-driven acceptance* – which we will here denote by κ. Lin and Kelly define the κ-rule as follows. If $(\Omega, \mathcal{P}(\Omega), \mu)$ is a discrete probability space and $q \in \mathbb{R}$

$(q \geq 1)$, we have:

$$\kappa_q : \Delta_{\mathfrak{A}} \to \mathfrak{A}, \text{, defined as}$$
$$\kappa_q(\mu) := \left\{ \omega_i \in \Omega \;\middle|\; \frac{\mu(\omega_i)}{\max_{\omega \in \Omega} \mu(\omega)} \geq \frac{1}{q} \right\}$$

Alternatively, they show we can characterise the acceptance rule and the resulting revision as an order-minimisation operation:

$$\kappa_q(\mu) = \min(\prec_\mu), \text{ where } \omega_i \prec_\mu \omega_j \text{ if and only if } \frac{\mu(\omega_i)}{\mu(\omega_j)} > q$$
$$\kappa_q(\mu)^* X := \kappa_q(\mu_X) = \min(\prec_{\mu(\cdot \mid X)}) = \min(\prec_\mu \restriction X) \text{ for all } X \text{ with } \mu(X) > 0$$

Lin and Kelly obtain the following completeness result for κ-models:

Theorem 3.1.3 (**System P completeness**, Lin [38], Lin and Kelly [26])
Let $\Gamma \cup \{\Phi\}$ a finite set of flat conditionals of the form $\varphi \mathrel{|\!\sim} \psi$, and $\mathcal{K}$ the class of κ-models. Then,

$$\Gamma \vdash_{\mathsf{P}} \Phi \textit{ if and only if } \Gamma \models_{\mathcal{K}} \Phi.$$

System P has been a long-time favourite amongst the systems of nonmonotonic logic [28]. A probabilistic semantics for it was already present in Adams' early work deriving from his Ph.D. thesis [1]. Lin's result provides a semantics for system P that is significantly less cumbersome than Adams' original '$\delta - \epsilon$' account, and more intuitive: it employs the Ramsey test to directly relate conditional probabilities to conditional beliefs. It is interesting to see that we can obtain simple probabilistic semantics for system P in a relatively natural way, via a well-chosen acceptance rule.

Naturally, an analogous question arises in the context of Leitgeb's acceptance principle. What is the logic of Bayesian reasoning based on the τ-rule – the logic of probabilistically stable revision? It would be particularly interesting to see if we can identify the resulting logic as a known member of the well-studied family of KLM-style nonmonotonic logics [28, 49]. As we shall see, however, the logic of stability-based acceptance is a rather unusual beast.

Firstly, as we will see next, probabilistically stable revision validates certain strong monotonicity principles, while failing even mild instances of case-reasoning. This already places the resulting logic outside the main classical systems of non-monotonic reasoning. Importantly, this logic does not admit *preferential* semantics, whereby the revision operation is represented as a minimisation operator for an underlying plausibility order (see [28]; we shall discuss this more in detail in §3.2.3). This is in sharp contrast with Lin and Kelly's κ-rule: their completeness result relies on a characterisation of the κ-rule as an

order-minimisation operator, which allows them to associate the class of κ-models with preferential structures with partial orders, known to provide semantics for system P [28].

As we cannot associate probabilistically stable revision operators with preferential models based on plausibility orders, no such translation into preferential semantics is readily available. Thus preferential models are of no help our main task of giving a qualitative structural description of probabilistically stable revision; in what follows, we shall instead solve the problem by appealing to selection function semantics and the theory of comparative probability orders. This being said, a useful starting point consists in investigating various salient patterns of inference validated by stable revision operators: we turn to this now.

3.2 The logic of Leitgeb acceptance

We begin with a brief comparison of the logic generated by the τ-rule with well-known systems from the nonmonotonic logic literature. We note two properties of the τ-rule which make the resulting logic rather unusual: the τ-generated consequence relations validate Rational Monotonicity, while they do not validate the (Or) rule. We briefly discuss the significance of these facts and compare probabilistically stable revision with AGM revision.

We then turn to a discussion of the representation problem. Our task is to find a class of purely 'qualitative' (non-probabilistic) structures that capture the behaviour of the τ-generated revision operators, so as to reveal the key structural properties of probabilistically stable reasoning. We introduce *strongest-stable-set operators* – a convenient way to conceive of probabilistically stable revision – and motivate the use of *selection structures* to describe their behaviour. In the remainder of the section, we clarify certain straightforward, but informative geometrical aspects of the representation problem and make a first connection to the theory of comparative probability orders.

3.2.1 Some preliminary observations

We want to capture the following relation. Take some probability space of the form $(\Omega, \mathfrak{A}, \mu)$, and consider the consequence relation $\mid\!\sim_\mu$ defined directly on $\mathfrak{A}$ as:

$$A \mid\!\sim_\mu B \text{ if and only if } \tau(\mu_A) \subseteq B \text{ or } \mu(A) = 0,$$

i.e., the strongest stable proposition conditional on A entails B or, in other words, the agent's belief set contains B after learning A. Each relation $\mid\!\sim_\mu$ (or, equivalently, each qualitative revision generated by τ) represents a doxastic state together with 'contingency plans'. The current unconditional beliefs $\tau(\mu)$ are given by all A such that $\Omega \mid\!\sim_\mu A$ – that is, all propositions that $\mid\!\sim_\mu$-follow from the tautology. All other entailments of the form $A \mid\!\sim_\mu B$ specify the agent's contingency plan for revision.

Here, we work with a finite sample space Ω (size n) and an algebra $\mathfrak{A}$ over it, which we assume to be the whole powerset. Note that we also avoid issues concerning valuations and the definability of sets of states in Ω via Boolean formulae: we carry on treating propositions as subsets of Ω.

We can immediately note some interesting features of the resulting consequence relation. First of all, it directly follows from this setup that (Left Equivalence) and (Right Weakening) hold, as the τ-rule operates directly on an algebra of propositions. Since belief states are always closed under deduction, it follows that (And) is validated, as well. Next, it follows from the definition of stability that $\tau(\mu_A) \subseteq A$, and so (Ref) is validated, too. But now there are two particularly interesting aspects to this version of nonmonotonic consequence. Firstly, note that Rational Monotonicity is validated:

Observation 3.2.1
The τ-rule satisfies (RM)*: that is, for any measure μ, we have that*

$$\textit{If } A \not\sim_\mu B^c \textit{ and } A \mathrel{|\!\sim_\mu} C, \textit{ then } A \cap B \mathrel{|\!\sim_\mu} C.$$

Proof. We need to show

$$\text{If } \tau(\mu_A) \not\subseteq B^c, \text{ and } \tau(\mu_A) \subseteq C, \text{ then } \tau(\mu_{A\cap B}) \subseteq C.$$

Assume $\tau(\mu_A) \not\subseteq B^c$ and $\tau(\mu_A) \subseteq C$. This entails $\tau(\mu_A) \cap B \neq \emptyset$, and moreover $\tau(\mu_A) \cap B \subseteq C$. We prove that $\tau(\mu_{A\cap B}) \subseteq \tau(\mu_A) \cap B$. It is enough to show that $\tau(\mu_A) \cap B$ is stable with respect to $\mu_{A\cap B}$: the desired inclusion then follows since $\tau(\mu_{A\cap B})$ is the $\subseteq$-least $\mu_{A\cap B}$-stable set. So let $Y \in \mathfrak{A}$ such that $(\tau(\mu_A) \cap B) \cap Y \neq \emptyset$ and $\mu_{A\cap B}(Y) > 0$. We need to show that

$$\mu_{A\cap B}\Big(\tau(\mu_A) \cap B \,\Big|\, Y\Big) > t.$$

Note that $\mu_{A\cap B}(Y) > 0$ entails $\mu_A(B \cap Y) > 0$, and since $\tau(\mu_A) \cap (B \cap Y) \neq \emptyset$, the μ_A-stability of $\tau(\mu_A)$ entails $\mu_A(\tau(\mu_A) \,|\, B \cap Y) > \mathrm{t}$. But then

$$\begin{aligned}\mu_A\Big(\tau(\mu_A) \,\Big|\, B \cap Y\Big) &= \mu_A\Big(\tau(\mu_A) \cap B \,\Big|\, B \cap Y\Big) \\ &= \mu_{A\cap B}\Big(\tau(\mu_A) \cap B \,\Big|\, Y\Big) > t,\end{aligned}$$

as desired. So the set $\tau(\mu_A) \cap B$ is $\mu_{A\cap B}$-stable, and therefore $\tau(\mu_{A\cap B}) \subseteq \tau(\mu_A) \cap B \subseteq C$. □

Note that nothing in the proof of Observation 3.2.1 relies on our specific choice of threshold: as the proof shows, (RM) continues to hold for any choice of threshold.

Secondly, τ-acceptance does *not* satisfy the (Or) rule.

Observation 3.2.2

For any discrete set algebra $(\Omega, \mathfrak{A})$ *with* $|\Omega| \geq 3$*, there is a measure* μ *on it such that* $\mathrel{|\!\sim_\mu}$ *fails the* (Or) *rule.*

Proof. A simple counterexample: let $\Omega = \{\omega_1, \omega_2, \omega_3, \ldots, \omega_n\}$. Let $\mu = (0.4, 0.3, 0.3, 0, \ldots, 0)$, and set $A = \{\omega_3\}$, $B := \{\omega_1, \omega_2\}$, and $C := \{\omega_1, \omega_3\}$. For a threshold $t = 1/2$, we can easily compute $\tau(\mu_A) = A$, $\tau(\mu_B) = \{\omega_1\}$, and $\tau(\mu_{A\cup B}) = A \cup B$. So we have $A \mathrel{|\!\sim_\mu} C$ and $B \mathrel{|\!\sim_\mu} C$, but $A \cup B \not\mathrel{|\!\sim_\mu} C$. □

Restricting attention to the space $\{\omega_1, \omega_2, \omega_3\}$ in the example above, C is $\mathrel{|\!\sim_\mu}$-entailed by both A and its complement A^c, but is not $\mathrel{|\!\sim_\mu}$-entailed by the tautological disjunction $A \cup A^c$. In logical terms, this means that the τ-rule fails an even weaker principle, capturing the restriction of (Or) to mutually exclusive and exhaustive propositions:

$$\frac{\varphi \mathrel{|\!\sim} \psi \quad \neg\varphi \mathrel{|\!\sim} \psi}{\mathsf{T} \mathrel{|\!\sim} \psi} \ (\mathsf{exOr})$$

Given an algebra of propositions, how common are measures that fail the (Or) rule? Building on counterexamples like the above, one can rather easily show that for any discrete set algebra $(\Omega, \mathfrak{A})$ and any threshold t, there is an open neighbourhood of distributions $\mathcal{N}$ in the probability simplex such that, for any $\mu \in \mathcal{N}$, $\mathrel{|\!\sim_\mu}$ fails the (Or) rule. In this sense, the failure of (Or) is a non-negligible aspect of how the τ rule behaves[25].

These characteristics of τ-generated consequence relations are rather unusual, and they make it difficult to place, be it very approximately, the resulting logic of probabilistic stability on the family tree of known nonmonotonic logics. To begin with, most known nonmonotonic

[25]There is a natural connection between the (Or) rule and the oft-discussed principle of *conglomerability* (notably discussed – and rejected – by De Finetti [16]). We say a probability measure μ is (countably) *conglomerable* if, for any countable partition Π of the underlying probability space, we have

$$\mu(X) \in \big[\inf_{E\in\Pi} \mu(X \mid E), \sup_{E\in\Pi} \mu(X \mid E)\big]$$

for any $X \in \mathfrak{A}$. In other words, conglomerability requires that the unconditional probability of an event remains within the bounds fixed by the most extreme values its probability could take when conditioning on cells from the partition. That is, conglomerability requires the following: if learning any event $E \in \Pi$ yields a probability $\mu(\cdot \mid E)$ such that $\mu(X \mid E) \in [a, b]$ (for some fixed $a, b \in [0, 1]$ with $a \leq b$), then we should already have $\mu(X) \in [a, b]$ unconditionally. The analogy with the (Or) rule is rather immediate (but see Howson [22] who defends the view that the analogy with an infinitary Or-introduction rule is 'merely illusory' [22, p. 12], if this Or-rule is considered as a purely deductive inference rule).

Non-conglomerability can only occur if the measure μ fails countable additivity [51] and, with a few extra assumptions, Seidenfeld et al. have shown that this phenomenon generalises to higher cardinalities [52] (that is, whenever the measure is not κ-additive, conglomerability fails for partitions of cardinality at most κ). The failure of the (Or) rule can be seen as showing that the stability rule forces a finite, qualitative counterpart of non-conglomerability, even though the underlying measures are always assumed to be conglomerable; for we can have a partition $\Pi = \{A, A^c\}$ and an event $X \in \mathfrak{A}$ such that X has a μ_A-stable subset *and* a μ_{A^c}-stable subset, but *no* μ-stable subset.

systems in the literature, like McCarthy's circumscription logics [39], Reiter's default logics [40], systems P and R [28], as well as most logics of conditionals, extend system C. But a small modification of the counterexample to (Or) used in the proof of Observation 3.2.2 shows that the τ-generated consequence relations do not in general satisfy (Cut); hence, we are not dealing with a C-complying class of consequence relations (and of course, our consequence relations of the form $\mid\sim_\mu$ do not extend system P). Further, τ-consequence is always closed under (RM); but the (RM)-complying systems presented in the literature usually also satisfy the (Or) rule, which τ-consequence violates. This means that we are dealing with a notion of consequence which is, in one sense, unusually weak, in that it does not validate very common (and, arguably, rather intuitive) rules like (Or) or (Cut). In another sense, however, this notion of consequence is rather strong, for it validates (RM), itself considered a strong requirement on nonmonotonic consequence. All this indicates that we need to do some more work to isolate the correct rules of stability-based reasoning. Before we address this problem, however, let us very briefly point out two interesting differences between τ-generated revision and AGM-compliant operators.

3.2.2 AGM and τ-generated revision

In Chapter 2, we studied the AGM revision operators obtained via the τ-rule, which consisted in generating a system of stable sets from the prior, and then revising the doxastic state via the resulting ordering on states. There is a sense in which this procedure does not exploit much of the *dynamics* of probabilistic credences: revisions depend entirely on this initial ordering. By contrast, τ-generated revision operators faithfully describe the dynamics of probabilistically stable reasoning under Bayesian conditioning: the τ-rule is applied at each conditioning step, and the changes in the agent's beliefs closely mirror the changes that conditioning brings to the structure of stable sets – such as, crucially, the formation of new stable sets (spheres).

The most evident distinction between those two revision mechanisms is of course that τ-generated revisions, by their very design, commute with Bayesian conditioning, while the AGM procedure above does not, as we discussed at length in the previous chapter.

On the logical side, another difference between AGM and τ-generated revision is that the latter fails the (Or) rule, while our AGM-driven revision always satisfies (Or) (it is a straightforward exercise to show that this holds for any AGM operator). This failure is somewhat troubling, and some may see it as quite a damning feature of τ-generated revision.

The usual arguments for the (Or) rule present it as a commonsense desideratum for handling *case reasoning*. The rationale for case reasoning is best illustrated in the context of arguments for the (weaker) rule (exOr). Suppose we have $\varphi \mid\sim \psi$ and $\neg\varphi \mid\sim \psi$, while also having $\varphi \vee \neg\varphi \not\mid\sim \psi$. This means that the agent believes ψ conditionally on learning either φ

or $\neg\varphi$, but does not currently accept ψ. This seems to run counter to the following intuitive principle, expressed by Lin and Kelly:

> if you know that you will accept a proposition regardless what you learn, you should accept it already. [25, p. 960]

The rule can also be seen as analogous to Savage's decision-theoretic *sure-thing principle* [48]: if, in a decision problem, the agent knows she would select action a conditional on event X being true, *and* she would select action a conditional on its negation X^c, then she ought to select action a outright.

A natural way to excuse the failure of the (Or) rule consists in marking the distinction between *hypothetical* conditionals and *actual update* conditionals. On the first reading, the expression $\varphi \mathrel{|\!\sim} \psi$ captures a contingency plan that is transparent to the agent herself: under this reading, the agent knowingly commits to accepting ψ in both cases φ and $\neg\varphi$. Refusing to accept ψ appears incongruous, since she also knows that one of φ and $\neg\varphi$ already obtains (even though she may not know which).

On the second reading, in the conditional $\varphi \mathrel{|\!\sim} \psi$, the antecedent φ ought to be read as a *truly dynamic* operator: in a spirit closer to dynamics logics [45, 5], we could conceive of learning φ as an *event* taking place. Following the dynamic tradition, we can suggestively write this informational event as $[!\varphi]$ to distinguish it from φ, the mere proposition *that* φ *holds*. Then the formula $\varphi \mathrel{|\!\sim} \psi$ – properly read as $[!\varphi]\psi$ – expresses that the informational state of the agent is such that, after *actually, truthfully learning* φ, she accepts ψ. It may be the case that the information events $[!\varphi]$ and $[!\neg\varphi]$ both lead her to accept ψ; but, at the current stage, neither event actually happened, so nothing prompted the agent to adopt that belief. On this reading, there is no rationality failure on the part of the agent: for one thing, she need not believe that either event $[!\varphi]$ or $[!\neg\varphi]$ will occur (even though she knows that either φ or $\neg\varphi$ is the case). For another, the outcomes of learning events need not be transparent to her.

In what follows, we will remain neutral on this interpretative issue. In particular, we will not pursue in detail this second reading here: it is most fruitful to first understand the behaviour of basic stability-induced conditionals, regardless of matters of interpretation, before we can resort to more expressive logics that capture the distinction between propositions φ and their corresponding learning events $[!\varphi]$. In any case, while the failure of (Or) may be unsettling, it also renders the logic of probabilistic stability an unusual and interesting object of study.

On the other hand, probabilistically stable revision can outperform AGM revision in probabilistic learning tasks. This important difference between AGM operators and probabilistically stable revision emerges when considering simple statistical learning scenarios. Consider the following:

Example 3.2.3

There are two urns: one urn with only white balls (call it W*), and one with black and white balls in equal proportions (call it* BW*). The learner is presented with urn* W*, but she does not know which of the two urns she was given: her priors assign probability 1/2 to each urn, and assumes the draws are independent and identically distributed (with the obvious Bernoulli distribution corresponding to each urn: the urn* W *corresponds to the parameter* $p = 1$ *for the probability of a white ball, and* BW *corresponds to the parameter* $p = 1/2$*). She repeatedly draws balls from the urn, with replacement, and adjusts her credences accordingly via conditioning. In this context, the learner can select a belief revision procedure to be followed in parallel to the updating of her subjective probability measure. Here we compare two such procedures.*

The first procedure is the one discussed in Chapter 2, which consists in generating a system of spheres form the prior, and then revising her doxastic state via the resulting AGM revision operation. The second one is τ*-generated revision: at each draw of a ball from the urn, the agent conditions on the current evidence, after which she applies the* τ*-rule to determine her doxastic state. Now, the question is: which of those two procedures will allow the agent to learn the correct urn? That is, as the agent draws white ball after white ball, is there a time at which she will come to believe that the urn is* W*, and stick to this belief forever after?*

We represent each possible sequence of draws as a binary sequence, with a draw of a white ball denoted by 0 and a black ball by 1. We assume that each experiment involves a finite number k *of draws, and we consider what happens when we increase the number of draws in the experiment*[26]*. At each stage* k*, our space contains basic propositions of the form* (U, x)*, with* $\mathsf{U} \in \{\mathsf{W}, \mathsf{BW}\}$ *and* $x \in \{0,1\}^k$*, representing the draw of sequence* x *from urn* U*. Thus, the hypothesis* W *is identified with* $\{(\mathsf{W}, x) \mid x \in \{0,1\}^k\}$*.*

Now, consider the AGM reasoner. Note that, given her prior distribution μ*, there exist no stable sets* X *of measure less that 1. This is because, in order to exceed the threshold* t*, any stable set* S *would have to contain at least one state of the form* (BW, x)*, for some* $x \in \{0,1\}^k$*, where*

$$\mu(\mathsf{BW}, x) = \mu(\mathsf{BW}) \cdot \mu(x \mid \mathsf{BW}) = \frac{1}{2} \cdot \frac{1}{2^k} = \frac{1}{2^{k+1}}.$$

[26] This is a cumbersome way of modeling the learning process here. Of course, it is much more natural to model this example using a single infinite space, e.g. by taking a sample space $\{\mathsf{W}, \mathsf{BW}\} \times 2^\omega$ and an algebra generated by basic opens of the form $(S, [\![x]\!])$ with $S \in \{\mathsf{W}, \mathsf{BW}\}$ and $[\![x]\!] = \{X \in 2^\omega \mid X \text{ extends } x\}$. However, in the infinite case there are some subtleties to address concerning the fact the space is nonatomic, which trivialises the notion of stability (as any set with measure below one has defeaters – see brief discussion in Chapter 4). There is also no least set of measure 1, and so we need to modify the notion of acceptance slightly. In order to avoid those difficulties, it is more convenient here to model the scenario as involving increasingly large but finite spaces representing increasing numbers of draws.

Since $\mu(S) < 1$, there must also be a state of the form (BW, y) in S^c, where we also have $\mu(\mathsf{BW}, y) = 1/2^{k+1}$. But then

$$\mu(S \mid S^c \cup \{(\mathsf{BW}, x)\}) \leq \frac{\mu(\mathsf{BW}, x)}{\mu(\mathsf{BW}, x) + \mu(\mathsf{BW}, y)} = 1/2,$$

and so the proposition $S^c \cup \{(\mathsf{BW}, x)\}$ is a defeater for S.

Thus, the AGM-complying reasoner starts by believing only the strongest probability 1 proposition, which is consistent with $(\mathsf{BW}, 0 \ldots 0)$ (that is, it is consistent with the proposition that only white balls are drawn from the mixed urn BW). But then the BW hypothesis will never be eliminated by AGM revision, since any piece of information received by further draws – namely, any sequence of 0's – remains logically consistent with BW. Thus, the mixed-urn hypothesis is never set aside and so W cannot be learnt.

Using the τ-generated revision, on the other hand, allows the reasoner to learn that the urn is W. After n trials, the learner observes a sequence $0^{(n)} := (0, \ldots, 0)$ of n white balls. Her credence in W is then

$$\mu(\mathsf{W} \mid 0^{(n)}) = \frac{1/2}{1/2 + 1/2^{n+1}} = \frac{1}{1 + 1/2^n}$$

and thus, $\lim_{n\to\infty} \mu(\mathsf{W} \mid 0^{(n)}) = 1$. At some point, we will pass the stability threshold t and have $\mu(\mathsf{W}, 0^{(k)}) > \frac{t}{1-t} \cdot \mu(\mathsf{BW}, 0^{(k)})$: the agent comes to believe W, her strongest stable proposition being the singleton set $\{(\mathsf{W}, 0^{(k)})\}$.

As this example illustrates, there exist cases where AGM-driven revision never learns the correct outcome, while stability reasoning does.

3.2.3 Towards a representation theorem: qualitative models

The τ-generated revisions take the form $\tau(\mu) \mapsto \tau(\mu_X)$, where μ is the agent's subjective probability distribution and X a new revision input. Keeping the initial doxastic state $\tau(\mu)$ implicit in the background, we can fully characterise each such revision as a map $X \mapsto \tau(\mu_X)$ (sending a proposition $X \in \mathfrak{A}$ – the revision input – to another proposition $\tau(\mu_X)$ – the strongest accepted proposition, representing the updated belief state). Each such map can be seen as a *strongest-stable-set operator*, sending each X to the strongest ($\subseteq$-least) stable set given X. An important step towards solving the representation problem is to characterise those revisions in a purely qualitative way: that is, to describe all maps $X \mapsto \tau(\mu_X)$ in a way that does not depend on the underlying probability measures μ.

A natural and rather prominent model for qualitative revision operators is given by *order-based* revisions: these are revisions that are defined by order-minimisation and "qualitative" conditioning. Starting from an order $\leq$ defined on Ω, the agent's accepted propositions are

given by the set $\min(\leq)$. Revision by an event X amounts to restricting $\leq$ to X and taking the collection of all $\leq$-minimal elements of the restricted order. In short, order-based revisions are of the form $\min(\leq) \mapsto \min(\leq\restriction X)$, where $\leq$ is an order defined on Ω, and X the event learnt by the agent. Many qualitative revision operators admit natural characterisations in terms of order-based revisions – not least among which are AGM operators as well as Lin and Kelly's κ-generated revision. It is natural to ask if one could give such an order-based characterisation for the class τ-generated revision operators: that is, whether each map taking X to $\tau(\mu_X)$ can be described as an operation of the form $X \mapsto \min(\leq\restriction X)$. Unfortunately, it is easy to see that this cannot be done:

Example 3.2.4
Let $t=1/2$, $\Omega=\{\omega_1,\omega_2,\omega_3\}$ with μ given by (0.4,0.35,0.25) – that is, $\mu(\omega_1)=0.4$, $\mu(\omega_2)=0.35$ and $\mu(\omega_3)=0.25$. Let $X:=\{\omega_1,\omega_2\}$. Note that $\tau(\mu)=X$: the hypothesis X is already accepted by the agent. Now suppose the agent now learns X with certainty: we have the updated probabilities given by $(0.5\bar{3}, 0.4\bar{6}, 0)$ and the the new belief set is given by $\{\omega_1\} \neq X$. But any order $\leq$ on Ω for which $\min(\leq)=X$ will also satisfy $\min(\leq\restriction X)=X$, and so the order-based revision by X will not change the belief set.

What this example illustrates is that τ-generated revision cannot be *tracked* using order-based revision: that is, there is no way to translate each probability measure μ into an order $\leq$ on Ω so that each operation $\min(\leq) \mapsto \min(\leq\restriction X)$ coincides with the revision $\tau(\mu) \mapsto \tau(\mu_X)$.

This is one reason why here we follow a more fruitful approach, which consists in employing *selection structures*. Selection structures are of the form $\mathfrak{M}=(\Omega,\mathfrak{A},\sigma)$, with $(\Omega,\mathfrak{A})$ a set-algebra of propositions, and a map $\sigma:\mathfrak{A}\to\mathfrak{A}$ called a *selection* function. The problem consists in imposing the right axioms on σ so that it behaves like a τ-generated revision map. Then, for any A, $\sigma(A)$ represents the strongest probabilistically stable proposition after conditioning on A (so that we can say that σ *selects* the strongest accepted proposition conditional on A).

To each selection structure $\mathfrak{M}=(\Omega,\mathfrak{A},\sigma)$ corresponds a relation $\mathrel{|\!\sim_\sigma}$ defined as:

$$A \mathrel{|\!\sim_\sigma} B \text{ if and only if } \sigma(A)\subseteq B.$$

We want to impose axioms on selection functions such that, for each selection structure $\mathfrak{M}=(\Omega,\mathfrak{A},\sigma)$, there is a probability measure μ on $(\Omega,\mathfrak{A})$ such that $\mathrel{|\!\sim_\mu}=\mathrel{|\!\sim_\sigma}$. This means that μ *represents* σ, in the sense that:

$$\begin{aligned}\forall A\in\mathfrak{A},\ \sigma(A)=\tau(\mu_A) &\text{ if } \mu(A)\neq 0,\\ \text{and } \sigma(A)=\emptyset \qquad &\text{if } \mu(A)=0.\end{aligned}$$

Yet another description of our problem is as follows: for each probability space $(\Omega, \mathfrak{A}, \mu)$ taken together with τ, consider the map $\tau(\mu_{(\cdot)}) : \mathfrak{A} \to \mathfrak{A}$ taking each proposition X to $\tau(\mu_X)$, the strongest accepted proposition conditional on X (here we can slightly change the definition of τ and conveniently assume that $\tau(\mu_X) = \emptyset$ whenever $\mu(X) = 0$). We want to show that the class of selection structures $(\Omega, \mathfrak{A}, \sigma)$ coincides with the class of structures $(\Omega, \mathfrak{A}, \tau(\mu_{(\cdot)}))$. Identifying the representable selection structures amounts to fully characterising probabilistically stable revision operators, in the following sense. Each probability measure generates a particular epistemic state $(K(\mu), *)$ where $K(\mu) := \tau(\mu)$ represents the current belief set of the agent and $* : \mathfrak{A} \to \mathfrak{A}$ is a revision operator mapping each revision input X to the revised state $K^*X := \tau(\mu_X)$. Each such epistemic state $(K(\mu), *)$ can be described by a single selection structure $(\Omega, \mathfrak{A}, \sigma)$, where $\sigma(\Omega)$ represents the unconditional beliefs $K(\mu)$ and $\sigma(X)$ the revised state K^*X. We want to find an axiomatic description of the class of selection structures that correspond exactly to epistemic states $(K(\mu), *)$, where $*$ is a strongest-stable-set operator.

In what follows, we will build our way towards such a representation result:

- We begin by providing a geometric characterisation of consequence relations, which establishes that each τ-generated consequence relation can be uniquely identified by a specific system of linear inequalities. This allows to identify certain important properties that the selection function σ must satisfy in order to be probabilistically representable.

- Next, we focus on a representation for the special case of Leitgeb's rule with strict threshold $t = 1/2$. This case is a natural choice form the logical side, since it can be seen as the most 'qualitative' version of stability-based acceptance, as advocated by Leitgeb [32]. We will see that the representation problem for the case $t = 1/2$ bears a close connection with the theory of comparative probability orders. We exploit this connection in our representation theorem. We first prove a useful lemma giving sufficient conditions for the (simultaneous) probabilistic representability of two comparative probability orders (one strict, and one non-strict) on a finite algebra; then, we derive from this lemma our desired representation result for selection structures. As we will see, the key condition allowing the probabilistic representability of a selection rule is a special version of the Scott axiom from the theory of comparative probability [50].

3.2.4 The geometry of τ-generated revision.

We consider the case of finite probability spaces of the form $(\Omega, \mathcal{P}(\Omega), \mu)$, with $\Omega = \{\omega_1, \ldots, \omega_n\}$ a finite sample space. We provide a geometric characterisation of the represen-

tation problem, which helps isolate relevant structural properties of selection functions. We begin with the following observation:

Proposition 3.2.5
Let $(\Omega, \mathcal{P}(\Omega), \mu)$ a finite probability space, and $t \in [0.5, 1)$ a threshold. Then for any A such that $\mu(A) > 0$, we have that $\tau(\mu_A) = B$ if and only if the following hold:

(i) $\forall \omega \in B,\ \mu(\omega) > \frac{t}{1-t} \cdot \mu(A \setminus B)$

(ii) $\forall X \subset B,\ \exists \omega \in X,\ \mu(\omega) \leq \frac{t}{1-t} \cdot \mu(A \setminus X)$.

Proof. We sketch it only, as it easily follows from the definition of stability (a proof is implicit in Leitgeb's discussion of an algorithm for generating stable sets in [30, p.1363]). The key fact is that a proposition X is (μ, t)-stable exactly when there are no *defeater states* – that is, $\omega \in X$ such that $\mu(\omega \,|\, X^c \cup \{\omega\}) \leq t$. Then we use the fact that $\mu(\omega) \leq \frac{t}{1-t} \cdot \mu(A \setminus X)$ is equivalent to $\mu_A(\omega \,|\, X^c \cup \{\omega\}) \leq t$. Condition (i) says that no $\omega \in B$ is a defeater w.r.t A: so B is indeed μ_A-stable. Condition (ii) says that each strict subset of B has at least one defeater w.r.t A: so no strictly smaller subset is μ_A-stable. The two conditions together mean that B is indeed the $\subseteq$-least μ_A-stable set. □

Note that either side of the biconditional entails $B \subseteq A$, so we can rewrite Proposition 3.2.5 in the following equivalent form:

$$\tau(\mu_A) = B \text{ if and only if } B \subseteq A \text{ and}$$

(i) $\forall \omega \in B,\ \mu(\omega) > \frac{t}{1-t} \cdot \mu(A \setminus B)$;

(ii) $\forall X \subset B,\ \exists \omega \in X,\ \mu(\omega) \leq \frac{t}{1-t} \cdot \mu(A \setminus X)$.

This very elementary observation will be useful in building the bridge with qualitative models.

Proposition 3.2.5 points to an important fact about the behaviour of the τ-rule. By definition, each $\mid\!\sim_\mu$ is clearly determined by the values of $\tau(\mu_A)$ for each A. In essence, Proposition 3.2.5 reduces the problem of determining $\tau(\mu_A)$ to that of comparing inequalities of the form $\mu(\omega) > \frac{t}{1-t} \cdot \mu(X)$ and $\mu(\omega) \leq \frac{t}{1-t} \cdot \mu(X)$ for certain sets X.

This reasoning gives us a useful geometric characterisation of the $\mid\!\sim_\mu$ relations. Consider the simplex Δ^{n-1} of all distributions over Ω. For any μ, $\rho \in \Delta^{n-1}$, we would like to know when $\mid\!\sim_\mu = \mid\!\sim_\rho$ – i.e., when two distributions generate the same consequence relations. Write $\mu \sim \rho$ whenever this is the case. Since we want each selection function σ to represent one particular consequence relation, this means that we want axioms for selection functions such that, for any σ, the set of distributions in Δ^{n-1} representing σ is an equivalence class of $\sim$.

What do those equivalence classes look like, and how do we find the selection functions that pick out exactly those classes?

The proposition above allows to visualise each $\sim$-equivalence class as a certain convex polytope in Δ^{n-1}. Define the following:

Definition 3.2.6 (**Fixed-odds hyperplanes**)
Let $\mathcal{H}_n$ be the collection of hyperplanes in $\mathbb{R}^n$ defined by equations of the form

$$x_i = \frac{t}{1-t}\Big(\sum_{x_j \in X} x_j\Big), \text{ where } X \text{ is a set of variables such that } x_i \notin X.$$

We call $\mathcal{H}_n$ the collection of fixed-odds hyperplanes.

We are interested in the way that $\mathcal{H}_n$ partitions the probability simplex Δ^{n-1}. For each $\mu \in \Delta^{n-1}$, it is enough to look at equalities of the form $\mu(\omega_i) = \frac{t}{1-t} \cdot \mu(X)$ for sets $X \subseteq \Omega$ not containing ω_i. We have:

Observation 3.2.7
Let μ, $\rho \in \Delta^{n-1}$. We have $\mid\!\sim_\mu = \mid\!\sim_\rho$ if and only if μ and ρ lie in the same region of the hyperplane arrangement $\mathcal{H}_n$ and have the same support[27].

Proof. Suppose μ and ρ do not lie in the same region of $\mathcal{H}_n$. Then there is some hyperplane with equation $x_i = \frac{t}{1-t}\Big(\sum_{x_j \in X} x_j\Big)$, $x_i \notin X$, that separates them. In terms of probabilities, this means that there is some ω_i and set X with $\omega_i \notin X$ such that (wlog):

$$\mu(\omega_i) > \frac{t}{1-t}\mu(X) \quad \text{and} \quad \rho(\omega_i) \leq \frac{t}{1-t}\rho(X)$$

This entails that $\mu(\omega_i \,|\, X \cup \{\omega_i\}) > t$, which in turn entails that $X \cup \{\omega_i\} \mid\!\sim_\mu \{\omega_i\}$, while by the same reasoning $X \cup \{\omega_i\} \not\mid\!\sim_\rho \{\omega_i\}$, which means $\mid\!\sim_\mu \neq \mid\!\sim_\rho$. Conversely, suppose μ and ρ lie in the same region of $\mathcal{H}_n$. This means that they lie on the same side of every hyperplane in $\mathcal{H}_n$. We then have that for every state ω_i and set X such that $\omega_i \notin X$, we have

$$\mu(\omega_i) > \frac{t}{1-t}\mu(X) \quad \text{iff} \quad \rho(\omega_i) > \frac{t}{1-t}\rho(X) \tag{3.1}$$

This just means that $\mu(\omega_i \,|\, X \cup \{\omega_i\}) > t$ holds if and only if $\rho(\omega_i \,|\, X \cup \{\omega_i\}) > t$. Now let propositions A, $B \subseteq \Omega$ such that $A \mid\!\sim_\mu B$. We assume $\mu(A) \neq 0$ (otherwise, we immediately have $A \mid\!\sim_\rho B$ – this is because μ and ρ have the same support, and therefore if $\mu(A) = 0$ entails $\rho(A) = 0$). Thus $A \mid\!\sim_\mu B$ means $\tau(\mu_A) \subseteq B$. Suppose towards a contradiction that $\tau(\rho_A) \not\subseteq B$. Then, by Proposition 3.2.5, either property (i) or (ii) must fail for ρ. If (i) fails,

[27] That is, μ and ρ agree on which propositions have measure 0.

then

$$\exists \omega \in B,\ \rho(\omega) \leq \frac{t}{1-t} \cdot \rho(A \setminus B).$$

But, by (1), this means that $\mu(\omega) \leq \frac{t}{1-t} \cdot \mu(A \setminus B)$, and so (i) does not hold for μ, contradicting the fact that $\tau(\mu_A) \subseteq B$. Similarly, if (ii) fails, then we have

$$\exists X \subset B,\ \forall \omega \in X,\ \rho(\omega) > \frac{t}{1-t} \cdot \rho(A \setminus X).$$

But then (1) entails that, for all $\forall \omega \in X,\ \mu(\omega) > \frac{t}{1-t} \cdot \mu(A \setminus X)$, which by Proposition 3.2.5 again contradicts the fact that $\tau(\mu_A) \subseteq B$. So both (i) and (ii) hold for ρ, and therefore we have $\tau(\rho_A) \subseteq B$, hence $A \mathrel{|\!\sim_\rho} B$ as required. This concludes the proof. □

Thus we have characterised the way in which $\mathrel{|\!\sim_\mu}$ relations divide the probability simplex: each is determined by hyperplane equations of the form given above[28]. Thus:

- For each sample space of size n, we get a hyperplane arrangement $\mathcal{H}_n$, given by $n \times (2^{n-1} - 1)$ equations of the form $\mu(\omega) = \frac{t}{1-t} \cdot \mu(X)$ for pairs (ω, X) such that $\omega \notin X$.
- Each relation $\mathrel{|\!\sim_\mu}$ can be identified uniquely by checking in which region of $\mathcal{H}_n$ the point μ is.

The significance of this characterisation is that it gives us a strategy for finding the right axioms for selection functions.

3.2.5 Representing selections

Knowing, for each pair (ω, X) with $\omega \notin X$, which of $\mu(\omega) > \frac{t}{1-t} \cdot \mu(X)$ or $\mu(\omega) \leq \frac{t}{1-t} \cdot \mu(X)$ holds amounts to knowing what $\mathrel{|\!\sim_\mu}$ is. But this simply corresponds to checking whether $\mu(\omega \mid X \cup \{\omega\}) > t$ or not. In turn, it is straightforward to see that $\mu(\omega \mid X \cup \{\omega\}) > t$ if and only if $\tau(\mu_{X \cup \{\omega\}}) = \{\omega\}$ (that is, $\{\omega\}$ is the strongest stable proposition with respect to distribution $\mu_{X \cup \{\omega\}}$). This immediately suggests the following desiderata.

Firstly, we want to define selection functions that mirror the behaviour of the τ-rule shown in Proposition 3.2.5. That is, we want σ such that $\sigma(A) = B$ if and only if $B \subseteq A$ and

(i′) $\forall b \in B,\ \sigma((A \setminus B) \cup \{b\}) = \{b\}$;

(ii′) $\forall X \subset B,\ \exists x \in X,\ \sigma((A \setminus X) \cup \{x\}) \neq \{x\}$.

[28] Regions of hyperplane arrangements (chambers) do not contain points on the hyperplanes themselves (they are open regions). What of distributions that lie on the hyperplanes? We simply add all distributions μ satisfying $\mu(\omega_i) = \frac{t}{1-t}\mu(X)$ to the adjacent regions satisfying $\mu(\omega_i) < \frac{t}{1-t}\mu(X)$. Then Observation 3.2.7 still holds for this extended notion of 'regions' (via the same argument). This takes care of all distributions: we have fully partitioned the simplex by $\sim$-equivalence classes.

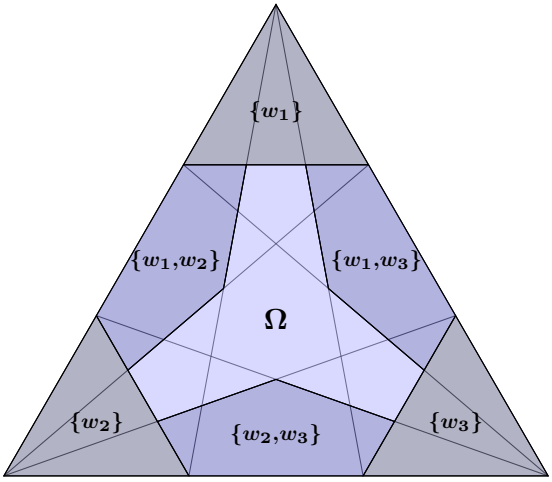


Figure 3.2: Fixed-odds lines for Leitgeb's τ-rule with $t = 2/3$ and $|\Omega| = 3$. Each region of the hyperplane arrangement determines a specific consequence relation $\mid\!\sim$. As before, the colored regions represent acceptance zones – here, they correspond to sets of probability distributions which agree on the unconditional strongest stable set.

Secondly, given σ, we want to be able to translate all statements of the form $\sigma((A \setminus B) \cup \{b\}) = \{b\}$ into a set of linear inequalities that admits a solution in Δ^{n-1}. We construct a system of linear inequalities L_σ as follows: for each pair (ω, X) as above,

- whenever $\sigma(X \cup \{\omega\}) = \{\omega\}$, add the constraint $\mu(\omega) > \frac{t}{1-t} \cdot \mu(X)$,
- otherwise, add the constraint $\mu(\omega) \leq \frac{t}{1-t} \cdot \mu(X)$.

We need to ensure that the resulting system of linear inequalities L_σ, together with the constraint that μ lie in the simplex Δ^{n-1}, admits a solution.

If these desiderata are satisfied, then we have fully characterised the class of τ-generated revision maps, and therefore the class of consequence relations $\mid\!\sim_\mu$. We can argue as follows: suppose a measure μ satisfies all linear inequalities in the resulting system L_σ. Then, in particular, whenever $\sigma(A) = B$, the properties (i′) and (ii′) hold, and since μ satisfies the resulting inequalities, the corresponding properties (i) and (ii) hold as well on the probabilistic side; hence, $\tau(\mu_A) = B$. Conversely, if $\tau(\mu_A) = B$, then $B \subseteq A$, and both (i) and (ii) hold. Via our translation, this means that properties (i′) and (ii′) hold as well, which yields $\sigma(A) = B$. Thus, if we can show that L_σ always admits a solution for any σ, we are done.

Here are some axioms for selection functions which are sound with respect to this probabilistic interpretation[29]:

[29] Axiom (S1) may look surprising here, since we should allow nonempty sets X such that $\sigma(X) = \emptyset$, representing sets of measure 0. The idea is that we can exclude all such propositions from the domain of σ, restrict attention to what σ does on this restricted domain, represent it via a *regular* probability measure μ on this smaller algebra, and simply let $\mu(X) = 0$ for all such X. So, without loss of generality, we can focus

(S1) $\sigma(X) = \emptyset$ only if $X = \emptyset$;

(S2) $\sigma(X) \subseteq X$;

(S3) If $\sigma(A) \cap B \neq \emptyset$, then $\sigma(A \cap B) \subseteq \sigma(A) \cap B$;

(S4$_n$) For any n: if $\sigma(A \cup X_i) = X_i$ for all $i \leq n$, then $\sigma(A \cup \bigcup_{i \leq n} X_i) \subseteq \bigcup_{i \leq n} X_i$;

(S5) For any A, B, C which are pairwise disjoint, if $\sigma(A \cup B) \subseteq A$ and $\sigma(B \cup C) \subseteq B$, then $\sigma(A \cup C) \subseteq A$.

Note that (S2) and (S3) suffice for σ to validate (Ref) and (RM), respectively. Now, the axioms (S1)-(S4) suffice for selection functions to mirror Proposition 3.2.5 on the qualitative side:

Proposition 3.2.8
If $\sigma : \mathfrak{A} \to \mathfrak{A}$ satisfies (S1)-(S4), then $\sigma(A) = B$ if and only if $B \subseteq A$ and

(i′) $\forall b \in B$, $\sigma((A \setminus B) \cup \{b\}) = \{b\}$;

(ii′) $\forall X \subset B$, $\exists x \in X$, $\sigma((A \setminus X) \cup \{x\}) \neq \{x\}$.

Proof. Suppose $\sigma(A) = B$. Then $B \subseteq A$ because of (S2). If $A = \emptyset$, then by (S1) so is B, and then (i′) and (ii′) hold vacuously as B contains neither elements nor strict subsets. So we can assume $A, B \neq \emptyset$. We show (i′). Take $b \in B$. Now we have $\sigma(A) \cap \big((A \setminus B) \cup \{b\}\big) \neq \emptyset$, since it is equal to $B \cap \big((A \setminus B) \cup \{b\}\big) = \{b\}$. So by (S3), we must have

$$\begin{aligned} \sigma\Big(A \cap \big((A \setminus B) \cup \{b\}\big)\Big) &\subseteq \sigma(A) \cap ((A \setminus B) \cup \{b\}) \\ \sigma\Big((A \setminus B) \cup \{b\}\Big) &\subseteq B \ \cap ((A \setminus B) \cup \{b\}), \\ \text{so} \qquad \sigma\Big((A \setminus B) \cup \{b\}\Big) &\subseteq \{b\}. \end{aligned}$$

Now by (S1), $\sigma\big((A \setminus B) \cup \{b\}\big) \neq \emptyset$, so it is equal to $\{b\}$, as desired.

To show that (ii′) holds, proceed by contradiction. Suppose that $\exists X \subset B$, $\forall x \in X$, $\sigma((A \setminus X) \cup \{x\}) = x$. X is a finite set, so write $X = \{x_1, \ldots, x_k\}$. We then have $\forall i \leq k$, $\sigma((A \setminus X) \cup \{x_i\}) = x_i$. So, by (S4$_k$), we have that

$$\sigma((A \setminus X) \cup \bigcup_{i \leq k} \{x_i\}) \subseteq \bigcup_{i \leq k} \{x_i\}.$$

The left-hand side then is simply $\sigma((A \setminus X) \cup X) = \sigma(A)$(by (S2) we have $B \subseteq A$, hence $X \subset A$), while the right-hand side is X. So we have $\sigma(A) \subseteq X$. But we know $\sigma(A) = B$ and $X \subset B$, which contradicts $\sigma(A) \subseteq X$. So (ii′) holds after all.

on representing $\mid\!\sim_\mu$ for regular measures μ.

For the other direction, suppose (i′) and (ii′) hold. We show this entails $\sigma(A) = B$. First, list elements $b_1, ..., b_k$ of B and use $(\mathsf{S4}_k)$ as before to conclude $\sigma(A \cup B) \subseteq B$, and since $B \subseteq A$ we have $\sigma(A) \subseteq B$. Now assume $B \not\subseteq \sigma(A)$. Since we already know $\sigma(A) \subseteq B$, this means $\sigma(A) \subset B$. By (ii′), we have that

$$\exists a \in \sigma(A), \quad \sigma\Big((A \setminus \sigma(A)) \cup \{a\}\Big) \neq \{a\}$$

But this is impossible, since – as we have just shown above using $(\mathsf{S3})$ – in general $\sigma(A) = S$ entails $\forall s \in S$, $\sigma((A \setminus S) \cup \{s\}) = \{s\}$. So $B \subseteq \sigma(A)$ after all, and since we know $\sigma(A) \subseteq B$ we can conclude $\sigma(A) = B$. □

The question of ensuring the consistency of the system of linear inequalities L_σ that we obtain from σ is more involved. It can be seen as a special and somewhat more intricate case of a representation problem for comparative probability orders. We consider the particular case for a threshold $t = 1/2$.

3.2.6 Connection with comparative probability orders

From now on, we deal with the case of the τ-rule with threshold $t = 1/2$. To make the comparison with comparative probability more salient, we employ here a more suggestive notation. Define the relation $\succ_\sigma$ between states $\omega \in \Omega$ and sets $X \subseteq \Omega$ as follows: for each pair (ω, X) with $\omega \notin X$, let

$$\omega \succ_\sigma X \text{ if and only if } \sigma(X \cup \{\omega\}) = \omega.$$

(When the selection function σ is clear form the context, we will omit the subscript). We are given a system of inequalities as follows: for each pair (ω, X), we have that either $\omega \succ_\sigma X$ or $\neg(\omega \succ_\sigma X)$. Each expression $\omega \succ_\sigma X$ translates into the constraint $\mu(\omega) > \mu(X)$, while each expression $\neg(\omega \succ_\sigma X)$ translates into $\mu(\omega) \leq \mu(X)$. The question is: what axioms do we need to impose on $\succ_\sigma$ for it to be probabilistically representable?

The problem we have to solve is one of representation of a partial comparative probability order. In the theory of comparative probability, one usually starts with a full ordering $\preceq$ on an algebra of events $\mathcal{P}(\Omega)$, and the task consists in finding a probability measure which represents $\preceq$. We say that a measure μ represents $\preceq$ if and only if, for any $A, B \subseteq \Omega$,

$$A \preceq B \Leftrightarrow \mu(A) \leq \mu(B).$$

Under what circumstances is such an order representable? One of the most important classical results in the early theory of comparative probability is a general answer to this question.

Theorem 3.2.9 (Kraft-Pratt-Seidenberg [27], Scott [50])
Let $(\Omega, \mathcal{P}(\Omega))$ a finite set algebra, and $\preceq$ a reflexive total order on $\mathcal{P}(\Omega)$. There is a probability measure μ on $(\Omega, \mathcal{P}(\Omega))$ representing $\preceq$ if and only if the following hold for all A, B, $C \subseteq \Omega$:

(Q1) $\Omega \not\preceq \emptyset$;

(Q2) $\emptyset \preceq A$;

(Q3) *If* $(A \cup B) \cap C = \emptyset$, *then* $(A \preceq B \Leftrightarrow A \cup C \preceq B \cup C)$;

(Q4) *If* $(A_i)_{i \leqslant n}$ *and* $(B_i)_{i \leqslant n}$ *are balanced sequences and* $\forall i < n$, $A_i \preceq B_i$, *then* $A_n \succeq B_n$.

The last requirement plays a crucial role, and it is worth taking a moment here to explain its meaning. The notion of two sequences of events being *balanced* is to be understood as follows: given two such sequences $(A_i)_{i \leqslant n}$ and $(B_i)_{i \leqslant n}$, we write $(A_i)_{i \leqslant n} \geq_0 (B_i)_{i \leqslant n}$ whenever

$$\sum_{i \leqslant n} \mathbb{1}_{A_i} \geqslant \sum_{i \leqslant n} \mathbb{1}_{B_i}.$$

This means that for each $\omega \in \Omega$, ω is in at least as many A_i's as B_i's: in other words, we have $\big|\{i \leqslant n \,|\, \omega \in A_i\}\big| \geqslant \big|\{i \leqslant n \,|\, \omega \in B_i\}\big|$.

When $(A_i)_{i \leqslant n} \geq_0 (B_i)_{i \leqslant n}$ and $(A_i)_{i \leqslant n} \leq_0 (B_i)_{i \leqslant n}$, we write

$$(A_i)_{i \leqslant n} \equiv_0 (B_i)_{i \leqslant n}$$

and we say that $(A_i)_{i \leqslant n}$ and $(B_i)_{i \leqslant n}$ are *balanced* sequences. That is, $(A_i)_{i \leqslant n}$ and $(B_i)_{i \leqslant n}$ are balanced whenever for each $\omega \in \Omega$, $|\{i \leqslant n \,|\, \omega \in A_i\}| = |\{i \leqslant n \,|\, \omega \in B_i\}|$. This means that for each state ω, the number of occurrences of ω in the A_i sets is the same as the number of occurrences of ω in the B_i sets. To put it in terms of indicator functions:

Definition 3.2.10 (**Balanced sequences**)
Let $(\Omega, \mathfrak{A})$ a finite set algebra, and $(A_i)_{i \leqslant n}$ and $(B_i)_{i \leqslant s}$ two sequences of events from $\mathfrak{A}$. The sequences are balanced *if and only if*

$$\sum_{i \leqslant n} \mathbb{1}_{A_i} = \sum_{i \leqslant n} \mathbb{1}_{B_i}.$$

We then write $(A_i)_{i \leqslant n} \equiv_0 (B_i)_{i \leqslant n}$.

Scott's well-known proof of Theorem 3.2.9 [50] appeals to a hyperplane-separation theorem. Scott identifies each set X with its characteristic vector $\mathbb{1}_X$ and shows, via a crucial application of Scott's axiom (Q4), that the sets $\{\mathbb{1}_A - \mathbb{1}_B \,|\, A \succ B\}$ and $\{\mathbb{1}_A - \mathbb{1}_B \,|\, A \preceq$

B and $B \preceq A\}$ can be separated by a hyperplane with equation $\overrightarrow{v} \cdot \mathbf{x} = 0$. Then, letting

$$\mu(X) := \frac{\overrightarrow{v} \cdot \mathbb{1}_X}{\overrightarrow{v} \cdot \mathbb{1}_\Omega}$$

gives the desired probability representation.

Our case, however, is a little more intricate. We have an ordering $\succ_\sigma$ that is of a very specific kind: it is non-total, and its domain is restricted: it relates singletons (more generally, atoms in the event algebra) on one side to sets on the other. Furthermore, we need to deal with strict and non-strict constraints simultaneously – both of which are given as primitives – rather than deriving one set of constraints from the other (as it is usually done for total probability orders).

We can begin by observing some plausible candidates for representability conditions. The following properties of the relation $\succ$ are sound with respect to our desired probabilistic interpretation.

(No-swap) If $\omega \succ X$ and $Y \subseteq X$, then $\omega \succ Y$.

(Sub) If $\omega \succ X$ and $v \in X$, $v \succ Y$ then $\omega \succ (X \setminus \{v\}) \cup Y$.

(Sum$^-$) If $\omega_1 \not\succ B_1$ and $\omega_2 \not\succ B_2$ for $B_1 \cap B_2 = \emptyset$, then $\forall\omega(\omega \succ B_1 \cup B_2 \rightarrow \omega \succ \{\omega_1, \omega_2\})$.

(Sep) If $\exists X, \exists\omega$ such that $\omega \not\succ X \cup \{v_1\}$ and $\omega \succ X \cup \{v_2\}$, then $v_1 \succ \{v_2\}$.

(Scott$'$) If $(A_i)_{i\leq m}$ and $(B_i)_{i\leq m}$ are balanced sequences and $(\omega_i)_{i\leq m}$ a sequence of states, then $(\forall i \leq m,\, \omega_i \succ A_i) \rightarrow (\exists i \leq m,\, \omega_i \succ B_i)$.

In what follows, we apply the hyperplane separation techniques form the theory of comparative probability orders, and rely a method analogous to Scott's proof from [50]. Given that we are dealing with a mixed-constraints case, the relevant hyperplane-separation result to be employed here is the Motzkin Transposition Theorem[30] [42]. Here, one should also expect a Scott-like axiom to do most of the work: as it turns out, this is indeed the case, but what is needed is an axiom stronger than the property (Scott$'$) listed above. In what follows, we employ this strategy to prove a probabilistic representation theorem for selection structures.

3.3 Representation

In this section, we solve the representation problem for selection structures: that is, we give necessary and sufficient conditions for a selection function to be representable by a (regular)

[30] See, for instance, [49, p. 33] or [55].

probability measure. We begin by proving a general result which gives sufficient conditions for two partial orders on a finite algebra (one strict and one non-strict) to be simultaneously weakly representable by a probability measure (Proposition 3.3.2). We briefly discuss the relation between our conditions and those given by Fishburn [17] for the weak representation of strict comparative probability orders. We then use this general representation result to prove the representation theorem for selection structures (Theorem 3.3.8).

3.3.1 Weak representation of comparative probability orders

We begin by proving Proposition 3.3.2, which gives sufficient conditions for the joint weak representability of two relations on a finite algebra (one strict, one non-strict). We shall appeal to the following result:

Theorem 3.3.1 (**Motzkin Transposition Theorem**, Motzkin [42])
Let $\mathbf{M}_1$ *be a matrix in* $\mathbb{Q}^{k\times n}$*, and* $\mathbf{M}_2$ *a matrix in* $\mathbb{Q}^{p\times n}$*. Then, either there is a vector* $\vec{\mu}\in\mathbb{R}^n$ *such that*

$$\mathbf{M}_1\cdot\vec{\mu}\geqslant\vec{0}$$
$$\mathbf{M}_2\cdot\vec{\mu}>\vec{0}$$

or else there exist vectors $\alpha\in\mathbb{Q}^k$*,* $\beta\in\mathbb{Q}^p$ *such that*

(a) $\alpha^T\mathbf{M}_1+\beta^T\mathbf{M}_2=\vec{0}$*;*
(b) $\alpha\geqslant\vec{0}$*,* $\beta\geqslant\vec{0}$ *and* $\beta_i>0$ *for some coordinate* $i\leqslant p$*.*

We now prove our proposition:

Proposition 3.3.2 (**Weak representation**)
Let $(\Omega,\mathfrak{A})$ *a finite algebra, and let* $<$ *and* $\leqslant$ *be two partial relations on* $\mathcal{P}(\Omega)$ *such that for any* A*,* $B\subseteq\Omega$*:*

(A0) $\forall\omega\in\Omega$*,* $\{\omega\}>\emptyset$*;*
(A1) $A\leqslant B\Rightarrow A\not>B$*;*
(A2) $A<B\Rightarrow A\leqslant B$*;*
(Scott) *If* $(A_i)_{i\leqslant n}\equiv_0(B_i)_{i\leqslant n}$ *and* $\forall i\leqslant n$*,* $A_i\geqslant B_i$*, then* $\forall i\leqslant n$*,* $A_i\leqslant B_i$*.*

Then, there is a regular probability measure μ *on* $\mathfrak{A}$ *such that, for all* A*,* $B\subseteq\Omega$*,*

- $A<B\Rightarrow\mu(A)<\mu(B)$*;*
- $A\leqslant B\Rightarrow\mu(A)\leqslant\mu(B)$*.*

Proof. Let $\Omega=\{\omega_1,\ldots,\omega_n\}$. Consider the following two sets of vectors representing inequalities:

$$\Gamma := \{\mathbb{1}_A - \mathbb{1}_B \mid A \geqslant B\}$$
$$\Sigma := \{\mathbb{1}_A - \mathbb{1}_B \mid A > B\},$$

and we write $|\Gamma| = k$ and $|\Sigma| = p$. Take the following two matrices: $\mathbf{M}_\Gamma$ has as rows (transposes of) vectors in Γ, while the rows of matrix $\mathbf{M}_\Sigma$ are (transposes of) vectors in Σ. We write them as

$$\mathbf{M}_\Gamma = \begin{pmatrix} (\mathbb{1}_{A_1} - \mathbb{1}_{B_1})^T \\ \vdots \\ (\mathbb{1}_{A_k} - \mathbb{1}_{B_k})^T \end{pmatrix} \quad \text{and} \quad \mathbf{M}_\Sigma = \begin{pmatrix} (\mathbb{1}_{A_{k+1}} - \mathbb{1}_{B_{k+1}})^T \\ \vdots \\ (\mathbb{1}_{A_{k+p}} - \mathbb{1}_{B_{k+p}})^T \end{pmatrix}$$

We now prove that there is a vector $\vec{\mu} \in \mathbb{R}^n$ such that

$$\begin{aligned} \mathbf{M}_\Gamma \cdot \vec{\mu} &\geqslant \vec{0} \\ \mathbf{M}_\Sigma \cdot \vec{\mu} &> \vec{0} \end{aligned} \tag{3.2}$$

Assume towards a contradiction that there is no such $\vec{\mu}$. The matrices $\mathbf{M}_\Gamma$ and $\mathbf{M}_\Sigma$ are rational valued, since they only contain entries in $\{-1, 0, 1\}$. Then by Motzkin's Transposition Theorem, there exists vectors α, β with non-negative rational entries such that

$$\alpha^T \mathbf{M}_\Gamma + \beta^T \mathbf{M}_\Sigma = \vec{0} \tag{3.3}$$

and β has at least one positive coordinate. Now, we can in fact assume those are vectors in $\mathbb{N}$, by multiplying the entries by a common denominator. We can then rewrite the above in full as

$$(\alpha_1, \ldots, \alpha_k) \begin{pmatrix} (\mathbb{1}_{A_1} - \mathbb{1}_{B_1})^T \\ \vdots \\ (\mathbb{1}_{A_k} - \mathbb{1}_{B_k})^T \end{pmatrix} + (\beta_1, \ldots, \beta_p) \begin{pmatrix} (\mathbb{1}_{A_{k+1}} - \mathbb{1}_{B_{k+1}})^T \\ \vdots \\ (\mathbb{1}_{A_{k+p}} - \mathbb{1}_{B_{k+p}})^T \end{pmatrix} = \vec{0}$$

A piece of notation: given an integer m, let us write mA to denote the sequence consisting of the set A repeated m times. Given this, the sequence of sets

$$(\alpha_1 A_1, \ldots, \alpha_k A_k, \beta_1 A_{k+1}, \ldots \beta_p A_{k+p})$$

contains exactly α_1 copies of the set A_1, followed by α_2 copies of the set A_2, etc. (respectively, β_j copies of A_{k+j}). Now we claim that

$$(\alpha_1 A_1, \ldots, \alpha_k A_k, \beta_1 A_{k+1}, \ldots \beta_p A_{k+p}) \equiv_0 (\alpha_1 B_1, \ldots, \alpha_k B_k, \beta_1 B_{k+1}, \ldots \beta_p B_{k+p}). \tag{3.4}$$

The two sequences in (3.4) both have length $l = \sum_i \alpha_i + \sum_j \beta_j$. So let us write the two sequences in (3.4) as $(A_i^*)_{i \leqslant l}$ and $(B_i^*)_{i \leqslant l}$.

The fact that those two sequences are balanced follows from the equality (3.3). Note that

$$\begin{aligned}\alpha^T \mathbf{M}_\Gamma + \beta^T \mathbf{M}_\Sigma &= \sum_{i \leqslant k} \alpha_i (\mathbb{1}_{A_i} - \mathbb{1}_{B_i})^T + \sum_{j=k+1}^{p} \beta_j (\mathbb{1}_{A_j} - \mathbb{1}_{B_j})^T \\ &= \big(\sum_{i \leqslant k} \alpha_i \mathbb{1}_{A_i}^T + \sum_{j=k+1}^{p} \beta_j \mathbb{1}_{A_j}^T\big) - \big(\sum_{i \leqslant k} \alpha_j \mathbb{1}_{B_j}^T + \sum_{j=k+1}^{p} \beta_j \mathbb{1}_{B_j}^T\big) = \vec{0}\end{aligned}$$

In this last equality, the vector $\big(\sum_{i \leqslant k} \alpha_i \mathbb{1}_{A_i}^T + \sum_{j=k+1}^{p} \beta_j \mathbb{1}_{A_j}^T\big)$ is a $(1 \times n)$ vector $\mathbf{v} = (v_1, \ldots, v_n)$ where $v_j = |\{i \leqslant l \,|\, \omega_j \in A_i^*\}|$: that is, the j-th entry of $\mathbf{v}$ counts the number of A_i^*'s in which the state ω_j occurs. By this reasoning, the last line above states that, for any $\omega \in \Omega$, we have $|\{i \leqslant l \,|\, \omega \in A_i^*\}| = |\{i \leqslant l \,|\, \omega \in B_i^*\}|$. So the sequences in (3.4) are indeed balanced.

Now not only are the two sequences in (3.4) balanced, but by design of the matrices, we also have that for any A_i in the sequence, we have $A_i \geqslant B_i$: this is obvious for all pairs A_i, B_i with $i \leqslant k$ – i.e., the pairs taken from Γ (that is, such that $\mathbb{1}_{A_i} - \mathbb{1}_{B_i} \in \Gamma$). This also holds for pairs A_j, B_j for $j \geqslant k$ (pairs from Σ): this is because we know that $A_i > B_i$ by choice of Σ, so by axiom (A2) we also have $A_i \geqslant B_i$. So we have:

For any A_i occuring in the left-hand side sequence in (3.4), the corresponding B_i on the right-hand side is such that $A_i \geqslant B_i$.

Now we can appeal to (Scott) to conclude that we also must have $A_j \leqslant B_j$ for all coordinates j. But remember that $\mathbf{M}_\Sigma$ was non-empty, because we know from axiom (A0) that the relation $>$ admits strict inequalities of the form $\omega > B$. Furthermore, at least one such strict inequality must occur on the right-hand side in (3.4) since we know that β admits at least one strictly positive coordinate (by (b) of Motzkin's Transposition Theorem). This means that there is a pair A, B such $A > B$, the set A occurs somewhere on the left-hand side of (3.4), and B occurs at the same coordinate on the right-hand-side. In particular then, the application of (Scott) above entails that we must have $A \leqslant B$. But, by axiom (A1), $A \leqslant B$ entails $A \not> B$, contradicting the fact that $A > B$.

Thus, the assumption that such vectors α and β exist leads to contradiction. By Motzkin's theorem, we can conclude that there is some vector $\vec{\mu} \in \mathbb{R}^n$ solving $\mathbf{M}_\Gamma \cdot \vec{\mu} \geqslant \vec{0}$ and $\mathbf{M}_\Sigma \cdot \vec{\mu} > \vec{0}$. Now, because of axiom (A0), we know that the vector $\vec{\mu}$ satisfies

$$\vec{\mu} \cdot (\mathbb{1}_\omega - \mathbb{1}_\emptyset) > 0$$

for any $\omega \in \Omega$; but this simply means $\vec{\mu} \cdot \mathbf{e}_i = \vec{\mu}_i > 0$ for all i. So all coordinates of $\vec{\mu}$ are strictly positive. We can then define the following function μ^* on $\mathfrak{A}$:

$$\text{for any } X \subseteq \Omega, \text{ let } \mu^*(X) := \frac{\vec{\mu} \cdot \mathbb{1}_X}{\vec{\mu} \cdot \mathbb{1}_\Omega}.$$

Then μ^* is the desired probability distribution. It is function $\mu^* : \mathfrak{A} \to [0,1]$ since we have re-normalised by dividing by $||\vec{\mu}||$. Additivity follows from the definition: for disjoint sets A, B we have

$$\mu^*(A \cup B) = \frac{\vec{\mu} \cdot \mathbb{1}_{A \cup B}}{\vec{\mu} \cdot \mathbb{1}_\Omega} = \frac{\vec{\mu} \cdot (\mathbb{1}_A + \mathbb{1}_B)}{\vec{\mu} \cdot \mathbb{1}_\Omega} = \frac{\vec{\mu} \cdot \mathbb{1}_A + \vec{\mu} \cdot \mathbb{1}_B}{\vec{\mu} \cdot \mathbb{1}_\Omega} = \mu^*(A) + \mu^*(B).$$

Lastly, we have already checked that it is a *regular* distribution using axiom (A0). That μ^* respects both the strict and non-strict inequalities imposed by the orderings $>$ and $\geqslant$ follows immediately from (3.2). □

This result will be the key step in our representation theorem for selection functions.

Before we move on to the representation problem for selection structures, it is instructive to compare Proposition 3.3.2 with a theorem by Fishburn from [17]. There, Fishburn gives necessary and sufficient conditions for the weak representation of a strict qualitative probability order on a finite algebra.

Theorem 3.3.3 (Fishburn [17])
Let Ω a finite set and $\mathfrak{A}$ an algebra over it, and $\prec$ a binary relation on $\mathfrak{A}$. Then there is a measure μ on $(\Omega, \mathfrak{A})$ which weakly represents *the relation $\prec$, meaning*

$$\forall A, B \in \mathfrak{A},\ A \prec B \Rightarrow \mu(A) < \mu(B)$$

if and only if the following holds:

(F) *If $(A_i)_{i \leq n} \geq_0 (B_i)_{i \leq n}$ (with $n > 0$) and $A_i \prec B_i$ for all $i < n$, then $A_n \not\prec B_n$.*

The conditions given in Proposition 3.3.2 ensure the weak representation of the $<$ relation on $\mathfrak{A}$. In particular then, conditions (A0)-(A2) and Scott must entail (F). We can verify 'by hand' this is indeed true:

Proposition 3.3.4
Let $\mathfrak{A}$ an algebra over a finite set Ω, and $<$ a binary relation on $\mathfrak{A}$ which satisfies (A0)-(A2) *and* (Scott). *Then $<$ satisfies* (F).

Proof. Suppose towards a contradiction that (F) fails: that is, we have some A_i, B_i such that $(A_i)_{i \leq n} \geq_0 (B_i)_{i \leq n}$ and $A_i < B_i$ for *all* $i \leqslant n$. We derive a contradiction. For any $\omega \in \Omega$, let

$$\eta_A(\omega) := \big|\{i \leqslant n \,|\, \omega \in A_i\}\big|$$
$$\eta_B(\omega) := \big|\{i \leqslant n \,|\, \omega \in B_i\}\big|$$

Since $(A_i)_{i\leq n} \geq_0 (B_i)_{i\leq n}$, this means that for any $\omega \in \Omega$, we have

$$\eta_A(\omega) \geqslant \eta_B(\omega).$$

We construct a sequence $(a_1, ..., a_k)$ of states in Ω, as follows: for each $\omega \in \Omega$, whenever $\eta_A(\omega) = \eta_B(\omega) + p$ for some integer $p > 0$, add p copies of ω to the sequence (in whatever order). Then the sequence $(a_1, ..., a_k)$ contains exactly $\eta_A(\omega) - \eta_B(\omega)$ copies of each state $\omega \in \Omega$.

Writing $(A_i^*)_{i\leq n+k} := (A_1, \ldots, A_n, \emptyset, \ldots, \emptyset)$ and $(B_i^*)_{i\leq n+k} := (B_1, \ldots, B_n, \{a_1\}, \ldots, \{a_k\})$, we immediately have that the two sequences are balanced: we have

$$\eta_{B^*}(\omega) = \eta_B(\omega) + \big(\eta_A(\omega) - \eta_B(\omega)\big) = \eta_A(\omega) = \eta_{A^*}(\omega),$$

so that each 'extra' occurrence of a state in the sets A_i^* is matched by an occurrence in one of sets B_i^*. Thus we know:

$$\underset{i\leq n+k}{(A_i^*)} \equiv_0 \underset{i\leq n+k}{(B_i^*)}.$$

From (A0) we know that for each a_i in $(a_1, ..., a_k)$, we have $\{a_i\} > \emptyset$. We then have the following:

$$\begin{array}{ccccccc} \big(A_1, & \ldots, & A_n, & \emptyset, & \ldots, & \emptyset\big) \\ \wedge & \ldots & \wedge & \wedge & \ldots & \wedge \\ \big(B_1, & \ldots, & B_n, & \{a_1\}, & \ldots, & \{a_k\}\big) \end{array}$$

We can use (A2) and write:

$$\underset{i\leq n+k}{(A_i^*)} \equiv_0 \underset{i\leq n+k}{(B_i^*)} \text{ and } \forall i \leq n+k,\ A_i^* \leqslant B_i^*.$$

Now, by (Scott) we have that

$$\forall i \leq n+k,\ A_i^* \geqslant B_i^*.$$

But now, since we have $A_1 = A_1^*$ and $B_1 = B_1^*$, we obtain both $A_1 < B_1$ and $A_1 \geqslant B_1$, contradicting (A1). □

This verifies that we can easily derive Fishburn's axiom (F) directly from our conditions (A0)-(A2) and Scott.

Proposition 3.3.2 extends Fishburn's result by giving sufficient conditions for the simultaneous weak representation of two relations, one strict and one non-strict. We can now move on to giving the solution to our representation problem.

3.3.2 Representation theorem for selection structures

We begin by defining an order relation induced by selection functions.

Definition 3.3.5 (**The $\succcurlyeq^*_\sigma$ order**)
Given a selection structure $(\Omega, \mathfrak{A}, \sigma)$ and $X \cup \{\omega\} \subseteq \Omega$, write $\omega \succ_\sigma X$ whenever $\sigma(X \cup \{\omega\}) = \{\omega\} \cap X^c$, and $X \succeq_\sigma \omega$ whenever $\sigma(X \cup \{\omega\}) \neq \{\omega\} \cap X^c$. Extend this to a relation $\succcurlyeq^*_\sigma$ given by the condition:

$$A \succcurlyeq^*_\sigma B \Leftrightarrow A \succ_\sigma B \text{ or } A \succeq_\sigma B.$$

A few words about this order. First note that we have

$$\succcurlyeq^*_\sigma = \big\{(\{\omega\}, X) \,\big|\, \omega \succ_\sigma X\big\} \cup \big\{(X, \{\omega\}) \,\big|\, \omega \not\succ_\sigma X\big\}$$

with $\omega \in \Omega$, $X \subseteq \Omega$. This means that $A \succcurlyeq^*_\sigma B$ is defined exactly when at least one of A, B is a singleton (atom) in the algebra. It is undefined otherwise. Intuitively, $A \succcurlyeq^*_\sigma B$ means that either A is a singleton 'dominating' the set B[31] or that B is a singleton which fails to dominate A. An alternative way to define this ordering is by means of the following equivalences:

$$\omega \succ_\sigma X \Longleftrightarrow \omega \notin X \text{ and } \sigma(X \cup \{\omega\}) = \{\omega\},$$
$$X \succeq_\sigma \omega \Longleftrightarrow \sigma(X \cup \{\omega\}) \neq \{\omega\} \text{ or } \omega \in X$$

(The notation for all special order relations can be found summarised in the Appendix.) The following is immediate:

Proposition 3.3.6
Let σ be a selection function. Then the following holds:

(A1) $A \preccurlyeq^*_\sigma B \Rightarrow A \not\succ_\sigma B$.
(A2) $A \prec_\sigma B \Rightarrow A \preccurlyeq^*_\sigma B$.

The Scott axiom plays a key role in the representation theorem above. In the context of selection rules, it is indeed a very powerful property: firstly, we check that it is sound with respect to our probabilistic interpretation.

Proposition 3.3.7
The (Scott) *axiom for $\succcurlyeq^*_\sigma$ is probabilistically sound. That is, given any finite probability space $(\Omega, \mathcal{P}(\Omega), \mu)$, define $\sigma : \mathcal{P}(\Omega) \to \mathcal{P}(\Omega)$ by $\sigma(A) := \tau(\mu_A)$ for threshold $t = 1/2$. Then the* (Scott) *property holds for σ.*

[31] In the sense that $\mu(A) > \mu(B)$ for any measure that represents σ.

Proof. Let $\succcurlyeq^*_\sigma$ an order obtained from the selection function σ as above, so that μ weakly represents both $\succ_\sigma$ and $\succcurlyeq^*_\sigma$. We assume the measure μ is regular (recall we are only interested in regular measures, as we can restrict attention to $\mathrm{supp}(\mu)$ without loss of generality). Assume that $(A_i)_{i\leqslant n} \equiv_0 (B_i)_{i\leqslant n}$ and $\forall i \leqslant n$, $A_i \succcurlyeq^*_\sigma B_i$. We show that $\forall i \leqslant n$, $A_i \preccurlyeq^*_\sigma B_i$[32].

First, note that we cannot have any strict relation $A_i \succ_\sigma B_i$, as it would immediately entail

$$\sum_{i=1}^{n} \mu(A_i) > \sum_{i=1}^{n} \mu(B_i).$$

But this would contradict the fact that $(A_i)_{i\leqslant n} \equiv_0 (B_i)_{i\leqslant n}$, since the fact that these sequences are balanced entails

$$\sum_{i=1}^{n} \sum_{\omega\in A_i} \mu(\omega) = \sum_{i=1}^{n} \mu(A_i) = \sum_{i=1}^{n} \mu(B_i) = \sum_{i=1}^{n} \sum_{\omega\in B_i} \mu(\omega).$$

By definition of $\succcurlyeq^*_\sigma$ above, we can have $A_i \not\succ_\sigma B_i$ and $A_i \succcurlyeq^*_\sigma B_i$ only if B_i is a singleton $B_i = \{b_i\}$ such that $\sigma(A_i \cup \{b_i\}) \neq \{b_i\}$. So all B_i's must be singletons.

Next we note that all A_i's must be singletons as well. Firstly, all A_i's must be nonempty, by regularity of μ: for otherwise $\emptyset = A_i \succcurlyeq^*_\sigma \{b_i\}$ means that $\sigma(\emptyset \cup \{b_i\}) \neq \{b_i\}$, while regularity enforces $\sigma(\{\omega\}) = \{\omega\}$ for any state. Now suppose that some A_i contains different states $\{a_1, ..., a_k\}$. We have $\{a_1, ..., a_k\} \succcurlyeq^*_\sigma b_i$ and since the sequences are balanced, each of those a_i's must appear as a singleton $\{b_i\}$ in the sequence $(B_i)_{i\leqslant n}$. Now a contradiction follows by a counting argument: count all *occurrences* of elements in the sets A_j and do the same the sets B_j. Since all B_j's are singletons we have

$$\sum_{\omega\in\Omega} |\{j \leqslant n \,|\, \omega \in B_j\}| = \sum_{j\leqslant n} |B_j| = n.$$

Each occurrence of a state ω in any of the A_j must be matched by an occurrence in one of the B_j. But there is at least one such occurrence for each A_j, since they are nonempty, and strictly more than one occurrence for $A_i = \{a_1, ..., a_k\}$, so $\sum_{\omega\in\Omega} |\{j \leqslant n \,|\, \omega \in A_j\}| > n$. This entails

$$\sum_{i\leqslant n} \mathbb{1}_{A_i} > \sum_{i\leqslant n} \mathbb{1}_{B_i},$$

contradicting the fact that $(A_i)_{i\leqslant n}$ and $(B_i)_{i\leqslant n}$ are balanced. So all A_i's must indeed be

[32]Here the only subtlety is the following: given the premises, it is immediate that we must have $\mu(A_i) \leqslant \mu(B_i)$ for all $i \leqslant n$. This in itself does not entail $A_i \preccurlyeq^*_\sigma B_i$, however: we only have that $A \succcurlyeq^*_\sigma B$ entails $\mu(A) \geqslant \mu(B)$, but the converse direction may not hold. For instance, it could be that $\mu(A_n) = \mu(B_n)$ and B_n is an atom in the algebra while A_n is not, in which case we cannot have $A_n \preccurlyeq^*_\sigma B_n$, as we can have neither $A_n \preceq_\sigma B_n$ (due to the domain restrictions of $\preceq$, which only allow this to hold when A_n is a singleton) nor $A_n \prec_\sigma B_n$ (as this contradicts $\mu(A_n) = \mu(B_n)$). Thus we must make sure that his never occurs in balanced sequences.

singletons as well.

So the sequences $(A_i)_{i\leqslant n}$ and $(B_i)_{i\leqslant n}$ are really sequences of singletons $(a_i)_{i\leqslant n}$ and $(b_i)_{i\leqslant n}$ such that $\mu(a_i) \geqslant \mu(b_i)$ for all $i \geqslant n$, and since they are balanced it follows that $\mu(a_i) \leqslant \mu(b_i)$. Now note that, given the definition of $\succcurlyeq^*_\sigma$, whenever a_i and b_i are singletons then $\mu(a_i) \leqslant \mu(b_i)$ entails $a_i \preccurlyeq^*_\sigma b_i$: for if $a_i \not\preccurlyeq^*_\sigma b_i$, then $\sigma(\{a_i, b_i\}) = a_i \setminus \{b_i\} \neq \emptyset$, which means $a_i \succ_\sigma b_i$ and, by weak representation, $\mu(a_i) > \mu(b_i)$. So, we can conclude that $A_i \preccurlyeq^*_\sigma B_i$ holds for all $i \leqslant n$, as desired. □

One may also verify, although we omit the argument here, that the (Scott) axiom imposed on $\succcurlyeq^*_\sigma$, together with properties (S1) and (S2) as introduced in Proposition 3.2.8, entails *all* of the desired properties for selection functions introduced in the previous section (page 66)[33].

We can now prove the representation theorem:

Theorem 3.3.8

Representation theorem for selection structures

Let $(\Omega, \mathcal{P}(\Omega), \sigma)$ be a selection structure satisfying the following:

(S1) $\sigma(X) = \emptyset$ *only if* $X = \emptyset$

(S2) $\sigma(X) \subseteq X$

(S3) *If* $\sigma(A) \cap B \neq \emptyset$, *then* $\sigma(A \cap B) \subseteq \sigma(A) \cap B$

($\mathsf{S4}_n$) *For any* n: *if* $\sigma(A \cup X_i) = X_i$ *for all* $i \leq n$, *then* $\sigma(A \cup \bigcup_{i\leq n} X_i) \subseteq \bigcup_{i\leq n} X_i$

(Scott) *If* $(A_i)_{i\leqslant n} \equiv_0 (B_i)_{i\leqslant n}$ *and* $\forall i \leqslant n$, $A_i \succcurlyeq^*_\sigma B_i$, *then* $\forall i \leqslant n$, $A_i \preccurlyeq^*_\sigma B_i$.

Then there is a (regular) probability measure representing σ. Conversely, for any probability space $(\Omega, \mathcal{P}(\Omega), \mu)$ with μ a regular measure, the strongest stable set operator $\sigma_\mu : X \mapsto \tau(\mu_X)$ satisfies axioms (S1) – ($\mathsf{S4}_n$) *and* (Scott).

Proof. The second part – the probabilistic soundness of the axioms – is straightforward (and has, for the most part, been verified in this chapter; the only remaining scheme to check, ($\mathsf{S4}_n$), is indirectly shown to be sound in section 3.3.3). We show that the axioms suffice for probabilistic representability. Let $(\Omega, \mathcal{P}(\Omega), \sigma)$ a selection structure as above. Observe that the relations $\succ_\sigma$ and $\succcurlyeq^*_\sigma$ satisfy all of the conditions in Proposition 3.3.2: the (Scott) property is given. Next, (S1) ensures that $\sigma(\{\omega\}) \neq \emptyset$ for all $\omega \in \Omega$, so that (S2) ensures $\sigma(\{\omega\}) = \{\omega\}$, so the following holds:

$$\textbf{(A0)} \quad \forall \omega \in \Omega,\ \omega \succ_\sigma \emptyset.$$

[33] That is, if σ satisfies (S1) and (S2), and the induced order $\succcurlyeq^*_\sigma$ satisfies the (Scott) axiom, then $\succ_\sigma$ satisfies *(No-swap)*, *(Sub)*, *(Sum$^-$)*, *(Sep)* and *(Scott$^-$)*: further, the property (S5), introduced just before Proposition 3.2.8, also follows.

Further, by Proposition 3.3.6, we have

$$\textbf{(A1)} \quad A \preccurlyeq^*_\sigma B \Rightarrow A \not\succ_\sigma B$$

$$\textbf{(A2)} \quad A \prec_\sigma B \Rightarrow A \preccurlyeq^*_\sigma B$$

Given this, Proposition 3.3.2 entails that there exists a regular probability measure μ such that for any A, $B \subseteq \Omega$:

- $A \prec_\sigma B \Rightarrow \mu(A) < \mu(B)$
- $A \preccurlyeq^*_\sigma B \Rightarrow \mu(A) \leqslant \mu(B)$

By definition of $\prec_\sigma$ and $\preceq_\sigma$, this entails that we have, for any $\omega \in \Omega$ and $X \subseteq \Omega$ with $\omega \notin X$:

- If $\sigma(X \cup \{\omega\}) = \{\omega\}$ (equivalently, $X \prec_\sigma \omega$) then $\mu(X) < \mu(\omega)$
- If $\sigma(X \cup \{\omega\}) \neq \{\omega\}$, then $\omega \preccurlyeq^*_\sigma X$ and so $\mu(\omega) \leqslant \mu(X)$

This means that the measure μ agrees with σ on all pairs $(\omega, X) \in \Omega \times \mathfrak{A}$, and so solves the system consisting of all σ-generated inequalities in L_σ. By Observation 3.2.7, the system L_σ uniquely identifies a consequence relation $\mathrel{|\!\sim_\mu}$. The selection function σ satisfies all of $(\mathsf{S1}) - (\mathsf{S4})$ so, by Proposition 3.2.8 (and the discussion immediately preceding it), we have that $\sigma(A) = B$ if and only if $\tau(\mu_A) = B$, and thus μ represents the selection function σ. $\square$

This gives the solution to our representation problem: the selection function σ is a strongest-stable-set operator (generated by some probability measure) if and only if it satisfies the properties $(\mathsf{S1})$, $(\mathsf{S2})$, $(\mathsf{S3})$, $(\mathsf{S4}_n)$ and (Scott). Thus Theorem 3.3.8 gives a full qualitative description of strongest-stable-set operators on finite probability spaces.

This concludes our discussion of the representation of the strongest-stable-set operator by means of selection functions. Before we move on, let us conclude with a few remarks about the axioms.

Scott axioms and Fishburn axioms for comparative probability. Firstly, note that none of the results in this section section relied on the fact that we worked with full powerset algebras on Ω: this simply made our notation more convenient, as we could refer to *singleton sets* $\{\omega\}$ for $\omega \in \Omega$, instead of talking about *atoms in the underlying algebra*. All of the the above results – and Theorem 3.3.8 in particular – hold for any pair $(\Omega, \mathfrak{A})$ where $\mathfrak{A}$ is a subalgebra of $\mathcal{P}(\Omega)$, as long we work in finite spaces and replace any mention of 'singletons in $\mathcal{P}(\Omega)$' with 'atoms in $\mathfrak{A}$'.

Secondly, in order to get a better grasp on the connection between selection functions and the theory of comparative probability orders, it will be useful to think about what

selection functions can tell us about the underlying probability comparisons. Consider a representable selection structure $(\Omega, \mathfrak{A}, \sigma)$ (that is, a selection structure satisfying the necessary and sufficient conditions from the representation theorem given above). Given a representable selection function σ, which probability inequalities of the from $\mu(A) > \mu(B)$ must hold for any measure μ representing σ? In what ways can we express the fact that $\mu(A) > \mu(B)$, using only the selection function σ? The following is immediate:

Observation 3.3.9
For any representable selection function σ and measure μ representing σ, we have that $\sigma(A \cup B) \subseteq A \setminus B$ entails $\mu(A) > \mu(B)$.

Proof. For any $\omega \in \sigma(A \cup B)$, by Proposition 3.2.5 we have $\mu(\omega) > \mu\big((A \cup B) \setminus \sigma(A \cup B)\big)$. Since $\sigma(A \cup B) \subseteq A \cap B^c$, we have $B \subseteq (A \cup B) \setminus \sigma(A \cup B)$ which entails $\mu(\omega) > \mu(B)$ for any $\omega \in \sigma(A \cup B)$. So $\mu\big(\sigma(A \cup B)\big) > \mu(B)$, and we have $\sigma(A \cup B) \subseteq A$, so $\mu(A) > \mu(B)$. □

Of course, we know that representable selection functions always satisfy $\sigma(A) \subseteq A$, and so the condition $\sigma(A \cup B) \subseteq A \cap B^c$ can be rewritten as $\sigma(A \cup B) \subseteq B^c$.

Now, in order to understand the relation between selection structures and their underlying comparative probability orders, we would like to express the fact that $\mu(A) > \mu(B)$ using only the language of selection functions[34]. The above gives us one sufficient condition; but we can say more. Consider the following case.

Definition 3.3.10 (**Separation order**)
Let $(\Omega, \mathfrak{A}, \sigma)$ a selection structure with $A, B, D \in \mathfrak{A}$. We say that D separates *A from B (written $A \rhd_D B$) whenever the following conditions hold:*

- $D \cap (A \cup B) = \emptyset$
- $\sigma(D \cup B) \subseteq B^c$
- $\sigma(D \cup A) \not\subseteq A^c$

In other words, D separates A from B whenever D is disjoint form both A and B, and in addition D dominates B, but does not dominate A. It is immediate then that we have:

Proposition 3.3.11
Let $(\Omega, \mathfrak{A}, \sigma)$ a representable selection structure and μ a probability measure representing σ. Then for any $A, B, D \in \mathfrak{A}$, if $A \rhd_D B$ then $\mu(A) > \mu(B)$.

We can combine the two observations above and define the following relation.

Definition 3.3.12 (**Dominance relation**)
Let $(\Omega, \mathfrak{A}, \sigma)$ a selection structure. We define the relation $\rhd \subseteq \mathfrak{A} \times \mathfrak{A}$ as follows:

[34] This is meant informally; we have not yet introduced a formal language. We will do so in the next section, at which point the present observations will become useful.

$$A \triangleright B \text{ if and only if } \sigma(A \cup B) \subseteq B^c \text{ or } A \triangleright_D B \text{ for some } D \in \mathfrak{A}.$$

This gives us another way to express the fact that $\mu(A) > \mu(B)$ holds for *any* representing probability function: namely $A \triangleright B$ entails $\mu(A) > \mu(B)$ for any μ representing the selection function σ. Thus the extended ordering $\triangleright$ indeed entails that one event must have higher probability than another.

In Proposition 3.3.4, we showed that the Scott axioms entailed the Fishburn axiom (F) for comparative probability orders. Here, since the Fishburn axiom (F) from Theorem 3.3.3 is necessary for any measure weakly representing the comparative ordering on propositions forced by σ, the resulting order relation must satisfy the corresponding form of the Fishburn axiom. In particular then, a generalised version of the Fishburn axiom must also hold for the order $\triangleright$.

Definition 3.3.13 (**Generalised Fishburn Axiom**)
A selection structure $(\Omega, \mathfrak{A}, \sigma)$ *satisfies the* Generalised Fishburn Axiom *if and only if, for any* $A, B \in \mathfrak{A}$*, we have:*

$$\text{Whenever } (A_i)_{i \leq n} \leq_0 (B_i)_{i \leq n} \text{ and } A_i \triangleright B_i \text{ for all } i \leqslant n-1, \text{ then } \neg(A_n \triangleright B_n).$$

And indeed, we can verify directly that the required properties for representation (namely (S1), (S3), ($S4_n$) and (Scott)) entail the Generalised Fishburn Axiom.

Proposition 3.3.14
Any representable selection structure $(\Omega, \mathfrak{A}, \sigma)$ *satisfies the Generalised Fishburn Axiom.*

Proof. Let $(\Omega, \mathfrak{A}, \sigma)$ a structure satisfying (S1), (S2), (S3), ($S4_n$) and (Scott). In what follows we shall continue to treat the algebra $\mathfrak{A}$ as a full powerset algebra, so that we can conveniently refer to $\mathfrak{A}$-atoms as singleton sets.

It is convenient now to introduce notation for the ordering relations that we will use in what follows. We write:

$$A \gg B \text{ if and only if } \sigma(A \cup B) \subseteq B^c,$$

so that we also have

$$A \triangleright B \text{ if and only if either } A \gg B \text{ or } A \triangleright_D B \text{ for some } D \in \mathfrak{A}.$$

Recall also the relation of dominance for singletons (atoms in the algebra), where for $\omega \in \Omega$ and $A \in \mathfrak{A}$ we write

$$\omega \succ_\sigma A \text{ if and only if } \sigma(A \cup \{\omega\}) = \{\omega\} \setminus A,$$

and $A \succeq_\sigma \omega$ otherwise. See the Appendix for a summary of all order relations employed here.

We begin by showing the following Lemma:

Lemma 3.3.15
If $A \triangleright_D B$ then there exists some $d \in \sigma(D \cup B)$ such that $d \succ_\sigma B \cup \big(D \setminus \sigma(D \cup B)\big)$ and $A \cup \big(D \setminus \sigma(D \cup B)\big) \succeq_\sigma d$.

Proof. Suppose that for all $d \in \sigma(D \cup B)$, we have $A \cup \big(D \setminus \sigma(D \cup B)\big) \not\succeq_\sigma d$. List all elements of $\sigma(D \cup B)$ as $\sigma(D \cup B) = \{d_1, ..., d_l\}$. By definition of the $\succeq_\sigma$ relation, we then have

$$\sigma\Big(A \cup \big(D \setminus \sigma(D \cup B)\big) \cup \{d_i\}\Big) = \{d_i\} \quad \text{for all } i \leq l.$$

By ($\mathsf{S4}_n$), we can write

$$\sigma\Big(A \cup \big(D \setminus \sigma(D \cup B)\big) \cup \bigcup_{i \leq l} \{d_i\}\Big) \subseteq \bigcup_{i \leq l} \{d_i\},$$

which simply means

$$\sigma(A \cup D) \subseteq \sigma(D \cup B),$$

so that we get $\sigma(A \cup D) \subseteq \sigma(D \cup B) \subseteq D \subseteq A^c$, and $\sigma(A \cup D) \subseteq A^c$ contradicts the third separation condtion; so $A \not\triangleright_D B$. The supposition is thus false and we can conclude that there exists a $d^\star \in \sigma(D \cup B)$ such that $A \cup \big(D \setminus \sigma(D \cup B)\big) \succeq_\sigma d^\star$.

Next, we need to show $\exists d \in \sigma(D \cup B)$ s.t. $d \succ_\sigma B \cup \big(D \setminus \sigma(D \cup B)\big)$. In fact this holds for *any* $d \in \sigma(D \cup B)$, as can be seen by a simple argument from property ($\mathsf{S3}$) (rational montonicity)[35]. In particular, this holds of the $d^\star$ that we have chosen above: so $d^\star$ is our desired witness. □

Here is what the Lemma says: suppose D directly dominates B but does not directly dominate A. Then there is some 'witness' $\mathfrak{A}$-atom $d \in \sigma(D \cup B)$ that directly dominates $B \cup \big(D \setminus \sigma(D \cup B)\big)$, but does not dominate $A \cup \big(D \setminus \sigma(D \cup B)\big)$[36].

Now assume that the Generalised Fishburn Axiom does not hold. We show that the (Scott) axioms fails too. The failure of the Fishburn axiom means that there exist two sequences of events $(A_i)_{i \leq n}$, $(B_i)_{i \leq n}$ such that

[35] We have

$$d \in \sigma(D \cup B) \cap (B \cup \{d\}) \neq \emptyset,$$

so by ($\mathsf{S3}$) we get

$$\sigma\big((D \cup B) \cap (B \cup \{d\})\big) \subseteq \sigma(D \cup B) \cap (B \cup \{d\}) = \{d\};$$

by ($\mathsf{S1}$), we know the set on the left-hand side is nonempty, so

$$\sigma\big((D \cup B) \cap (B \cup \{d\})\big) = \sigma\big(B \cup (D \setminus \sigma(D \cup B))\big) = \{d\}.$$

Thus we have $d \succ_\sigma B \cup \big(D \setminus \sigma(D \cup B)\big)$, for any $d \in \sigma(D \cup B)$.

[36] Note what this means in terms of probabilistic representation: d is an atom such that

$$\mu(B) + \mu\big(D \setminus \sigma(D \cup B)\big) < \mu(d) \leq \mu(A) + \mu\big(D \setminus \sigma(D \cup B)\big),$$

so in this sense d is an 'atomic' witness to the fact that $\mu(B) < \mu(A)$.

- $(A_i)_{i\leq n} \geq_0 (B_i)_{i\leq n}$ (that is, $\sum_{i=1}^n \mathbb{1}_{A_i} \geq \sum_{i=1}^n \mathbb{1}_{B_i}$), and
- $\forall i \leq n$, $B_i \rhd A_i$.

We can write down the two sequences of events as follows:

$$\begin{array}{ccccccc} (A_1, & \ldots, & A_k, & A_{k+1} & \ldots, & A_n) \\ \Mapsto & \ldots & \Mapsto & \triangle^{D_{k+1}} & \ldots & \triangle^{D_n} \\ \underbrace{(B_1, \quad \ldots, \quad B_k,}_{\text{I}} & & & \underbrace{B_{k+1} \quad \ldots, \quad B_n)}_{\text{II}} & & \end{array}$$

We group the pairwise comparisons as follows: group I consists of those pairs A_j, B_j where B_j directly dominates A_j, while group II consists of those pairs where some $D_j \in \mathfrak{A}$ separates A_j from B_j (and B_j does not directly dominate A_j). We transform this into a sequence of events violating the Scott axiom.

Step 1: for each pair A_j, B_j in group II, pick a witness d_j as given by the Lemma: i.e, $d_j \in \sigma(D_j \cup B_j)$ such that

$$d_j \succ_\sigma A_j \cup \big(D_j \setminus \sigma(D_j \cup A_j)\big)$$

and

$$B_j \cup \big(D_j \setminus \sigma(D_j \cup A_j)\big) \succeq_\sigma d_j.$$

Then, in the array of comparisons above, replace each inequality of the form $B_j \rhd_{D_j} A_j$ by two comparisons by means of the following transformation:

$$\begin{array}{ccccc} A_j & & \{d_j\} & & A_j \cup \big(D_j \setminus \sigma(D_j \cup A_j)\big) \\ \triangle^{D_j} & \rightsquigarrow & \curlywedge^\sigma & & \curlywedge^\sigma \\ B_j & & B_j \cup \big(D_j \setminus \sigma(D_j \cup A_j)\big) & & \{d_j\} \end{array}$$

Let $(A_i^1)_{i\leq m}$ and $(B_i^1)_{i\leq m}$ be the two sequences of events obtained after each such replacement: namely, each A_j in group II has been replaced by two events $(\{d_j\}, A_j \cup \big(D_j \setminus \sigma(D_j \cup A_j)\big))$ and each B_j with the corresponding $(B_j \cup \big(D_j \setminus \sigma(D_j \cup A_j)\big), \{d_j\})$. It is easy to see that we still have $(A_i^1)_{i\leq m} \geq_0 (B_i^1)_{i\leq m}$: at each replacement we simply added one copy of d_j and one copy of $D_j \setminus \sigma(D_j \cup A_j)$ to each side of the sequence, and so the balance of occurrences of each $\omega \in \Omega$ in the two sequences of events remains the same.

Step 2: we carry out a balancing argument analogous to the proof of Proposition 3.3.4. For each $\omega \in \Omega$ that has more occurrences in the sequence $(A_i^1)_{i\leq m}$ we add exactly

$$\eta_{A^1}(\omega) - \eta_{B^1}(\omega) = |\{i \leq m \,|\, \omega \in A_i^1\}| - |\{i \leq m \,|\, \omega \in B_i^1\}|$$

copies of a singleton $\{\omega\}$ to the sequence B^1, and add a corresponding occurrence of $\emptyset$ to the sequence A^1 with the same index. Let $(A_i^*)_{i\leq p}$ and $(B_i^*)_{i\leq p}$ the new sequences thus obtained. For each such added pair $A_i^* := \emptyset$ and $B_i^* := \{\omega\}$, the inequality $\{\omega\} \succ_\sigma \emptyset$ holds in the selection structure by (S1), and so we have $B_i^* \succ_\sigma A_i^*$ for all added sets. Further, the balancing step ensures that the sequences are balanced: $(A_i^*)_{i\leq p} \equiv_0 (B_i^*)_{i\leq p}$.

Step 3: we now have balanced sequences $(A_i^*)_{i\leq p}$ and $(B_i^*)_{i\leq p}$ such that for all $i \leq p$, either $B_i^* \succeq_\sigma A_i^*$ or $\sigma(A_j^* \cup B_j^*) \subseteq B_j^* \setminus A_j^*$, *and the latter holds for for some* $j \leq p$[37]. To conclude the proof, we only need to show that this entails the failure of the (Scott) axiom.

Once again, we divide these two sequences in two groups as follows:

$$\begin{array}{ccccccc} \big(A_1^*, & \ldots, & A_k^*, & A_{k+1}^* & \ldots, & A_p^*\big) \\ \Mapsto & \ldots & \Mapsto & \curlywedge^\sigma & \ldots & \curlywedge^\sigma \\ \underbrace{\big(B_1^*, \quad \ldots, \quad B_k^*,}_{\text{I}'} & & & \underbrace{B_{k+1}^* \quad \ldots, \quad B_p^*\big)}_{\text{II}'} & & \end{array}$$

Group I′ contains all A_j^*, B_j^* with $\sigma(A \cup B) \subseteq B_j^* \setminus A_j^*$: we know that there is at least one such pair (in particular, note that it includes all comparisons of the form $\{d_j\} \succ_\sigma A_j \cup \big(D_j \setminus \sigma(D_j \cup A_j)\big)$ obtained from the transformation performed in Step 1). Group II′ contains all non-strict $\preceq_\sigma$-inequalities of the form $A_j^* \preceq_\sigma B_j^*$: in particular, given the procedures performed in Step 1 and Step 2, this is exactly the collection of all and only comparisons of the form $B_j \cup \big(D_j \setminus \sigma(D_j \cup A_j)\big) \succeq_\sigma \{d_j\}$ obtained at the previous step.

Given each A_j^*, B_j^* in group I′, choose an arbitrary $b_j \in \sigma(A_j^* \cup B_j^*)$. Given the fact that $\sigma(A_j^* \cup B_j^*) \subseteq B_j^* \setminus A_j^*$, an application of (S3) shows that we have $\{b_j\} \succ_\sigma A_j^*$. Next, list (without repetitions) the elements of $B_j^* \setminus \{b_j\}$ as $\{\omega_1, \ldots, \omega_l\}$: for each ω_i we have $\omega_i \succ_\sigma \emptyset$. So we can perform the following transformation on the sequences

$$\begin{array}{ccccccc} A_j^* & & A_j^* & \emptyset & \ldots & \emptyset \\ \Mapsto & \rightsquigarrow & \curlywedge^\sigma & \curlywedge^\sigma & \ldots & \curlywedge^\sigma \\ B_j^* & & \{b_j\} & \{\omega_1\} & \ldots & \{\omega_l\} \end{array}$$

In other words, we replace each A_j^* with a sequence $\big(A_j^*, \emptyset, \ldots, \emptyset\big)$ and the corresponding B_j^* with the sequence $(\{b_j\}, \{\omega_1\}, \ldots, \{\omega_l\})$. The sequences $(A_i^2)_{i\leq k}$ and $(B_i^2)_{i\leq k}$ thus obtained are evidently balanced: the transformation only rearranged the positions of each element of Ω in the sequence, but did not modify their total number of occurrences. Now we have

- $(A_i^2)_{i\leq k} \equiv_0 (B_i^2)_{i\leq k}$

[37] To see why: note that even if group I above is empty, the splitting process carried out in Step 2 adds a strict $\succ_\sigma$ comparison between a A^*-set and a B^*-set.

- $\forall i \leq k$, $B_i^2 \succeq_\sigma A_i^2$
- $\exists j \leq k$, $B_i^2 \succ_\sigma A_i^2$

The second statement holds because we already have $A_j^* \preceq_\sigma B_j^*$ for all sets A_i^*, B_i^* in group II′; and for any pair A_i^2, B_i^2 introduced Step 3, we evidently have $B_i^2 \succ_\sigma A_i^2$. The last statement holds because we know that group I′ above is non-empty, so that at least one of the comparisons is strict. Those three properties violate the (Scott) axiom. So the failure of the Generalised Fishburn axiom entails the failure of (Scott). □

While entirely elementary, the argument is instructive in as much as it illustrates how the qualitative axioms for selection structures can be put to use to directly derive (without appealing to a geometric or algebraic argument) the condition on systems of linear inequalities that is implicitly captured by the Fishburn axiom. Note that the proof uses all the properties (S1), (S2), (S3), ($\mathsf{S4}_n$) and (Scott).

The Generalised Fishburn Axiom can be employed to characterise an interesting class of structures that approximate probabilistically stable revision.

Definition 3.3.16 (**Fishburn structures**)
A Fishburn *structure is a selection structure* $(\Omega, \mathfrak{A}, \sigma)$ *satisfying the following for all* $X, A, B \in \mathfrak{A}$*:*

(S1) $\sigma(X) = \emptyset$ *only if* $X = \emptyset$

(S2) $\sigma(X) \subseteq X$

(S3) *If* $\sigma(A) \cap B \neq \emptyset$*, then* $\sigma(A \cap B) \subseteq \sigma(A) \cap B$

($\mathsf{S4}_n$) *For any* n*: if* $\sigma(A \cup X_i) = X_i$ *for all* $i \leq n$*, then* $\sigma(A \cup \bigcup_{i \leq n} X_i) \subseteq \bigcup_{i \leq n} X_i$

(GFA) *Whenever* $(A_i)_{i \leq n} \leq_0 (B_i)_{i \leq n}$ *and* $A_i \triangleright B_i$ *for all* $i \leqslant n-1$*, then* $\neg(A_n \triangleright B_n)$.

We can observe the following:

Observation 3.3.17
Let $(\Omega, \mathfrak{A}, \sigma)$ *a Fishburn structure. There exists a measure that weakly represents the induced order* $\triangleright$*. If* μ *is any such measure, we have that for any* $A \in \mathfrak{A}$*, the event* $\sigma(A) \in \mathfrak{A}$ *is* $\mu(\cdot \mid A)$*-stable (for threshold* $t = 1/2$*).*

Proof. Given the Generalised Fishburn Axiom (GFA), by Theorem 3.3.3, there exists a measure μ on $\mathfrak{A}$ such that for any $X, Y \in \mathfrak{A}$ we have that $X \triangleright Y$ entails $\mu(X) > \mu(Y)$. Let $A \in \mathfrak{A}$ with $A \neq \emptyset$. We have $\sigma(A) \neq \emptyset$ by (S1). We show that $\sigma(A)$ is μ_A-stable. Firstly, $\sigma(A) \subseteq A$ by (S2). All we need to show is the following: for every $\omega \in \sigma(A)$, we have

$\mu_A(\omega \mid A \setminus \sigma(A)) > 1/2$. Let $\omega \in \sigma(A)$. Write $B := (A \setminus \sigma(A)) \cup \{\omega\}$. Then $\sigma(A) \cap B = \{\omega\} \neq \emptyset$ so by (S3) we have $\sigma(A \cap B) \subseteq \sigma(A) \cap B$, which means

$$\sigma\big([A \setminus \sigma(A)] \cup \{\omega\}\big) = \{\omega\}.$$

This entails $\{\omega\} \triangleright [A \setminus \sigma(A)]$, so $\mu(\omega) > \mu(A \setminus \sigma(A))$. We get $\mu_A(\omega \mid A \setminus \sigma(A)) > 1/2$, as desired. Note that the argument does not rely on the axiom ($\mathsf{S4}_n$). □

The upshot is that Fishburn structures capture a class of revision operators that *respect probabilistic stability*: given any (nonempty) revision input A, the strongest accepted proposition $\sigma(A)$ is stable with respect to the updated measure $\mu(\cdot \mid A)$. However, it need not be the *logically strongest* stable set, and the selection functions do not capture probabilistically stable revision in the form of strongest-stable-set operators.

Fishburn structures thus comply with the non-reductionistic stability rule discussed in the previous chapter (§ 2.3.1). There, the only criterion imposed on revisions was simply that the strongest accepted proposition be probabilistically stable with respect to the updated measure. In our discussion we noted the stability constraint *alone* was too weak to identify interesting revision operators (and to rule out trivial revisions which always select the least set with probability 1 after conditioning). By contrast, the revision operators captured by Fishburn structures constitute a relatively well-behaved family that complies with the stability requirement: Fishburn structures approximate probabilistically stable revision, in that they satisfy the strong monotonicity principle (S3) corresponding to Rational Monotonicity, as well as the axiom ($\mathsf{S4}_n$). Moreover, they are partially representable by a probability measure, in that the strict dominance order generated by the selection function is numerically representable. This wider class of revision operators also admits a simpler axiomatisation, as the $\triangleright$ relation – as opposed to the relation $\succcurlyeq_\sigma$ employed in the (Scott) axiom – does not depend on the property of being an atom in the algebra. We discuss next the role of the scheme ($\mathsf{S4}_n$) in our representation theorem and its relation to the (Or) rule.

3.3.3 Minimisation operators and the Or rule again

As we pointed out in Section 3.2.3, strongest-stable-set operators cannot be represented as a map $X \mapsto \min_R(X)$ for some order relation $R \subseteq \Omega^2$. In other words, probabilistically stable revision cannot be tracked using a *minimisation operator* for a plausibility relation. We also observed that strongest-stable-set operators, treated as selection functions, do not validate the (Or) rule. The former fact easily follows from the latter. As noted by e.g. van Benthem [60] and Rott [47] the following conditions are necessary and sufficient for representability as a minimisation operator.

Proposition 3.3.18 (van Benthem [60])
Let Ω a finite set. Given a function $\sigma : \mathcal{P}(\Omega) \to \mathcal{P}(\Omega)$, the following are equivalent:

- *σ satisfies the following properties for any $X_i \subseteq \Omega$:*

  (1) $\sigma(X) \subseteq X$

  (2) $\sigma(\bigcup_{i\leq n} X_i) \subseteq \bigcup_{i\leq n} \sigma(X_i)$

  (3) $\bigcap_{i\leq n} \sigma(X_i) \subseteq \sigma(\bigcup_{i\leq n} X_i)$

- *There is an asymmetric binary relation $R \subseteq \Omega^2$ such that for all X:*

$$\sigma(X) = \min_R(X) := \{\omega \in X \mid \neg\exists v \in X,\ R(v,\omega)\}$$

As a quick verification reveals, strongest-stable-set operators validate both (1) and (3), but fail (2). The failure of (2) is unsurprising: interpreting $X_i \mathrel{|\!\sim} X_j$ as $\sigma(X_i) \subseteq X_j$, the property corresponds to the (Or) rule.

It is worth noting, however that a weaker form of the (Or) rule does obtain for probabilistically stable revision:

Observation 3.3.19
For any representable selection function σ on an algebra $\mathfrak{A}$, we have that the following holds for any finite collection of events X_i $(i \leq n)$ in $\mathfrak{A}$:

$$\text{If } X_i \setminus X_j \subseteq \sigma(X_i) \text{ for all } i \neq j, \text{ then } \sigma(\bigcup_{i\leq n} X_i) \subseteq \bigcup_{i\leq n} \sigma(X_i). \tag{wO}$$

Proof. Let μ be a probability measure representing the selection function σ, so that $\sigma(X) = \tau(\mu_X)$ for all $X \in \mathfrak{A}$. Assume $X_i \setminus X_j \subseteq \sigma(X_i)$ for all $i, j \leq n$ with $i \neq j$. It is enough to show that $\bigcup_{i\leq n} \sigma(X_i)$ is stable with respect to the measure $\mu(\cdot \mid \bigcup_{i\leq n} X_i)$: this suffices, since $\sigma(\bigcup_{i\leq n} X_i)$ is the *strongest* stable set w.r.t $\mu(\cdot \mid \bigcup_{i\leq n} X_i)$. Let $\omega \in \bigcup_{i\leq n} \sigma(X_i)$[38], and consider the relative complement $\bigcup_{i\leq n} X_i \setminus \bigcup_{i\leq n} \sigma(X_i)$. Since $X_i \setminus X_j \subseteq \sigma(X_i)$ for all distinct i, j, this means that $\bigcup_{i\neq j}(X_i \setminus X_j) \subseteq \bigcup_{i\leq n} \sigma(X_i)$. So we get $\bigcup_{i\leq n} X_i \setminus \bigcup_{i\leq n} \sigma(X_i) \subseteq (\bigcup_{i\leq n} X_i) \setminus \bigcup_{i\neq j}(X_i \setminus X_j) = \bigcap_{i\leq n} X_i$.

This establishes that $\bigcup_{i\leq n} X_i \setminus \bigcup_{i\leq n} \sigma(X_i) \subseteq \bigcap_{i\leq n} X_i$; we can then also write

$$\bigcup_{i\leq n} X_i \setminus \bigcup_{i\leq n} \sigma(X_i) \subseteq X_j \setminus \sigma(X_j)$$

[38] Here again, we take a singleton to represent an atom for simplicity, but this is immaterial: the argument applies for any choice of $\mathfrak{A}$-atom in $\bigcup_{i\leq n} \sigma(X_i)$ instead of $\{\omega\}$.

for all $j \leq n$. But evidently, since $\omega \in \bigcup_{i \leq n} \sigma(X_i)$ we have $\omega \in \sigma(X_j)$ for some j: this means that $\mu(\omega) > \mu(X_j \setminus \sigma(X_j))$. In particular,

$$\mu(\omega) > \mu\big(\bigcup_{i \leq n} X_i \setminus \bigcup_{i \leq n} \sigma(X_i)\big)$$

where ω was arbitrary in $\bigcup_{i \leq n} \sigma(X_i)$. This establishes that $\bigcup_{i \leq n} \sigma(X_i)$ is stable. □

Since the condition $\sigma(\bigcup_{i \leq n} X_i) \subseteq \bigcup_{i \leq n} \sigma(X_i)$ captures the (Or) rule, we can see (wO) as a substantially weaker form of (Or)-style reasoning: it specifies that the (Or) rule can be applied to a set of antecedents $X_1, ..., X_n$ provided that they satisfy the side condition $X_i \setminus X_j \subseteq \sigma(X_i)$ for all $i \neq j$. This weak (wO) rule can be written in semantic form as follows (here we write its two-premise version):

$$\frac{X_i \setminus X_j \subseteq \sigma(X_i) \text{ for } i \neq j \quad X_1 \mathrel{|\!\sim} A \quad X_2 \mathrel{|\!\sim} A}{X_1 \cup X_2 \mathrel{|\!\sim} A} \ (\mathsf{wO}_2)$$

The side constraints $X_i \setminus X_j \subseteq \sigma(X_i)$ give conditions under which Or-type inferences are valid: all of the $X_i \setminus X_j$ states must be typical given X_i, in the sense of being in the selected subset. It is perhaps more informative to see this as a constraint on *atypical* states: namely, all of the *atypical* X_i states – those in $X_i \setminus \sigma(X_i)$ – must be in X_j.

In addition to outlining a very modest extent to which probabilistic stability obeys a version of case reasoning (or the sure thing principle), this rule can be used to obtain an alternative characterisation of Leitgeb structures: it is enough to replace the axiom $\mathsf{S4_n}$ by the property (wO). This is because the property already entails $\mathsf{S4_n}$.

Observation 3.3.20
Suppose a selection function σ satisfies the property (wO). Then it also satisfies $\mathsf{S4_n}$.

Proof. Suppose σ has property (wO). Suppose we have sets A, X_i $(i \leq n)$, such that $\forall i \leq n$, $\sigma(A \cup X_i) = X_i$. Writing $D_i := A \cup X_i$, we have, for each $i \neq j$, $D_i \setminus D_j = X_i \setminus (A \cup X_j)$. Now $\sigma(D_i) = \sigma(A \cup X_i) = X_i$, and so we can write $D_i \setminus D_j \subseteq \sigma(D_i)$. By (wO), we can conclude $\sigma(\bigcup_{i \leq n} D_i) \subseteq \bigcup_{i \leq n} \sigma(D_i)$, which is equivalent to $\sigma(\bigcup_{i \leq n} A \cup X_i) \subseteq \bigcup_{i \leq n} X_i$. We have shown $(\mathsf{S4_n})$. □

This observation entails the following reformulation of our representation theorem (Theorem 3.3.8):

Proposition 3.3.21
Let $(\Omega, \mathfrak{A}, \sigma)$ a selection structure. The following conditions are necessary and sufficient for σ to be a strongest-stable-set operator for some underlying (regular) probability measure on $\mathfrak{A}$:

(S1) $\sigma(X) = \emptyset$ *only if* $X = \emptyset$

(S2) $\sigma(X) \subseteq X$

(S3) *If* $\sigma(A) \cap B \neq \emptyset$, *then* $\sigma(A \cap B) \subseteq \sigma(A) \cap B$

(wO) *If* $X_i \setminus X_j \subseteq \sigma(X_i)$ *for all* $i \neq j$, *then* $\sigma(\bigcup_{i \leq n} X_i) \subseteq \bigcup_{i \leq n} \sigma(X_i)$.

(Scott) *If* $(A_i)_{i \leqslant n} \equiv_0 (B_i)_{i \leqslant n}$ *and* $\forall i \leqslant n$, $A_i \succcurlyeq^*_\sigma B_i$, *then* $\forall i \leqslant n$, $A_i \preccurlyeq^*_\sigma B_i$.

This reformulation of the representation theorem is slightly more perspicuous: probabilistically stable revision operations are characterised by reflexivity, rational monotonicity, a weaker form of the (Or) rule, and the (Scott)-type axiom for representability (which guarantees that the 'preference' ordering between atomic and other events implied by σ can be captured quantitatively). Next, we note an interesting connection between our representation problem and the theory of simple voting games.

3.3.4 Connection with simple voting games

Simple voting games are (simple) structures with various interesting properties, studied in game theory and combinatorics [11, 57]. A simple voting game is a pair $(P, \mathcal{W})$, where P usually represents a finite set of voters, and $\mathcal{W} \subseteq 2^P$ the set of winning coalitions, required to be closed under taking supersets. The game admits a quota representation whenever there is a quota $q \in \mathbb{R}$ and a weight function $m : P \to \mathbb{R}^+$ such that

$$A \in \mathcal{W} \Leftrightarrow \sum_{a \in A} m(a) \geq q,$$

which means that there is a way to assign a weight to each player, in such a way that a coalition is winning exactly when the collective weight of the players in the coalition reaches (or surpasses) the required quota q.

Our representation problem bears a close connection to the theory of simple games. Consider a selection structure $(\Omega, \mathfrak{A})$ with $\Omega = \{\omega_1, ..., \omega_n\}$. Let $\mathcal{D}_i := \{X \subseteq \Omega \,|\, \omega_i \succ_\sigma X\}$, the collection of sets dominated by ω_i. Each state ω_i generates a simple voting game $G_i = (\Omega, {\mathcal{D}_i}^c)$, where the closure-under-superset condition is satisfied thanks to the (No-swap) property (see page 66: the propery ensures that if $X \in {\mathcal{D}_i}^c$, $X \subseteq Y$, then $Y \in {\mathcal{D}_i}^c$). Now suppose μ is a representation for the induced order $\succcurlyeq^*_\sigma$. Then we have, for each ω_i, that

$$X \in {\mathcal{D}_i}^c \text{ iff } \mu(X) \geq \mu(\omega_i).$$

In other words, if μ is a probabilistic representation for $\succcurlyeq^*_\sigma$, it is also weight representation of each game G_i. More specifically, μ *simultaneously* represents all games G_i, where the quota for each G_i is $\mu(\omega_i)$. Conversely, each such simultaneous quota representation of the

associated system of games $\{G_i\}_{i\leq n}$ gives rise to a probability distribution representing the order $\succcurlyeq^*_\sigma$ (it suffices to normalise each weight by the weight of the grand coalition Ω). Thus, finding necessary and sufficient conditions for $\succcurlyeq^*_\sigma$ to be representable is equivalent to finding the exact conditions for the collection of games $\{G_i\}_{i\leq n}$ to be simultaneously representable in this sense (note that an important aspect of simultaneous representations is that the quotas themselves depend on the weight function).

We can give this problem a possible game-theoretic interpretation in terms of what we could call 'coordinated blocking games'. We first identify each state $\omega_i \in \Omega$ with a player. Each game G_i tells us which coalitions can block any decision that player ω_i supports. If the collection of games $\{G_i\}_{i\leq n}$ is simultaneously representable as in the above, we can say that the voting system is at least minimally coherent, in the sense that one can attribute weights to all players in a uniform manner[39].

In the context of simple games, the (Scott′) axiom above (p. 66) entails that one cannot transform a collection $A_1, ..., A_n$ of winning coalitions into a collection of losing coalitions by a sequence of pairwise exchanges of players from one coalition to another. We know from classical results on simple games (see [57]) that axioms (No-swap) and (Scott′) suffice for each individual game $\{G_i\}_{i\leq n}$ to be weight-representable. The method should be straightforward to those familiar with the comparative probability literature: we identify each coalition with its characteristic function, and the (Scott′) axiom guarantees that there is a hyperplane separating the winning from the losing coalitions. The normal vector to the separating hyperplane determines the desired weight function.

In our case, the full (Scott) axiom on $\succcurlyeq^*_\sigma$ suffices to ensure that there is a way of constructing representations of all the games G_i that are consistent with each other; e.g., that no separating hyperplane for G_i lumps together some winning and losing coalitions from another game G_j.

We can also interpret strongest-stable-set operators in the context of voting games. Consider a weighted voting game $(P, \mathcal{W})$ with weight representation $m : P \to \mathbb{R}^+$. For each player $p \in P$, the *projected* game $G_p = (P \setminus \{p\}, W_p)$ is given by $W_p := \{X \subseteq P \setminus \{p\} \mid \sum_{q\in X} m(q) \geq m(p)\}$. The probability function obtained by normalising m gives rise to a selection function $\sigma : \mathcal{P}(P) \to \mathcal{P}(P)$. Given any subset of players $A \subseteq P$, the function σ outputs the minimal $X \subseteq A$ such that $A \setminus X$ is a losing coalition in every reduced game G_p (with $p \in X$). The set $\sigma(A)$ represents the minimal coalition of players such that individual

[39] We can illustrate this with the following scenario. Imagine that you are a historian researching the voting protocols of an ancient civilization. You know that each province sent a delegate to vote in a national council, but you have no explicit information about how exactly the outcomes of the votes were determined. The only information available is records of some results of the votes which specify which coalitions were able to block which delegate (that is, you have a collection of games $\{G_i\}_{i\leq n}$ like the above). The working hypothesis is that the vote of each delegate was accorded a fixed *weight* – proportional, say, to the population of the delegate's province. In order to check if this hypothesis is at least consistent with your data, you must check if the games $\{G_i\}_{i\leq n}$ are simultaneously representable.

player $p \in \sigma(A)$ can block (or 'veto') the coalition $A \setminus \sigma(A)$. Suppose, for instance, that we play the following game. First, we restrict attention to a subset A of players. Then we play a *champion game* on A: a champion p is picked at random from some pre-selected subset $X \subseteq A$. The champion then votes against the entire opposition $A \setminus X$. Different choices of possible champion sets X yield different chances of winning against the opposition. A *decisive* team is a subset of players $X \subseteq A$ such that, no matter who is chosen from X as a champion, the opposition $A \setminus X$ loses against the champion. We can think of $\sigma(A)$ as the minimal decisive team in the champion game on A.

3.4 Axiomatising logics of stability

With our representation theorem at hand (Theroem 3.3.8), we now have a full characterisation of strongest-stable-set operators: we have thus identified key properties of structures encoding the behaviour of τ-generated revision – or alternatively, τ-generated conditional belief. We close this chapter with a few remarks on axiomatising the logic of probabilistically stable belief.

Consider the language $\mathcal{L}_{\mathsf{KBS}}$ is given by:

$$\varphi, \psi ::= p \mid \varphi \wedge \psi \mid \neg\varphi \mid \mathsf{B}^{\varphi}\psi \mid \mathsf{K}\varphi \mid \mathsf{S}^{\varphi}\psi$$

Here, aside from a standard knowledge operator K and a conditional belief operator B, we also have a selection operator S: the formula $\mathsf{S}^{\varphi}\psi$ is intended to capture the property that ψ *is the strongest stable proposition after conditioning on* φ.

Conditional belief expressions of the form $\mathsf{B}^{\varphi}\psi$ can of course be understood as a non-monotonic conditional $\varphi \mathrel{|\!\sim} \psi$. The departure from the simple syntax of flat conditionals of the form $\alpha \mathrel{|\!\sim} \beta$, as used in the context of non-monotonic logics, has a twofold motivation. On the one hand, it remains closer to standard modal presentations of doxastic logics, and thus facilitates a direct comparison with well-known systems. On the other hand, the introduction of the S-operator allows for a direct formulation of the scheme ($\mathsf{S4_n}$) of representable selection structures, for which the simple language of flat conditionals does not suffice[40], thus allowing to express this distinctive property of strongest-stable-set operators.

The semantics will be given by structures of the following form:

Definition 3.4.1 (**Models and satisfaction relation**)

[40]The difficulty lies in expressing exact equalities of the form $\sigma(A) = B$, as opposed to inclusions of the form $\sigma(A) \subseteq B$. In particular, note that the scheme ($\mathsf{S4_n}$) is *not* sound if we replace, in the antecedent, the equalities $\sigma(A \cup X_i) = X_i$ with inclusions $\sigma(A \cup X_i) \subseteq X_i$.

A $\mathcal{L}_{\mathsf{KBS}}$-model is a structure of the form

$$\mathfrak{M} = \Big(\Omega, \sim, \{\sigma_\omega\}_{\omega\in\Omega}, [\![\cdot]\!]\Big)$$

where

- Ω *is a set of states,*
- $\sim$ *is an equivalence relation,*
- $[\![\cdot]\!] : \mathsf{At}(\mathcal{L}_{\mathsf{KBS}}) \to \mathcal{P}(\Omega)$ *is a valuation assigning a set* $[\![p]\!] \subseteq \Omega$ *to each atomic formula* $p \in \mathcal{L}_{\mathsf{KBS}}$*,*
- *each* $\sigma_\omega : \mathfrak{A}_\omega \to \mathfrak{A}_\omega$ *is a selection function on* $\mathfrak{A}_\omega$*, where* $\mathfrak{A}_\omega$ *is the algebra generated on the equivalence class* $[\omega]_\sim$ *by the valuation:*

$$\mathfrak{A}_\omega := \big\{ [\![\varphi]\!] \cap [\omega]_\sim \,|\, \varphi \in \mathcal{L}_{\mathsf{KBS}},\ \varphi \text{ boolean} \big\}$$

Additionally, we also require:

- *If* $\omega \sim v$ *then* $\sigma_\omega = \sigma_v$

Write $[\![\varphi]\!]_\omega$ *for* $[\![\varphi]\!] \cap [\omega]_\sim$*. We extend* $[\![\cdot]\!]$ *to all of* $\mathcal{L}_{\mathsf{KBS}}$ *following the clauses:*

$$\begin{aligned}
[\![\varphi \wedge \psi]\!] =& [\![\varphi]\!] \cap [\![\psi]\!] \quad \textit{and} \quad [\![\neg\varphi]\!] = \Omega \setminus [\![\varphi]\!] \\
[\![\mathsf{B}^\varphi\psi]\!] =& \quad \Big\{\omega \in \Omega \,\Big|\, \sigma_\omega([\![\varphi]\!]_\omega) \subseteq [\![\psi]\!]_\omega\Big\}. \\
[\![\mathsf{S}^\varphi\psi]\!] =& \quad \Big\{\omega \in \Omega \,\Big|\, \sigma_\omega([\![\varphi]\!]_\omega) = [\![\psi]\!]_\omega\Big\}. \\
[\![\mathsf{K}\varphi]\!] =& \quad \Big\{\omega \in \Omega \,\Big|\, [\omega]_\sim \subseteq [\![\varphi]\!]\Big\}.
\end{aligned}$$

Satisfaction is defined as

$$\mathfrak{M}, \omega \vDash \varphi \Leftrightarrow \omega \in [\![\varphi]\!]$$

This definition gives rise to the following satisfaction relation:

$$\begin{aligned}
\mathfrak{M}, \omega \vDash \alpha \quad &\Leftrightarrow \quad \omega \in [\![\alpha]\!] \text{ for boolean } \alpha \\
\mathfrak{M}, \omega \vDash \mathsf{K}\varphi \quad &\Leftrightarrow \quad [\omega]_\sim \subseteq [\![\varphi]\!] \\
\mathfrak{M}, \omega \vDash \mathsf{B}^\varphi\psi \quad &\Leftrightarrow \quad \sigma_\omega([\![\varphi]\!]_\omega) \subseteq [\![\psi]\!]_\omega \\
\mathfrak{M}, \omega \vDash \mathsf{S}^\varphi\psi \quad &\Leftrightarrow \quad \sigma_\omega([\![\varphi]\!]_\omega) = [\![\psi]\!]_\omega
\end{aligned}$$

For better legibility, we will sometimes write $\mathsf{S}^{\varphi}\psi$ and $\mathsf{B}^{\varphi}\psi$ respectively as $\mathsf{S}(\psi \,|\, \varphi)$ and $\mathsf{B}(\psi \,|\, \varphi)$.

Note that this semantics does not truly allow to talk about nested B or S operators: while it allows expressions of the from $\mathsf{B}(\theta \,|\, \mathsf{B}^{\varphi}\psi)$ (that is, $\mathsf{B}^{\mathsf{B}^{\varphi}\psi}\theta$), each such a formula will be equivalent, in any selection structure, to either $\mathsf{B}^{\top}\theta$ or $\mathsf{B}^{\bot}\theta$, depending on whether $\mathsf{B}^{\varphi}\psi$ holds at the evaluation state in the model or not. With the stability operator, we can express various interesting properties: for example $\mathsf{S}^{\varphi}\varphi$ expresses that the proposition $[\![\varphi]\!]$ is a fixpoint of the stability operator.

The structures of interest to us are *Leitgeb models*:

Definition 3.4.2
The class $\mathbb{L}$ of τ-models (or Leitgeb structures*) consists of $\mathcal{L}_{\mathsf{KBS}}$-structures of the form*

$$(\Omega, \sim, \{\sigma_\omega\}_{\omega\in\Omega}, [\![\cdot]\!])$$

where Ω is a finite set, and each σ_ω is a selection function on $\mathfrak{A}_\omega$ satisfying the following properties:

(S1) *$\sigma(X) = \emptyset$ only if $X = \emptyset$*

(S2) *$\sigma(X) \subseteq X$*

(S3) *If $\sigma(A) \cap B \neq \emptyset$, then $\sigma(A \cap B) \subseteq \sigma(A) \cap B$*

($\mathsf{S4}_n$) *For any n: if $\sigma(A \cup X_i) = X_i$ for all $i \leq n$, then $\sigma(A \cup \bigcup_{i\leq n} X_i) \subseteq \bigcup_{i\leq n} X_i$*

(Scott) *If $(A_i)_{i\leqslant n} \equiv_0 (B_i)_{i\leqslant n}$ and $\forall i \leqslant n$, $A_i \succcurlyeq^*_\sigma B_i$, then $\forall i \leqslant n$, $A_i \preccurlyeq^*_\sigma B_i$.*

In other words, in the light of Theorem 3.3.8, we know that the class $\mathbb{L}$ of consists of selection structures where each σ_ω is representable (via the acceptance rule τ): this means that given a structure $\big(\Omega, \sim, \{\sigma_\omega\}_{\omega\in\Omega}, [\![\cdot]\!]\big)$ in $\mathbb{L}$, for each $\omega \in \Omega$ the structure of the form

$$([\omega]_\sim, \mathfrak{A}_\omega, \sigma_\omega)$$

is probabilistically representable. This corresponds simply to having a measure $\mu_{[\omega]}$ over each equivalence class $[\omega]_\sim$ which represents the underlying selection function σ_ω (recall that $\sim$-equivalent states all have the same selection function). In purely probabilistic terms, whenever $\mathfrak{M}, \omega \vDash \mathsf{B}^{\varphi}\psi$ in a Leitgeb model $\mathfrak{M}$, the formula $\mathsf{B}^{\varphi}\psi$ expresses that the strongest stable set, after conditioning on the event $[\![\varphi]\!] \cap [\omega]_\sim$, is a subset of $[\![\psi]\!] \cap [\omega]_\sim$. The event $[\![\varphi]\!]_\omega = [\![\varphi]\!] \cap [\omega]_\sim$ corresponds to the information available to the agent at state ω: it is the conjunction of the received information $[\![\varphi]\!]$ with the agent's background knowledge $[\omega]_\sim$.

Note also that the algebra of propositions is generated by a boolean valuation, so that each event in the algebra $\mathfrak{A}_\omega$ is definable by a boolean formula at ω. This is as it should be: as the algebra $\mathfrak{A}_\omega$ is intended to represent the internal doxastic state of an agent – the space of hypotheses she is considering – there is no need for events that are not explicitly definable by a formula in the agent's language.

Axioms. The problem of obtaining a complete logic of probabilistically stable conditional belief – or, equivalently, the non-monotonic logic of the τ rule – amounts to finding a syntactic characterisation of all and only $\mathcal{L}_{\mathsf{KBS}}$ formulae that are valid over the class $\mathbb{L}$ of Leitgeb models. In what follows, we make the first few steps towards an axiomatisation and highlight some of the difficulties involved.

We can first observe that the notion of two sequences of events being balanced can be captured using a boolean formula. This follows from the following observation, due to Domotor [14]:

Proposition 3.4.3 (Domotor [14, p. 40])
Let $(A_i)_{i\leq n}$, $(B_i)_{i\leq n}$ be two sequences of events in $\mathfrak{A}$. We have:

$$(A_1, \ldots, A_n) \equiv_0 (B_1, \ldots, B_n)$$

if and only if, for any $k \leq n$,

$$\bigcup_{1\leq i_1<\ldots<i_k\leq n} A_{i_1} \cap \ldots \cap A_{i_k} = \bigcup_{1\leq i_1<\ldots<i_k\leq n} B_{i_1} \cap \ldots \cap B_{i_k}$$

Proposition 3.4.3 allows us to express the fact that two sequences of events are balanced.

Definition 3.4.4
For α_i, $\beta_i \in \mathcal{L}_{\mathsf{KBS}}$, we write

$$(\alpha_1, \ldots, \alpha_n)\mathbb{E}(\beta_1, \ldots, \beta_n) := \bigwedge_{1\leq k\leq n} \Big(\bigvee_{1\leq i_1<\ldots<i_k\leq n} (\alpha_{i_1} \wedge \ldots \wedge \alpha_{i_k}) \leftrightarrow \bigvee_{1\leq i_1<\ldots<i_k\leq n} (\beta_{i_1} \wedge \ldots \wedge \beta_{i_k}) \Big)$$

$$(\alpha_1, \ldots, \alpha_n)\mathbb{E}^+(\beta_1, \ldots, \beta_n) := \bigwedge_{1\leq k\leq n} \Big(\bigvee_{1\leq i_1<\ldots<i_k\leq n} (\alpha_{i_1} \wedge \ldots \wedge \alpha_{i_k}) \rightarrow \bigvee_{1\leq i_1<\ldots<i_k\leq n} (\beta_{i_1} \wedge \ldots \wedge \beta_{i_k}) \Big)$$

We can verify that the balancedness of two sequences is expressible:

Proposition 3.4.5
Let $\mathfrak{M}$ an $\mathcal{L}_{\mathsf{KBS}}$-model, and $v \in \Omega$. We have

$$\mathfrak{M}, v \vDash (\alpha_1, \ldots, \alpha_n)\mathbb{E}(\beta_1, \ldots, \beta_n)$$

if and only if

$$([\![\alpha_1]\!]_v, \ldots, [\![\alpha_n]\!]_v) \equiv_0 ([\![\beta_1]\!]_v, \ldots, [\![\beta_n]\!]_v)$$

Proof. (For better legibility, we drop the subscript v henceforth). For the right to left direction, assume towards a contradiction that

$$\sum_{i\leq n} \mathbb{1}_{[\![\alpha_i]\!]} \neq \sum_{i\leq n} \mathbb{1}_{[\![\beta_i]\!]}.$$

As before, we write $\eta_\alpha(\omega) := \big|\{i \leq n \,|\, \omega \in [\![\alpha_i]\!]\}\big|$ for each $\omega \in \Omega$. Our assumption means that there is some $\omega \in \Omega$ such that $\eta_\alpha(\omega) \neq \eta_\beta(\omega)$. Without loss of generality, we have $\eta_\alpha(\omega) > \eta_\beta(\omega)$. We write $k = \eta_\alpha(\omega)$, and let $j_1, ..., j_k$ the collection of all distinct $j \leq n$ such that $\omega \in [\![\alpha_j]\!]$. Clearly then $\omega \in [\![\alpha_{j_1} \wedge ... \wedge \alpha_{j_k}]\!]$, and so

$$\omega \in \Big\| \bigvee_{1\leq i_1<\ldots<i_k\leq n} (\alpha_{i_1} \wedge ... \wedge \alpha_{i_k}) \Big\|.$$

Now $\eta_\alpha(\omega) > \eta_\beta(\omega)$ entails that for any set of k distinct indices $i_1, ..., i_k$, we have $\omega \notin [\![\beta_{i_1}]\!] \cap ... \cap [\![\beta_{i_k}]\!]$, since ω occurs in only $\eta_\beta(\omega)$ many of the $[\![\beta_i]\!]$'s, and $\eta_\beta(\omega) < k$. So we have that for any set of k distinct indices $i_1, ..., i_k$, $\omega \notin [\![\beta_{i_1} \wedge ... \wedge \beta_{i_k}]\!]$, and so

$$\omega \notin \Big\| \bigvee_{1\leq i_1<\ldots<i_k\leq n} (\beta_{i_1} \wedge ... \wedge \beta_{i_k}) \Big\|.$$

This means that ω witnesses that one of the conjuncts in the formula $(\alpha_1, \ldots, \alpha_n)\mathbb{E}(\beta_1, \ldots, \beta_n)$ does not hold in $\mathfrak{M}$. So we have $\mathfrak{M} \nVdash (\alpha_1, \ldots, \alpha_n)\mathbb{E}(\beta_1, \ldots, \beta_n)$.

Conversely, whenever $\sum_{i\leq n} \mathbb{1}_{[\![\alpha_i]\!]} = \sum_{i\leq n} \mathbb{1}_{[\![\beta_i]\!]}$,, meaning that we have $\eta_\alpha(\omega) = \eta_\beta(\omega)$ for each $\omega \in \Omega$, it is easy to see that whenever $\omega \in [\![\alpha_{j_1}]\!] \cap ... \cap [\![\alpha_{j_k}]\!] = [\![\alpha_{j_1} \wedge ... \wedge \alpha_{j_k}]\!]$ for some collection of distinct indices $j_1, ..., j_k$, there also exists some other set of k indices $i_1, ..., i_k$ such that $\omega \in [\![\beta_{i_1} \wedge ... \wedge \beta_{i_k}]\!]$, and vice versa. This ensures that for any $k \leq n$, the biconditional

$$\bigvee_{1\leq i_1<\ldots<i_k\leq n} (\alpha_{i_1} \wedge ... \wedge \alpha_{i_k}) \leftrightarrow \bigvee_{1\leq i_1<\ldots<i_k\leq n} (\beta_{i_1} \wedge ... \wedge \beta_{i_k})$$

holds. Thus we have $\mathfrak{M} \vDash (\alpha_1, \ldots, \alpha_n)\mathbb{E}(\beta_1, \ldots, \beta_n)$. □

With this characterisation at hand, we can now discuss the problem of axiomatising the logic of Leitgeb structures. Consider the axioms in Figure 3.3. The axioms in (S5) ensure that the K operator is a standard S5 modality (we keep the K operator as primitive for the sake of perspicuous notation: we can in principle treat $\mathsf{K}\varphi$ is as definable by $\mathsf{B}^{\neg\varphi}\bot$). The axioms (Pos) and (Neg) are axioms for positive and negative introspection, respectively,

System L

Axioms:

(S5) S5 axioms for K

(KtB) $\mathsf{K}\varphi \to \mathsf{B}^{\psi}\varphi$

(PosM) $\mathsf{M}^{\varphi}\psi \to \mathsf{K}\mathsf{M}^{\varphi}\psi$ for $\mathsf{M} \in \{\mathsf{B}, \mathsf{S}\}$

(NegM) $\neg\mathsf{M}^{\varphi}\psi \to \mathsf{K}\neg\mathsf{M}^{\varphi}\psi$ for $\mathsf{M} \in \{\mathsf{B}, \mathsf{S}\}$

(Reg) $(\mathsf{B}^{\varphi}\bot) \to \mathsf{K}\neg\varphi$

(Ref) $\mathsf{B}^{\varphi}\varphi$

(RW) $\big(\mathsf{B}^{\varphi}\psi \wedge \mathsf{K}(\psi \to \theta)\big) \to \mathsf{B}^{\varphi}\theta$

(And) $(\mathsf{B}^{\varphi}\psi \wedge \mathsf{B}^{\varphi}\theta) \to \mathsf{B}^{\varphi}(\psi \wedge \theta)$

(RM) $(\mathsf{B}^{\varphi}\theta \wedge \neg\mathsf{B}^{\varphi}\neg\psi) \to \mathsf{B}^{\varphi\wedge\psi}\theta$

(LEqM) $\mathsf{K}(\varphi \leftrightarrow \psi) \to \big(\mathsf{M}^{\varphi}\theta \leftrightarrow \mathsf{M}^{\psi}\theta\big)$ for $\mathsf{M} \in \{\mathsf{B}, \mathsf{S}\}$

(REqS) $\mathsf{K}(\varphi \leftrightarrow \psi) \to \big(\mathsf{S}^{\theta}\varphi \leftrightarrow \mathsf{S}^{\theta}\psi\big)$

(SB) $\mathsf{S}^{\varphi}\psi \to \mathsf{B}^{\varphi}\psi$

(SM) $(\mathsf{S}^{\alpha}\theta \wedge \mathsf{B}^{\alpha}\gamma) \to \mathsf{K}(\theta \to \gamma)$

(RMs) $\big(\mathsf{S}^{\varphi\wedge\gamma}\psi \wedge \mathsf{B}^{\varphi}\psi\big) \to \mathsf{S}^{\varphi}\psi$

(B4$_n$) $\Big(\bigwedge_{i\le n} \mathsf{S}(\psi_i \,|\, \varphi \vee \psi_i)\Big) \to \mathsf{B}\big(\bigvee_{i\le n}\psi_i \,|\, \varphi \vee (\bigvee_{i\le n}\psi_i)\big)$

(GF) $\Big(\mathsf{K}\big((\varphi_1,\ldots,\varphi_n)\mathbb{E}^{+}(\psi_1,\ldots,\psi_n)\big) \wedge \bigwedge_{i\le n-1}\big(\varphi_i \triangleright \psi_i\big)\Big) \to \neg(\varphi_n \triangleright \psi_n)$

(RT) $\big(\mathsf{B}^{\varphi\vee(\psi\vee\theta)}\neg(\psi\vee\theta) \wedge \mathsf{B}^{\psi\vee\eta}\neg(\varphi\vee\eta)\big) \to \mathsf{B}^{\varphi\vee(\eta\vee\theta)}\neg(\eta\vee\theta)$

Inference rules:

$$\frac{\varphi \quad \varphi \to \psi}{\psi}\ (\mathsf{MP}) \qquad \frac{\varphi}{\mathsf{K}\varphi}\ (\mathsf{NecK})$$

Figure 3.3: System L.

of the selection and conditional belief operators. In particular, note that they ensure that states in the same equivalence class in a model have the same selection function. (KtB) ensures that knowledge entails belief: that is, if a formula is true throughout the equivalence class (or, equivalenty, is assumed to be true in the underlying probability model), it remains believed no matter what the agent learns.

The axioms (Reg) through (RM) correspond to all the KLM-style axioms that we discussed at the start of this chapter, and verified the soundness of. It is worth remarking here that conditional belief operators B^{φ} are normal: from (And) and (RW) we can derive $\mathsf{B}^{\varphi}(\psi \to \theta) \to (\mathsf{B}^{\varphi}\psi \to \mathsf{B}^{\varphi}\theta)$; and from φ we can deduce $\mathsf{K}\varphi$ (by (NecK)), and (KtB) then entails $\mathsf{B}^{\psi}\varphi$. Axioms (LEqM) and (REqS) capture straightforward syntax-insensitivity properties of the B and S operators. Axioms (SB) and (SM) ensure that the S-operator captures exactly the strongest stable set (that is, the logically strongest accepted proposition in the algebra). Together they allow to derive the following:

$$\mathsf{S}^{\varphi}\psi \wedge \mathsf{S}^{\varphi}\theta \to \mathsf{K}(\psi \leftrightarrow \theta)$$

That is, they jointly guarantee that, conditional on a given event, the strongest stable set is unique.

Axiom (RMs) expresses the following property: if ψ is believed conditional of φ, but it is also the *strongest* stable belief conditional on some proposition stronger than φ, then ψ is already the strongest stable belief conditional on φ. This is intuitive, and not as redundant as it may first appear: for while its semantic version follows from (S3) (Rational Monotonicity), this particular form seems required nonetheless. To see why, a motivating example is the following: note the following valid implication over Leitgeb models: $\{\neg\mathsf{S}^{\top}p, \mathsf{B}^{\top}p\} \models_{\mathbb{L}} \neg\mathsf{S}^{p}\,p$. Semantically, this again follows from the property (S3) of the selection function. But the argument relies on applying the property directly to the set $\sigma([\![\top]\!])$: without the scheme (RMs), this reasoning cannot be readily carried out proof-theoretically unless we already have some formula γ defining the set: $[\![\gamma]\!] = \sigma([\![\top]\!])$. What is interesting to note is that adding to $\mathsf{S}^{\top}\gamma$ to $\{\neg\mathsf{S}^{\top}p, \mathsf{B}^{\top}p\}$, for *any* γ whatsoever, would allow to derive $\neg\mathsf{S}^{p}\,p$ by using (RM). However, the axioms do not guarantee that for every φ, we have some ψ with $\mathsf{S}^{\varphi}\psi$: in other words, they do not directly capture the fact that each $\sigma(A)$ is definable by a formula. With the axiom (RMs), we can remedy this[41]. A special case of that axiom scheme is

[41] This also suggests an important step in proving completeness: ensuring that we can obtain a maximal consistent set that is S-closed: that is, for every formula φ, there is some ψ such that the set contains $\mathsf{S}^{\varphi}\psi$. Since we are building finite models, we must make sure that this can be done for some 'logically finite' fragment of the language. The set $\{\neg\mathsf{S}^{\top}p, \mathsf{B}^{\top}p, \mathsf{S}^{p}p\}$ illustrates that, without axiom (RMs), we can have sets that are L-consistent but have no model *because* they cannot be extended to an S-closed set. This is because adding $\mathsf{S}^{\top}\gamma$, for *any* γ, would make this set L-inconsistent (use (RM)).

$\big(\mathsf{S}^{\top \wedge p} p \wedge \mathsf{B}^{\top} p\big) \to \mathsf{S}^{\top} p$, so that

$$\vdash_{\mathsf{L}} \neg \mathsf{S}^{\top} p \wedge \mathsf{B}^{\top} p \to \neg \mathsf{S}^{p} p,$$

and the worry is avoided.

The remaining three axioms are less usual and require more attention. First, the axiom scheme ($\mathsf{B4}_n$) corresponds to the scheme ($\mathsf{S4}_n$) of Leitgeb structures. Of particular interest is the scheme (GF):

$$\Big(\mathsf{K}\big((\varphi_1, \ldots, \varphi_n)\mathbb{E}^{+}(\psi_1, \ldots, \psi_n)\big) \wedge \bigwedge_{i \leq n-1} \big(\varphi_i \rhd \psi_i\big)\Big) \to \neg(\varphi_n \rhd \psi_n)$$

In this statement, the formula $\varphi \rhd \psi$ is meant to capture the semantic notion of a *dominance* relation, as introduced in Definition 3.3.12. Recall that the dominance relation between events in a selection structure is defined as follows:

$$A \rhd B \text{ if and only if } \sigma(A \cup B) \subseteq B^c \text{ or } A \rhd_D B \text{ for some } D \in \mathfrak{A}.$$

where $A \rhd_D B$ asserts that D *separates* A from B in the sense of Definition 3.3.10. Formally, the expression $\varphi \rhd \psi$ is an abbreviation defined as follows. First we introduce the following notation:

$$\varphi \rhd_\gamma \psi \;\Leftrightarrow\; \Big(\mathsf{K}\big(\gamma \to \neg(\varphi \vee \psi)\big) \wedge \neg \mathsf{B}^{\gamma \vee \varphi} \neg\varphi \wedge \mathsf{B}^{\gamma \vee \psi} \neg\psi\Big)$$

This formula asserts that γ separates φ from ψ. Then we define

$$\varphi \rhd \psi \;\Leftrightarrow\; \Big(\big(\mathsf{B}^{\varphi \vee \psi} \neg\psi\big) \vee \varphi \rhd_\gamma \psi\Big)$$

With this definition, a moment's reflection yields that (GF) captures the Generalised Fishburn axiom as discussed in the previous section:

Whenever $(A_i)_{i \leq n} \leq_0 (B_i)_{i \leq n}$ and $A_i \rhd B_i$ for all $i \leqslant n-1$, then $\neg(A_n \rhd B_n)$.

Note that each instance of (GF) can contain various γ's as separators inside the formula $\varphi_i \rhd \psi_i$. The expression $\varphi_i \rhd \psi_i$ does not capture a single formula but is itself a scheme, with one instance for each possible separator γ in the subformula $\varphi_i \rhd_{\gamma_i} \psi_i$[42].

What does this Generalised Fishburn mean for selection structures? Recall that an event $[\![\varphi]\!]$ directly dominates another event $[\![\psi]\!]$ whenever we have $\sigma([\![\varphi]\!] \cup [\![\psi]\!]) \subseteq [\![\psi]\!]^c$: this relation is easily captured by the formula $\mathsf{B}^{\varphi \vee \psi} \neg\psi$. Then, the formula $\varphi \rhd_\gamma \psi$ entails

[42]So, even for a fixed sequence of formulas $(\varphi_1, \ldots, \varphi_n, \psi_1, \ldots, \psi_n)$, there is still a separate instance of the scheme (GF) for each way of plugging in various possible γ_i's inside the expressions $\varphi_i \rhd \psi_i$; and for each way of plugging the $n-1$ separators γ_i $(i \leq n-1)$ in the premise of the conditonal, there are infinitely many instances of the scheme, one for each distinct γ_n that can occur in the conclusion $\neg(\varphi_n \rhd \psi_n)$. All in all, resorting to the (GF) scheme yields a rather complex axiomatisation.

that any probability measure representing the underlying selection function must satisfy $\mu([\![\varphi]\!]) > \mu([\![\psi]\!])$: this is witnessed by a *separator* $[\![\gamma]\!]$ – an event which directly dominates $[\![\psi]\!]$ but not $[\![\varphi]\!]$ (again, see our discussion of the Fishburn axiom above). By Fishburn's Theorem 3.3.3, (GF) ensures that the disjunctive order generated by *either* direct domination *or* the existence of separators is weakly probabilistically representable, in the sense that there exists a probability measure μ such that

$$\text{if } \mathfrak{M}, \omega \vDash \varphi \triangleright \psi \quad \text{then } \mu([\![\varphi]\!]_\omega) > \mu([\![\psi]\!]_\omega).$$

In short, (GF) ensures that the dominance relation $\triangleright$ on the underlying algebra of events is probabilistically representable.

Next, (GF) does *not* guarantee the transitivity of the induced order. But a (weaker) form of transitivity does hold for the domination relation. Consider the axiom (RT):

$$(\mathsf{RT}) \quad \big(\mathsf{B}^{\varphi \vee (\psi \vee \theta)} \neg(\psi \vee \theta) \wedge \mathsf{B}^{\psi \vee \eta} \neg(\varphi \vee \eta)\big) \to \mathsf{B}^{\varphi \vee (\eta \vee \theta)} \neg(\eta \vee \theta)$$

Informally, this axiom expresses the following: If $[\![\varphi]\!]$ dominates $[\![\psi \vee \theta]\!]$, and $[\![\psi]\!]$ dominates $[\![\eta]\!]$ in a part that does *not* intersect $[\![\varphi]\!]$, then $[\![\varphi]\!]$ dominates $[\![\eta \vee \theta]\!]$. In other words, we have *transitivity under substitution of dominated sets*. Starting from the set $[\![\psi]\!] \cup [\![\theta]\!]$, we can replace the dominated subset $[\![\psi]\!]$ by a set $[\![\eta]\!]$ that has even smaller probability, given that $[\![\psi]\!]$ dominates it; and since $[\![\varphi]\!]$ dominated $[\![\psi]\!] \cup [\![\theta]\!]$ already, it still dominates the resulting set $[\![\eta \vee \theta]\!] = [\![\eta]\!] \cup [\![\theta]\!]$. This however works only under the condition that the part of $[\![\psi]\!]$ that witnesses the domination of $[\![\psi]\!]$ over $[\![\eta]\!]$ – the part where the probability is concentrated – does not intersect $[\![\varphi]\!]$. The corresponding property of the strongest-stable-set operator is the following:

$$\text{If } \sigma(A \cup B \cup C) \subseteq (B \cup C)^c \text{ and } \sigma(B \cup D) \cap (A \cup D) = \emptyset, \text{ then } \sigma(A \cup D \cup C) \subseteq (D \cup C)^c$$

Writing the direct domination relation $\sigma(A \cup B) \subseteq B^c$ as $A \gg B$, we can express the same property in the following form:

$$A \gg (B \cup C) \text{ and } B \gg D \text{ entail } A \gg (D \cup C), \text{ provided that } \sigma(B \cup D) \subseteq A^c.$$

A clearer form of the axiom (RT) can be written in the form of a non-monotonic inference rule, which puts in evidence the fact that (RT) is a Horn rule.

$$\frac{\varphi \vee (\psi \vee \theta) \;\; |\!\sim \neg(\psi \vee \theta) \qquad \psi \vee \eta \;\; |\!\sim \neg(\varphi \vee \eta)}{\varphi \vee (\eta \vee \theta) \;\; |\!\sim \neg(\eta \vee \theta)}$$

By picking $\theta = \top$ in our axiom scheme, we get an intuitive instance of restricted transitivity of the domination relation:

$$\big(\mathsf{B}^{\varphi \vee \psi} \neg\psi \wedge \mathsf{B}^{\psi \vee \eta} \neg(\varphi \vee \eta)\big) \to \mathsf{B}^{\varphi \vee \eta} \neg\eta$$

This concludes our presentation of system L. An immediate observation is that none of the axioms seem to directly capture the (Scott) property of Leitgeb structures: we discuss this issue next.

The problem with atoms. We notice an immediate difficulty with expressing, in $\mathcal{L}_{\mathsf{KBS}}$, the (Scott) axiom for Leitgeb structures. Recall that the (Scott) property is expressed in terms of the order $\succcurlyeq^*_\sigma$, where $A \succcurlyeq^*_\sigma B$ holds if and only if one of the following obtains:

(1) A is a $\mathfrak{A}$-atom and $\sigma(A \cup B) \subseteq B^c$

(2) B is a $\mathfrak{A}$-atom and $\sigma(A \cup B) \not\subseteq A^c$

Expressing this relation in $\mathcal{L}_{\mathsf{KBS}}$ poses a challenge, as the property of being an atom in the underlying algebra of propositions is not definable in $\mathcal{L}_{\mathsf{KBS}}$. To what extend can we express the required property without explicit reference to atoms?

From the point of view of the underlying system of linear inequalities, if $\sigma(A \cup B) \subseteq B^c$ obtains, then we know that $\mu(A) > \mu(B)$ for any measure μ representing σ, regardless of whether or not A is an atom in the underlying algebra. So it is straightforward to see that we get an equivalent axiomatisation if we remove the atom condition from (1), replacing the condition $A \succcurlyeq^*_\sigma B$ with the statement that either

(1′) $\sigma(A \cup B) \subseteq B^c$, or

(2) B is a $\mathfrak{A}$-atom and $\sigma(A \cup B) \not\subseteq A^c$

Then the first disjunct becomes expressible by a formula of the form $\mathsf{B}^{\alpha \vee \beta} \neg \beta$, which ensures $\sigma([\![\alpha]\!] \cup [\![\beta]\!]) \subseteq [\![\beta]\!]^c$. However, we cannot in the same way remove the atom condition from (2): while $\sigma(A \cup B) \not\subseteq A^c$ entails the inequality $\mu(B) \leq \mu(A)$ if B is indeed a $\mathfrak{A}$-atom, this entailment does evidently not hold in general if B is an arbitrary event in $\mathfrak{A}$ (for a counterexample, consider a uniform measure over two disjoint sets A and B with $1 < |A| < |B|$).

Some remarks on completeness. We leave the problem of a complete axiomatisation for the class of Leitgeb models as an open question. It is worth noting that while the semantics given here is a little unusual (consider, in particular, the S operator), it has the advantage of employing a language close to well-studied modal doxastic logics, while being expressive enough to capture the characteristic properties of the stability operator.

The S modality can be seen as an *exact* belief operator: $\mathsf{S}^\varphi \psi$ means that ψ captures exactly all and only the agent's beliefs, after updating on φ. A similar modality has appeared in various contexts in the literature: a related operator has been used in artificial intelligence to capture the logic of 'only knowing' in nonmonotoning reasoning [36, 62], as well as in epistemic game theory – particularly in logical analyses of the Brandenburger-Keisler

paradox [10, 43], where the operator is used to model *assumptions* of players in a game model (understood as total descriptions of their belief sets). In both cases the modality simply expresses the fact that a formula holds in all and only successors of a state in a Kripke model: the logics resulting from its addition to a basic modal language admit a simple axiomatisation (see [23, 19]).

The main difference with our setup is that Leitgeb structures capture conditional beliefs, while known axiomatisations of this exact belief operator only consider the static, unconditional case. Here it is worthwhile to point out that the logics of exact *conditional* belief for selection functions – such as the one employed here – do not seem to have been studied. The conditional logic of probabilistic stability is a fitting domain of application[43].

A natural question is whether system L captures the conditional doxastic logic of Fishburn structures, with the axiom (GF) ensuring the partial representability of the selection function (in the sense of 3.3.17). Going further, a completeness result for Leitgeb structures will require particular care in constructing countermodels: for one, we would need to guarantee that the (finite) countermodel satisfies the (Scott) property to ensure full representability.

Lastly, we may of course have resort to more (or less) expressive languages for describing Leitgeb structures and its associated probabilistically stable revision operators. We sketch alternative semantics for the logic of probabilistic stability in the concluding chapter.

3.5 Summary

We explored the class of revision operators generated by Bayesian conditioning and Leitgeb's stability rule. We first studied the nonmonotonic consequence relations that the τ-rule generates and noted certain salient properties of τ-consequence that demarcate it from the standard systems proposed in the nonmonotonic logic literature. We addressed the problem of characterising probabilistically stable revision operators through selection function models. We gave a representation theorem for the class of selection function models corresponding to strongest-stable-set operators on finite probability spaces. This axiomatisation of strongest-stable-set operators helps identify important qualitative aspects of stability-based dynamics

[43]The axiomatisation of unconditional 'exact' modalities are made straighforward by the introduction of a complementary modality which quantifies over all inaccessible states in a Kripke model [19]. In the same way, it may be helpful to appeal to a complementary modality B which quantifies over the complement of selected states:

$$\mathfrak{M}, \omega \vDash \mathsf{B}^{\varphi}\psi \qquad \Leftrightarrow \qquad [\![\neg\psi]\!]_{\omega} \subseteq \sigma([\![\varphi]\!]_{\omega})$$

Each operator $\mathsf{S}^{\varphi}\psi$ is then definable by the formula $\mathsf{B}^{\varphi}\psi \wedge \mathsf{B}^{\varphi}\neg\varphi$. Note then that we can replace axiom ($\mathsf{B4}_n$) by the following scheme directly corresponding to the (wOR) property, as explained in Observation 3.3.20:

$$\Big(\bigwedge_{i\leq n,\ i\neq j} \mathsf{B}^{\varphi_i}(\varphi_i \to \varphi_j)\Big) \to \Big(\bigwedge_{i\leq n} \mathsf{B}(\theta \,|\, \varphi_i) \to \mathsf{B}\big(\theta \,|\, \bigvee_{i\leq n} \varphi\big)\Big)$$

.

of belief. We also noted a connection between the representation result and voting games.

With the representation theorem at hand, we also have a full description of the class of selection function models that can provide the semantics for a conditional doxastic logic of the stability rule. From the logical perspective, it appears that a richer syntax is needed in order to formulate interesting properties of the underlying conditional belief structures. As is common in probabilistic logics, the logical modeler is faced here with a trade-off between, one the one hand, a language expressive enough to capture what is distinctive about the stability rule and, on the other, a reasonably well-behaved and well-understood (modal) syntax that would keep the complexity of the resulting logic to a minimum. In any case, it is to be expected that standard canonical-model methods for proving modal completeness here will require additional transformation steps, particularly in order to enforce properties like the (Scott) axiom which, being algebra-dependent, is inexpressible in virtually any standard doxastic logic.

Various discussions about bridging probabilistic and qualitative accounts of belief have focused on the static representation of credal states, and many disputes among proponents of either models have focused on which kind of representation is too coarse- or fine-grained. Despite its faults and idiosyncrasies – such as the failure of (Or) – stability-generated revision not only allows to respect probabilistic dynamics by tracking Bayesian conditioning, but it does so by representing doxastic states at an interesting 'medium' level of granularity: specifying a consequence relation $\mid\!\sim_\mu$ (or, equivalently, a representable selection function) does not require, of course, the full specification of a distribution (nor even a full comparative probability ordering), but it is also less 'qualitative' than most simple models of qualitative revision. In this regard, we noted that τ-generated revision is not representable (or 'trackable') via preferential structures and order-minimisation operators. It crucially relies on capturing specific properties of comparative probability orders, of a more combinatorial flavour.

There is another bridging role that strongest-stable-set operators can play in the representation of uncertain inference. One may note *en passant* that each set of probability measures corresponding to a given strongest-stable-set operator – or, equivalently, a μ-consequence relation, encoding a 'belief state with contingency plan' – is a convex set of probability distributions (a 'credal set', in Bayesian parlance). By contrast, regions on the probability simplex corresponding to the distributions that agree only on *unconditional* beliefs are not necessarily convex. Since Bayesian authors often advocate the convexity requirement for credal sets [37], this may indicate that a Bayesian would be more inclined to see the *full* conditional belief structure generated by the τ rule – as given by a representable $\mid\!\sim$ or a selection function σ – as a legitimate 'qualitative' representation of an agent's belief state (rather than simply taking the raw belief set of the agent). In this sense, the τ-rule and its associated revision operator may be good point to start a conversation between Bayesian

and 'qualitative' reasoners.

4

Further directions and concluding remarks

Here ends our exploration of probabilistic stability, and the stability rule for acceptance.

We studied the tracking problem for the stability rule and showed how an information-theoretic perspective on the stability rule yields a simple yet satisfying bridge between AGM revision and Bayesian conditioning. This analysis restored a modest degree of harmony between probabilistic and qualitative dynamics of information states: a stability-complying agent who – through information loss or storage limitations – stores her information in a qualitative form, but accepts certain probabilistic norms for update (Bayesian conditioning) and the quantification of uncertainty (maximum entropy) will always comply with AGM revision operators.

Secondly, we examined in some detail the behaviour of probabilistically stable sets on finite spaces, and gave a complete characterisation of strongest-stable-set operators through our representation theorem. This result identifies exactly the selection function models for conditional belief operators induced by the stability rule (for threshold $t = 1/2$). We identified several important properties of the resulting non-monotonic logic. Along the way, we proved a useful theorem giving sufficient conditions for the joint representation of a pair of (respectively, strict and non-strict) comparative probability orders, and we pointed out an application of the representation theorem to simple voting games.

Throughout the text we have pointed out various natural and interesting questions that would deserve further investigation. We now close with several remaining topics and loose themes around acceptance rules that may pique the reader's curiosity.

Other logics of probabilistic stability. Other languages can be used to capture the logic of probabilistic stability. One particularly attractive option is to consider a logic with a *typicality* operator ∇ (see [8]), which we can interpret on a selection function model $\mathfrak{M} = (\Omega, \sigma, \sim, [\![\cdot]\!])$ as capturing exactly the selected states (the strongest stable event

conditional on the input). More precisely, we set

$$\mathfrak{M}, \omega \vDash \nabla\varphi \text{ if and only if } \omega \in \sigma_\omega([\![\varphi]\!])$$

Then the B and S operators are definable: we can write $\mathsf{B}^\varphi\psi \leftrightarrow \mathsf{K}(\nabla\varphi \rightarrow \psi)$ and $\mathsf{S}^\varphi\psi \leftrightarrow \mathsf{K}(\nabla\varphi \leftrightarrow \psi)$. This is an expressive language that can capture various interesting properties over Leitgeb models. For instance, consider the formula

$$\mathsf{K}(\nabla\varphi \leftrightarrow \nabla\psi)$$

which states that $\sigma([\![\varphi]\!]) = \sigma([\![\psi]\!])$: any hypothesis that is accepted after learning φ is also accepted after learning ψ. Or, in other words: φ and ψ have the same non-monotonic consequences.

Another useful fact about typicality operators – and one that may render the axiomatisation problem quite interesting – is that they can directly express *iterations* of the selection function. As a consequence, we can describe, in quite some detail, various fine features of the probability measure generating the selection function σ. For example: given $A \in \mathfrak{A}$, we define the σ-depth of A as $d(A) := \min\{n \in \mathbb{N} \mid \sigma^n(A) = A\}$. This gives us an approximate way to asses how concentrated the underlying probability measure is in A. Very roughly, an event of low σ-depth is one on which the measure is spread rather uniformly: a set of high σ-depth is one on which the measure is closer to being 'big-stepped' (to use terminology from [7]), i.e., with large probability gaps between individual atoms. Typicality operators can express the fact that an event $[\![\varphi]\!]$ has depth n, through the formula

$$\neg\mathsf{K}(\nabla^{n-1}\varphi \rightarrow \nabla^n\varphi) \wedge \mathsf{K}(\nabla^n\varphi \rightarrow \nabla^{n+1}\varphi)$$

What is the typicality logic of Leitgeb structures? We leave this as a task for another occasion. More generally, we note that typicality logics have been chiefly studied for selection functions that are representable as order-minimisation operators: there remain many open questions about axiomatising more general classes of selection functions, such as the strongest-stable-set operator, which are not 'trackable' by order-minimisation (for instance, they do not validate the axiom $\nabla\varphi \rightarrow \nabla\nabla\varphi$).

The definitional complexity of probabilistic acceptance rules. Acceptance rules can be classified by their definitional complexity in a sufficiently expressive language. A rough outline of how this can be done is as follows. Suppose we have a first-order language to talk about probability spaces, in which we can quantify over events in the algebra, and with enough resources to contain a modest amount of real arithmetic. For instance, take the first-order language of boolean algebras $\mathcal{L}_{BA}$ with a function symbol μ to be interpreted as a

measure on the algebra (as well as the usual boolean inclusion relation $\sqsubseteq$, boolean operations, and constants $\bot$ and $\top$ for the bottom and top elements of the algebra). Additionally, we take function symbols for basic arithmetical operations to be interpreted over the ordered real field $(\mathbb{R}, \times, +, \leq, 0, 1)$. Aside from ordinary $\mathcal{L}_{BA}$ formulas, we also allow formulas that are recursively built from basic expressions of the form $p_1(\vec{\mathsf{t}_1}) \leq p_2(\vec{\mathsf{t}_2})$, where p_i's are polynomial expressions in the signature $(\times, +, 0, 1)$ (with, say, constants for rationals) and the terms t_i consists of expressions $\mu(q)$, with q a term in $\mathcal{L}_{BA}$ (this is similar to the system of Fagin, Halpern and Meggido [15]).

We can interpret these expressions over structures of the form $(\mathbb{B}, \mathsf{P})$ where $\mathbb{B}$ is a (finite) boolean algebra and P a probability measure on it. Formulas in $\mathcal{L}_{BA}$ are evaluated in $\mathbb{B}$, and expressions of the form $p_1\big(\mu(q_1)\big) \leq p_2\big(\mu(q_2)\big)$ hold if the resulting inequalities are true once each $\mu(q)$ has been replaced by $\mathsf{P}([q])$ (here $[q] \in \mathbb{B}$ is the interpretation of the $\mathcal{L}_{BA}$ term q). As an example: $(\mathbb{B}, \mathsf{P}) \vDash \forall x\big((x \neq \bot) \to \mu(x) > \mathsf{0}\big)$ means that the measure P is regular: for all $a \in \mathbb{B}$, if a is not the bottom element $\bot_{\mathbb{B}}$ in $\mathbb{B}$, then $\mathsf{P}(a) > 0$.

Acceptance rules can then be classified by the definitional complexity of the set of accepted elements in the underlying algebra. For instance, the Lockean rule λ can be uniformly captured by an atomic formula:

$$B^{\lambda}_{\mu}(x) \Leftrightarrow \mu(x) \geq \mathsf{t}$$

where we take the parameter t as a constant symbol for $t \in \mathbb{Q}$. The τ rule can be captured by the following formula:

$$B^{\tau}_{\mu}(x) \Leftrightarrow \exists y(y \sqsubseteq x \wedge \mathsf{St}(y))$$

where $\mathsf{St}(y)$ is a stability predicate defined as

$$\mathsf{St}(y) = \forall z\big((z \neq \bot \wedge z \sqsubseteq y) \to (1 - \mathsf{t})\mu(z) > \mathsf{t}(1 - \mu(y))\big),$$

so that is $B^{\tau}_{\mu}(x)$ is equivalent to

$$\exists y\Big(y \sqsubseteq x \wedge \forall z\big((z \neq \bot \wedge z \sqsubseteq y) \to (1 - \mathsf{t})\mu(z) > \mathsf{t}(1 - \mu(y))\big)\Big).$$

The property of being an accepted proposition is a $\exists\forall$-condition. Similarly, the property of being the strongest stable set is a $\forall\exists$-formula, as it is given by

$$\mathsf{LS}(x) \Leftrightarrow \mathsf{St}(x) \wedge \forall y\big((y \sqsubseteq x \wedge y \neq x) \to \neg\mathsf{St}(y)\big).$$

This approach to classifying acceptance rules appears both natural and elementary. Can we have a useful classification of acceptance rules in terms of their definitional complexity?

As illustrated by the example above, this approach allows to state formally the more 'global' nature of the stability rule: for instance, as opposed to several other acceptance rules suggested in the literature, the acceptance of a hypothesis H does not merely depend on a *local* property (e.g., the value of $\mathsf{P}(H)$), but takes into account its interaction with all events consistent with H – a more general property of the probability space $(\mathbb{B}, \mathsf{P})$. Within such a formal framework for defining acceptance rules, one can employ standard tools from logic to investigate questions of this kind. On the one hand, we can ask how 'local' a given rule is: how much information about the entire measure do we need to have access to in order to determine whether a given hypothesis is accepted? What is the complexity of doing so? On the other, we can try and appeal to known preservation theorems to relate various invariance properties of an acceptance rule to the syntactic shape of its defining formula.

This approach may also allow to relate the definitional complexity of the property B^{α}_{μ} to the computational complexity of checking the condition (often, a linear program) expressed by the definition. In this manner we might, in some cases, establish a ready connection between the definitional complexity of an acceptance rule and the complexity of decision problems for the resulting doxastic logic.

Of course, for more sophisticated rules, we may need a language more complex than comparisons of polynomial expressions over $\mathbb{R}$ (and quantification over a boolean algebra). Nonetheless, even this simple framework could already provide a fruitful perspective on many acceptance rules known in the literature.

Acceptance rules and information loss. See our discussion at the very end of Section 2.4.

Observation sets and relativised stability. In section 2, we motivated probabilistic stability as a notion of robustness under any possible (consistent) new evidence that the agent can receive. Yet, in many learning situations, not every event in the probability space represents information that is relevant or accessible to the agent – for instance, some pieces of information may not be directly observable in a particular experimental setup, and so some propositions may be unlearnable. Thus, not every proposition should count as a potential defeater. From this perspective, stability under any possible information is too stringent a requirement. This suggests a natural generalisation of the stability rule which relies on weakening the requirements for what counts as stable. Rather than requiring stability with respect to *any* proposition consistent with the hypothesis, we can require stability only with respect to a distinguished set of events in the probability space. This distinguished collection of events should correspond to the evidence that the agent considers relevant in the learning scenario: it can for instance correspond to information that the agent can directly learn or observe (e.g., what the agent can measure as an outcome of the experiment being carried

out). A restriction such as this one is rather common in formal learning theory, where the space of admissible evidence is often restricted in a similar way.

We can thus relativise the notion of probabilistic stability – and, more generally, the notion of an acceptance rule – to an evidence set. This idea (already suggested, e.g., in [34]) can be cashed out as follows. Define an *observation space* as a pair $(S, \mathcal{E})$ where S is a probability space $(\Omega, \mathfrak{A}, \mathsf{P})$ and $\mathcal{E}$ – *the evidence set* – is a subcollection of events in the algebra $\mathfrak{A}$. The set $\mathcal{E}$ is the collection of propositions that count as relevant *evidence*: for instance, it can consist of the class of all propositions that the agent can learn in the learning scenario at hand – e.g., possible *observational data*.

For a simple example, consider a probabilistic learning problem in which data is sequentially sampled form some underlying sample space Ω. In such a context, it is standard to identify the observable data $\mathcal{E}$ with all possible finite data sequences that can be sampled in the course of the experiment. For an elementary instance of an observation space, take, e.g., the coin-tossing space ($\{0,1\}^{\mathbb{N}}$ with its natural Borel σ-algebra) equipped with a Borel probability measure and the evidence set $\mathcal{E}$ consisting of all its cylinder sets[44], corresponding to finitary sample observations. An *augmented* acceptance rule is then an operator mapping each observation space to a belief set – some collection of accepted propositions. We can thus consider an *augmented stability rule*, defined in a manner analogous to the simple stability rule, but based instead on the notion probabilistic stability relative to the evidence set. We say that a hypothesis H is $\mathcal{E}$-stable if and only if $\mathsf{P}(H \mid E) > t$ for all $E \in \mathcal{E}$ such that $E \cap H \neq \emptyset$. For instance, if $\mathcal{E} = \{\Omega\}$, we recover the Lockean rule; if $\mathcal{E}$ consists of the entire algebra of events, we recover Leitgeb's original rule.

Defining acceptance rules relative to an observation space $(S, \mathcal{E})$ yields a more fine-grained notion of acceptance that is tailored to *learning problems* (rather than bare probability spaces). This invites a variety of questions about probabilistic reasoning and uncertain acceptance in this learning-theoretic context: how should the notion of rational acceptance depend on the relevant evidence? Are there any augmented acceptance rules that are well-behaved when applied to natural learning problems? In particular, how does the augmented stability rule fare in this context?

Stability for continuous distributions and statistical learning. In this thesis we have only investigated the behaviour of the stability rule on discrete probability spaces. This is for a good reason: a simple argument shows that the notion of probabilistically stable sets trivialises for continuous distributions – and more generally, on all atomless probability

[44] That is, we have

$$\mathcal{E} := \{[\![s]\!] \mid s \in \{0,1\}^*\}$$

where, for each string s, the cylinder set $[\![s]\!]$ is defined as $\{X \in \{0,1\}^{\mathbb{N}} \mid s \sqsubset X\}$.

spaces – as only sets of measure one are stable[45]. More generally, it is a problematic feature of many acceptance rules that their applicability is limited to purely atomic spaces. For distributions with a little more structure – including typical models of statistical learning, featuring continuous distributions – the rules are either not well-behaved, or not defined at all.

The notion of stability relativised to an evidence set, as introduced in the previous paragraph, can be put to use here. By weakening the requirements for stability, it allows for a non-trivial notion of acceptance that can be applied in the context of continuous distributions and more realistic models of statistical learning. The move to observation spaces is also helpful as it yields a generalisation of the stability rule which makes explicit the role of the observational protocol the agents finds herself reasoning about. While the stability rule no longer trivialises on observation spaces, it becomes a non-trivial matter to find minimal conditions on the observation space that would guarantee the conjunctivity of the rule. A general characterisation of conjunctivity on observation spaces is worth exploring.

It is worth mentioning, however, that the augmented stability rule also suffers from serious difficulties. In particular, Bayesian statistical learning problems constitute one important class of observation spaces (models of sequential learning via i.i.d sampling), on which the generalised stability rule is stuck in a dilemma between non-triviality and conjunctivity. In [41] we show that, in the context of a standard (parametric) Bayesian learning model, the stability rule yields a notion of acceptance that is either trivial (only hypotheses with probability 1 are accepted) or fails to be conjunctive (accepted hypotheses are not closed under conjunctions). The first problem chiefly affects statistical hypotheses; the second one chiefly affects predictive hypotheses about future outcomes. The failure of conjunctivity for the stability rule is particularly salient, as it affects a wide class of consistent Bayesian priors and learning models with exchangeable random variables. These observations highlight a serious tension between (1) being responsive to evidence and (2) having conjunctive beliefs induced by the stability rule. We also show that a similar phenomenon affects probabilistic reasoners with continuous priors in a rather general context: the generalised stability rule is trivial on the class of Borel observation spaces, which capture the structure of learning problems where the observable events constitute a topology on the underlying Borel probability space. This severely limits the rule's scope of applicability.

Impossibility theorems for probabilistic acceptance and voting theory. What explains the scarcity (or, indeed, lack) of well-behaved conjunctive acceptance rules for continuous distributions? One possible answer points to invariance properties of acceptance

[45]Recall a probability space $(\Omega, \mathfrak{A}, \mu)$ is *atomless* if for any $X \in \mathfrak{A}$ with $\mu(X) > 0$ there is some measurable $Y \subset X$ with $0 < \mu(Y) < \mu(X)$. For any event X with $0 < \mu(X) < 1$, we can thus easily find a defeater of the form $X^c \cup D$ where D is a subset of X of sufficiently small measure.

rules. A result by Smith [54] shows that acceptance rules obeying certain desirable closure and invariance principles (closure under finite conjunction, invariance under coarsening and automorphisms) face a dilemma on atomless spaces: they are bound to be either trivial (in the sense of only accepting hypotheses that have probability 1) or inconsistent[46]. We show elsewhere [41] that a related limitative result obtains for augmented stability rules on Borel observation spaces. The result affects acceptance rules satisfying similar invariance conditions and admits a learning-theoretic interpretation: reasonable rules (conjunctive, invariant under coarsening and automorphisms of the observation space) cannot learn hypotheses that are, from a topological and learning-theoretic perspective, rather tame (e.g., both verifiable and falsifiable by the evidence available in the learning problem). This perspective partly clarifies why, in the context of learning with continuous distributions, there seems to be no perfect general-purpose rule for uncertain acceptance, even when acceptance depends on a relevant evidence set: under mild invariance constraints on acceptance rules, aggregating uncertain information leads to inconsistency.

These observations invite the quest to find a satisfactory general diagnosis of the tension between the high-probability requirement on belief, the closure of belief under conjunction, and invariance principles for acceptance. First, is there an appropriate weakening of the above invariance properties that would leave space for a well-behaved acceptance principle, at least on most common learning problems based on atomless spaces?

A broader question asks what lessons should be drawn from these results as to the *question-dependence* of uncertain acceptance. Consider, for instance, the following difference between statistical hypothesis testing and acceptance rules. Classical hypothesis testing typically takes place within a pre-determined question: that is, a partition of our probability space into disjoint and exhaustive hypotheses (consider, for instance, testing a null hypothesis against an alternative hypothesis). The acceptance or rejection of a hypothesis is decided against the backdrop of a finite (often binary) decision space (e.g., 'is ticket number 224 the winning ticket?', or 'is the bias of the coin greater than 0.5?'). No individual decision problem of this kind poses a conjunctivity problem. Problems for conjunctive acceptance occur when the learner asks herself what her answer would be had she asked a different question (and repartitioned the space accordingly); for the answers given to distinct (yet isomorphic) problems need not be logically compatible with each other, nor does their conjunction need preserve high probability. If the space admits enough symmetries taking one problem to another, inconsistencies easily arise. From this perspective, atomless spaces pose a particular problem precisely because they admit many symmetries: they can be coarsened into many distinct isomorphic partitions ('questions'), the answers to which, taken together, yield an

[46]Smith's original theorem is a little weaker, and the published proof [54] contains a minor mistake (a misapplication of Zorn's Lemma). See [41] for a strengthening of Smith's result and a correction of the original proof.

inconsistent set of accepted hypotheses. It is worth asking whether explanations along these lines really do get to the heart of the matter; is it fair to say that conjunctive acceptance rules fall victim to lottery-like paradoxes because they seek to answer 'too many' questions at once?

Another direction to investigate consists in comparing invariance conditions for acceptance rules with fairness conditions in voting and judgment aggregation. Both Smith's impossibility theorem and our limitative results mentioned above, while relying on specific topological and measure-theoretic properties of the underlying probability spaces, fundamentally all employ arguments of a distinct voting-theoretic flavour, reminiscent of impossibility theorems from judgment aggregation and social choice theory. Clarifying the mathematical and conceptual relationships between these two frameworks could be a rewarding task, from the perspectives of logic and probability, as well as social choice theory. This is indeed an avenue pursued, in the context of finite voting scenarios, by Dietrich and List [13].

The algebra of observation sets. In the context of augmented acceptance rules defined on observation spaces, another natural question concerns invariance properties of accepted hypotheses under combinations of evidence (observation) sets. Many natural invariance conditions – such as the ones used in Smith's impossibility theorem [54] – can be formulated as preservation under transformations of the underlying observation space; but, in studying more refined properties of acceptance rules, we can also naturally investigate the behaviour of acceptance rules under transformations applied to evidence sets on a fixed probability space. Consider for instance k observation sets $\mathcal{E}_1, \ldots, \mathcal{E}_k$ on a probability space S, and let $\odot$ an k-ary set operation on observations sets – e.g., take $\odot(\mathcal{E}_1, \mathcal{E}_2) = \mathcal{E}_1 \cap \mathcal{E}_2$. What are reasonable constraints on the relationship between the individual belief sets $K_\alpha(S, \mathcal{E}_i)$ and the collection $K_\alpha\big(S, \odot_i(\mathcal{E}_1, \ldots, \mathcal{E}_k)\big)$? This setting suggests a connection with evidence dynamics in neighbourhood semantics for modal logic [58, 59]. In particular, if we see each observation space as a model, each acceptance rule gives rise to a distinct belief modality, obeying different invariance conditions under transformations of the evidence space. Are there some reasonable structural restrictions on observation spaces that would allow us to classify well-behaved acceptance rules by their invariance conditions? Can this be done in a way that permits informative axiomatisations of the resulting doxastic logics?

Acceptance rules in game and decision theory. Aside from their logic and invariance conditions, *ye shall know good acceptance rules by their fruit*: how well do Bayesian agents perform at decision tasks, or in game-theoretic situations, by applying a particular acceptance rule? An element of response for the stability rule is provided by Leitgeb in [35], who shows how, in a simplified 'qualitativised' decision problem, following actions that the stability

rule accepts to be useful (in a precise sense) can be rationalised as (an approximation of) maximising expected utility. When is a given acceptance rule rationalisable as a form of utility maximisation?

More generally, there are several contexts – decision theoretic, game-theoretic, and learning-theoretic – where one studies the performance of broadly probabilistic agents (guided by probabilistic credences), but which could rather naturally lend themselves to an analysis in terms of acceptance rules. How is an agent's performance affected by employing an acceptance rule in making a decision under uncertainty, or choosing a strategy in a game?

Consider, for instance, iterated or extensive form games. A game proceeds in stages where, at every stage, every player must choose an action: the sequence of actions taken by both players determines their outcome (e.g., a numerical payoff, or a binary win/lose outcome). Each player has a probability distribution which encodes her uncertainty about the *type* of her opponents (what kind of player are they? What strategies are they likely to choose?), which also translates into uncertainty about outcomes in future rounds of the game. These credences can then be translated, through an acceptance rule, into downright categorical *beliefs* that will guide the player's future actions. How does the choice of an acceptance rule affect a player's final performance? When does an acceptance rule outperform another? In particular, these are contexts where the hypothesis space – e.g., hypotheses about the *types* of opposing players – is distinct from the evidence space (observations about the opponent's moves so far). Here using augmented acceptance rules, defined on observation spaces, is appropriate.

In each of these settings, studying the performance of particular acceptance rules can clarify what happens to agents who are not full-blooded Bayesians with access to sharp credences, but operate on the basis of a qualitative representation of their informational state.

A note on Bayesian and frequentist acceptance. Bayesian credible intervals and (frequentist) confidence intervals are two constructions closely related to a notion of acceptance: both are designed to output reasonable hypotheses about the value of an unknown parameter in statistical hypothesis testing. One may be tempted to see a conceptual affinity between probabilistic stability and each of these notions. Can they be reconstructed in terms of probabilistic stability? There is a point here that, although elementary, is worth clarifying, as it highlights an important conceptual component of the stability rule.

Suppose we have a parametric learning problem: data is generated by i.i.d sampling from an unknown generating distribution. The generating distribution is one of many possible distributions of the form Q_θ, each determined by a different parameter value θ. The agent has a prior probability over a range Θ of possible values of the parameter θ, representing her credences over which distributions Q_θ she believes are more or less likely to be generating

the data. As more samples are observed, the agent updates her prior by conditioning on the sampling data, leading to new credences about the values of the parameter for the generating distribution. For simplicity (and in order to avoid technical complications with continuous distributions), let us assume that both the parameter space Θ and the sample space Ω are finite, and that the number of samples is fixed in advance (to, say, k observations). In this way we can model the problem using a finite product space $S = (\Theta \times \Omega^k, \mathcal{B}_\Theta \times \mathcal{B}, \mathsf{P})$, with P the joint measure over the product space $\mathcal{B}_\Theta \times \mathcal{B}$, and a threshold $t \in (0.5, 1)$ (with $\mathcal{B}_\Theta$ and $\mathcal{B}$ algebras over the parameter space Θ and the sample space Ω, respectively). Sampling data is represented by a finite sequence of samples $\sigma \in \Omega^*$.

Both credible intervals and confidence intervals seem affiliated with a stability notion. One way to capture this is as follows. A credible interval is a random interval $I : \Omega^k \to \mathcal{B}(\Theta)$ that is 'stable' under any data $\sigma \in \Omega^k$: that is, for any $\sigma \in \Omega^k$, we have $\mathsf{P}(I(\sigma) \,|\, \sigma) > t$. In short: for any data that we observe, we will conjecture a range of values for the parameter that (under our updated probability) has probability at least t.

A confidence interval is a random interval $I : \Omega^k \to \mathcal{B}(\Theta)$ such that the hypothesis $[\theta^* \in I_\sigma] \subseteq \Omega^k$ is high conditional on any parameter value $\theta^* \in \Theta$: that is, for any $\theta^* \in \Theta$, we have $\mathsf{P}(\theta^* \in I_\sigma \,|\, \theta^*) > t$. Here, for a fixed θ^*, the set $[\theta^* \in I_\sigma] = \{\sigma \in \Omega^k \,|\, \theta^* \in I(\sigma)\}$ represents the event that we will observe data σ such that the resulting conjecture $I(\sigma)$ is consistent with θ^*. In short: for any possible value of the parameter, if that value were true, we would have a high ($> t$) probability of observing data that would yield a conjecture consistent with the true parameter value.

So there is temptation here to view credibility intervals as stable sets relative to the evidence set $\mathcal{E}_b := \{\Theta \times [\![\sigma]\!] \,|\, \sigma \in \Omega^k\}$ – possible sampling data – and confidence intervals as stable sets relative to the evidence set $\mathcal{E}_f := \{\{\theta\} \times \Omega^{\mathbb{N}} \,|\, \theta \in \Theta\}$ – possible values of the parameter. But this is misleading. The important difference, of course, is that the ordinary notion of probabilistic stability is defined for fixed hypotheses, whereas credible/confidence intervals are both *random objects* that depend on the data being observed.

More precisely, under the ordinary notion of $\mathcal{E}$-stability for some evidence set $\mathcal{E}$, a hypothesis being $\mathcal{E}$-stable is independent of any particular $E \in \mathcal{E}$ being conditioned on: we fix a hypothesis H and, once this is done, we ask whether this H retains high probability under conditioning on any consistent information E. But in the case of random intervals (both for credible and confidence intervals), the hypothesis being evaluated is *dependent* on the particular evidence $E \in \mathcal{E}$ that we condition on: schematically, each E that we condition on determines a *distinct* hypothesis H_E, and we require $\mathsf{P}(H_E \,|\, E) > t$. The difference can be put in terms of a change in quantifier order: $\mathcal{E}$-stability requires to pick an H such that no matter the $E \in \mathcal{E}$, we have $\mathsf{P}(H \,|\, E) > t$. Confidence and credible intervals require that, for any *given* E, we pick some H_E – dependent on E – such that $\mathsf{P}(H_E \,|\, E) > t$.

In terms of acceptance rules, it is more appropriate to describe the two constructions as conjecturing methods that are required to respect the *Lockean* rule: given every evidence $E \in \mathcal{E}$, we want to output a conjecture H_E that is accepted by the Lockean rule for the conditional measure $\mathsf{P}(\cdot \mid E)$. The difference is that credible intervals require this for hypotheses about the parameter values conditional on data, while confidence intervals require it for hypotheses about data conditional on parameter values. However, since the content of the hypothesis changes with the evidence we are conditioning on, there is no sense in which either construction depends on the notion of probabilistic stability or resilience.

Conclusion. As has been pointed out by Baltag and Smets [6], capturing acceptance as a probabilistic invariance condition aligns itself well with a suggestive view in information dynamics – at the root of a battle-tested research tradition – that seeks to characterise salient informational states by the transformations that leave them invariant. We saw in this thesis that probabilistic stability, as it is with all things, can both hold promise and bring trouble. Despite their simplicity, the notion of probabilistic stability and its associated acceptance rule are a source of attractive logical and mathematical questions, offer a promising philosophical program relating all-out-beliefs to numerical credences (and subjective probability to chance [53]), and raise poignant methodological questions about logical models of statistical and probabilistic reasoning.

It is indeed a fundamental tension in epistemology and in science that we want our methods to be, in the face of change, both responsive and resilient. The hope is that our investigation of the stability rule offers an instructive perspective on the risks and rewards of resilience. We leave this – until new information comes in – as a final word.

Bibliography

[1] Adams, E. W. and Levine, H. P. (1975). "On the Uncertainties Transmitted from Premisses to Conclusions in Deductive Inferences", *Synthese*, 30: 429-460. (page 49)

[2] Alchourrón, C., Gärdenfors, P. and Makinson, D. (1985). "On the logic of theory change: partial meet contraction and revision functions", *Journal of Symbolic Logic*, 50(2): 510-530. (page 5, 8)

[3] Arló-Costa, H. and Pedersen, A.P. (2012). "Belief and Probability: A General Theory of Probability Cores", *International Journal of Approximate Reasoning*, 53(3): 293-315. (page 10)

[4] Arló-Costa, H. and Pedersen, A.P. (2011). "Belief Revision", in Horsten, L. and Pettigrew, R. (eds.), *Continuum Companion to Philosophical Logic*, Continuum Press. (page 8, 13)

[5] Baltag, A., Moss, L. and Solecki, S. (1998). "The logic of public announcements, common knowledge, and private suspicions", in I. Gilboa (ed.), *TARK VII: Proceedings of the 7th Conference on Theoretical Aspects of Rationality and Knowledge*, Evanston, Illinois, USA: 43-56. (page 54)

[6] Baltag, A. and Smets, S. (2014). "On the Trails of Logical Dynamics" in Baltag, A. and Smets, S. (eds.), *Johan van Benthem on Logic and Information Dynamics*, Springer-Verlag. (page 110)

[7] Benferhat , S., Dubois, D. and Prade, H. (1999). "Possibilistic and standard probabilistic semantics of conditional knowledge bases", *Journal of Logic and Computation*, 9(6): 873-895. (page 101)

[8] Booth, R., Meyer, T., and Varzinczak, I. (2012). "PTL: A Propositional Typicality Logic", in *Logics in Artificial Intelligence* (Lecture Notes in Computer Science), 7519, Springer-Verlag Berlin Heidelberg: 107-119. (page 100)

[9] Boyd, S. and Vandenberghe, L. (2004), *Convex Optimization*, Cambridge University Press. (page 38)

[10] Brandenburger, A and Keisler, H. J. (2006)."An Impossibility Theorem on Beliefs in Games" *Studia Logica* 84: 211-240. (page 97)

[11] Chevyrev, I., Searles, D. and Sliinko, A. (2013), "On the Number of Facets of Polytopes Representing Comparative Probability Orders", *Order*, 30(2): 749-761. (page 85)

[12] Delgrande, J. (2012), "Revising Beliefs on the Basis of Evidence", *International Journal of Approximate Reasoning*, 53(3): 396-412. (page 6)

[13] Dietrich, F. and List, C. (2014), "From degrees of belief to binary beliefs: Lessons from judgment-aggregation theory", Munich Personal RePEc Archive paper No. 80844, https://mpra.ub.uni-muenchen.de/80844/1/MPRA_paper_80844.pdf (page 107)

[14] Domotor, Z. (1969), "Probabilistic relational structures and their applications", Technical Report 144, Institute for Mathematical Studies in the Social Sciences, Stanford University. (page 90)

[15] Fagin, R., Halpern, J. and Meggido, N., (1990). "A Logic for Reasoning about Probabilities", *Information and Computation* 87, 78-128. (page 102)

[16] de Finetti, B., (1972), *Probability, Induction and Statistics*, London: Wiley. (page 52)

[17] Fishburn, P. (1969). "Weak Qualitative Probability on Finite Sets", *The Annals of Mathematical Statistics*, 40(6): 2118-2126. (page 67, 70)

[18] Foley, R. (1993). *Working Without a Net*, Oxford: Oxford University Press. (page 9)

[19] Gargov, G., Passy, S. and Tinchev, T. (1987). "Modal Environment for Boolean Speculations" in *Mathematical Logic and its Applications*, Skordev, D. (ed.), Springer-Verlag. 253-263. (page 97)

[20] Grove, A. (1988). "Two modellings for theory change", *Journal of Philosophical Logic*, 17: 157-170. (page 8, 12, 13)

[21] Halpern, J. (2003). *Reasoning About Uncertainty*, Cambridge: MIT Press. (page 8, 30, 31)

[22] Howson, C. (2008). "De Finetti, Countable Additivity, Consistency and Coherence", *British Journal for the Philosophy of Science*, 59: 1-23. (page 52)

[23] Humberstone, L. (1987). "The Modal Logic of 'All and Only"', *Notre Dame Journal of Formal Logic*, 28(2): 177-188. (page 97)

[24] Jaynes, E. T (1957). "Information Theory and Statistical Mechanics", *Physical Review Series II*, 106(4): 620-663. (page 30)

[25] Kelly, K. and Lin, H. (2012). "Propositional Reasoning that Tracks Probabilistic Reasoning", *Journal of Philosophical Logic*, 41(6): 957-981. (page iii, 3, 5, 6, 9, 15, 17, 43, 44, 46, 48, 54)

[26] Kelly, K. and Lin, H. (2012). "A geo-logical solution to the lottery-paradox, with applications to conditional logic", *Synthese*, 186(2): 531-575. (page 2, 17, 44, 46, 47, 48, 49)

[27] Kraft, C., Pratt, J., and Seidenberg, A. (1959). "Intuitive Probability on Finite Sets", *The Annals of Mathematical Statistics*, 30(2): 408-419. (page 65)

[28] Kraus, S., Lehmann, D., and Magidor, M. (1990). "Nonmonotonic Reasoning, Preferential Models and Cumulative Logics", *Artificial Intelligence*, 44(1-2): 167-207. (page 45, 47, 49, 50, 53)

[29] Kyburg, H. (1970). *Probability and Inductive Logic*, Toronto: Macmillan. (page 2, 6)

[30] Leitgeb, H. (2013). "Reducing belief simpliciter to degrees of belief", *Annals of Pure and Applied Logic*, 164: 1338-1389. (page iii, 2, 3, 6, 9, 11, 12, 22, 25, 59)

[31] Leitgeb, H. (2012). "A Joint Theory of Belief and Probability: Comparing the Theories", talk given at the *MCMP Round Table on Acceptance* on February 3rd, 2012. Slides available at http://www.mcmp.philosophie.uni-muenchen.de/events/archive/2012/add_act/r_t_accept/round_table/comp_kevin_hanti_fin.pdf. (page 21, 25)

[32] Leitgeb, H. (2014). "Belief as a simplification of probability, and what it entails", in Baltag, A. and Smets, S. (eds.), *Johan van Benthem on Logic and Information Dynamics*, Springer-Verlag. (page 2, 58)

[33] Leitgeb, H. (2014). "The Stability Theory of Belief", *Philosophical Review*, 123 (2): 131-171. (page 2, 10)

[34] Leitgeb, H. (2014). "A Stability Theory of Belief and Degrees of Belief", slides available at http://www.math.lmu.de/ ptc14/Leitgebslides.pdf. (page 104)

[35] Leitgeb, H. (2017). *The Stability of Belief*, Oxford University Press. (page 27, 28, 107)

[36] Levesque, H. J. (1990). "All I know: A study in autoepistemic logic", *Artificial Intelligence* 42, 263-309. (page 96)

[37] Levi, I. (1980). *The Enterprise of Knowledge: An Essay on Knowledge, Credal Probability, and Chances*, Cambridge (MA): MIT Press. (page 44, 98)

[38] Lin, H. (2011). "A New Theory of Acceptance that Solves the Lottery Paradox and Provides a Simplified Probabilistic Semantics for Adams' Logic of Conditionals", MSc Thesis, Carnegie Mellon University. (page 48, 49)

[39] McCarthy, J. (1980). "Circumscription, a form of non monotonic reasoning", *Artificial Intelligence*, 13: 27-39. (page 53)

[40] Makinson, D. (1988). "General theory of cumulative inference", in Reinfrank, M., de Kleer, J., Ginsberg, M. and Sandewall, E. (eds), *Proceedings of the Second International Workshop on Non-Monotonic Reasoning* (Lecture Notes in Artificial Intelligence), 346, Springer-Verlag. (page 53)

[41] Mierzewski, K. (manuscript) "Probabilistic stability and acceptance on atomless spaces", Stanford University. (page 105, 106)

[42] Motzkin, T. S. (1951). "Two consequences of the transposition theorem on linear inequalities", *Econometrica*, 19(2): 184-185. (page 66, 67)

[43] Pacuit, E. (2007). "Understanding the Brandenburger-Keisler Paradox", *Studia Logica* 86: 435-454. (page 97)

[44] Paris J.B. (1994). *The uncertain reasoner's companion: a mathematical perspective*, Cambridge Tracts in Theoretical Computer Science 39, Cambridge: Cambridge University Press. (page 8, 31)

[45] Plaza, J. (2007). "Logics of public communications", *Synthese*, 158(2): 165-179. (page 54)

[46] Roman, S. (1997). *Coding and Information Theory*, Graduate Texts in Mathematics: Springer-Verlag. (page 8)

[47] Rott, H. (2001). *Change, Choice and Inference*, Oxford Logic Guides, Oxford University Press. (page 82)

[48] Savage, L. J. (1954). *The foundations of statistics*, John Wiley & Sons Inc., New York. (page 54)

[49] Schrijver, A. (2013). *A Course in Combinatorial Optimization*, Lecture notes, Universiteit van Amsterdam. Available at http://homepages.cwi.nl/ lex/files/dict.pdf. (page 49, 66)

[50] Scott, D. (1964). "Measurement structures and linear inequalities", *Journal of Mathematical Psychology*, 1(2): 233-247. (page 58, 65, 66)

[51] Seidenfeld, T., Schervish, M.J. and Kadane, J.B. (1998). "Non-conglomerability for finite-valued finitely additive probability", *The Indian Journal of Statistics*, 60(3): 476-491. (page 52)

[52] Seidenfeld, T., Schervish, M.J. and Kadane, J.B. (2017). "Non-conglomerability for countably additive measures that are not κ-additive", *The Review of Symbolic Logic*, 10(2): 284-300. (page 52)

[53] Skyrms, B. (1977). "Resiliency, Propensities, and Causal Necessity", *The Journal of Philosophy*, 74(11): 704-711. (page 2, 110)

[54] Smith, M. (2010). "A Generalised Lottery Paradox for Infinite Probability Spaces", *British Journal for the Philosophy of Science*, 61 (4): 821-831. (page 106, 107)

[55] Stoer, J. and Witzgall, C. (1970). "Convexity and optimization in finite dimensions I", *Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen mit besonderer Berüsichtigung der Anwendungsgebiete*, 163, Berlin: Springer-Verlag. (page 66)

[56] Sundaram, R. K. (1996). *A first course in optimization theory*, Cambridge University Press. (page 37)

[57] Taylor, A. and Zwicker, W. (1999). *Simple Games: Desirability Relations, Tradings, Pseudoweightings*, Princeton (NJ): Princeton University Press. (page 4, 47, 85, 86)

[58] van Benthem, J. and Pacuit, E., (2011). "Dynamic Logics of Evidence-Based Beliefs", *Studia Logica*, 99: 61-92. (page 107)

[59] van Benthem, J., Fernández-Duque, D., and Pacuit, E., (2014). "Evidence and plausibility in neighborhood structures", *Annals of Pure and Applied Logic*, 165: 106-133. (page 107)

[60] van Benthem, J., (1989). "Semantic Parallels in Natural Language and Computation", in *Logic Colloquium '87*, *Studies in Logic and the Foundations of Mathematics*, Volume 129, Elsevier: 331-375. (page 82, 83)

[61] van Fraassen, B. (1980). *The scientific image*, Oxford: Oxford University Press. (page 6)

[62] Waaler, A., Klüwer, J., Langholm, T. and Lian, E., (2007). "Only knowing with degrees of confidence", *Journal of Applied Logic* 5(3), 492-518. (page 96)

Appendix: order relations

Order relations on selection structures $(\Omega, \mathfrak{A}, \sigma)$	
Order	Definition
$A \gg B$	$\sigma(A \cup B) \subseteq A \setminus B$
$A \mathrel{\triangleright_D} B$	$D \cap (A \cup B) = \emptyset$ with $D \gg B$ and $D \not\gg A$
$A \mathrel{\triangleright} B$	$A \gg B$ or $A \mathrel{\triangleright_D} B$
$\omega \succ_\sigma B$ $(\omega \in \Omega)$	$\{\omega\} \gg B$, that is $\sigma(B \cup \{\omega\}) = \{\omega\} \cap B^c$
$B \succeq_\sigma \omega$	$\omega \not\succ_\sigma B$, that is $\sigma(B \cup \{\omega\} \neq \{\omega\} \cap B^c$
$A \succcurlyeq^*_\sigma B$	$(A \succ_\sigma B$ or $A \succeq_\sigma B)$. Equivalently: either $[A$ is a $\mathfrak{A}$-atom and $A \gg B]$ or $[B$ is a $\mathfrak{A}$-atom and $B \not\gg A]$

When $\mathfrak{A} = \mathcal{P}(\Omega)$, we have

$$\succcurlyeq^*_\sigma := \big\{(\{\omega\}, X) \,\big|\, \omega \succ_\sigma X\big\} \cup \big\{(X, \{\omega\}) \,\big|\, \omega \not\succ_\sigma X\big\}$$

Order relations in $\mathcal{L}_{\mathsf{KBS}}$		
Abbreviation	Definition	Semantic property
(none)	$\mathsf{B}^{\varphi \vee \psi} \neg\psi$	$[\![\varphi]\!] \gg [\![\psi]\!]$
$\varphi \mathrel{\triangleright_\gamma} \psi$	$\mathsf{K}\big(\gamma \to \neg(\varphi \vee \psi)\big) \wedge \neg \mathsf{B}^{\gamma \vee \varphi} \neg\varphi \wedge \mathsf{B}^{\gamma \vee \psi} \neg\psi$	$[\![\varphi]\!] \mathrel{\triangleright_{[\![\gamma]\!]}} [\![\psi]\!]$
$\varphi \mathrel{\triangleright} \psi$	$\big(\mathsf{B}^{\varphi \vee \psi} \neg\psi\big) \vee \varphi \mathrel{\triangleright_\gamma} \psi$	$[\![\varphi]\!] \mathrel{\triangleright} [\![\psi]\!]$