Belief and Counterfactuals.
A Study in Means-End Philosophy

Franz Huber

Under contract with Oxford University Press.

Please do not cite, quote, or distribute without explicit permission of the author!
# Contents

1 **Introduction** ........................................... 1

2 **Belief First** ........................................ 17  
   2.1 Ideal doxastic agents .............................. 17  
   2.2 Belief and ends ................................. 20  
   2.3 Conditional belief and belief revision ......... 22

3 **Belief Revision** ...................................... 29  
   3.1 The AGM theory of belief revision .............. 29  
   3.2 Systems of spheres ............................... 37  
   3.3 Iterated belief revision ......................... 41

4 **Conditional Belief** ............................... 51  
   4.1 Ranking theory: static rules .................... 51  
   4.2 Ranking theory: dynamic rules ................. 58  
   4.3 Iterated belief revision revisited ............. 65

5 **Why Should I?** ...................................... 75  
   5.1 The consistency argument ....................... 75  
   5.2 The Consistency Argument Continued .......... 83  
   5.3 The Consistency Argument Concluded .......... 92  
   5.4 Hypothetical imperatives ....................... 97  
   5.5 Conditional obligation and conditional belief 103  
   5.6 Proofs ........................................... 109

6 **Applications in Epistemology** .................. 121  
   6.1 Conceptual belief change and logical learning 121  
   6.2 Learning indicative conditionals .............. 127
6.3 In defense of rigidity .................. 133
Acknowledgments

Chapters 3 and 4 rely on and, with permission of Wiley, reuse material from Huber (2013b; c).
Chapter 1

Introduction

This book is about belief, counterfactuals, and their relation. Belief is central to epistemology. It occurs when an agent such as a person or computer program holds something to be true. Counterfactuals figure prominently in metaphysics. They are if-then claims, or conditionals, about what would have been the case if certain conditions had obtained. The aspect of their relation that I will focus on is a question in the epistemology of metaphysics.

In the way I will engage with it, epistemology is a normative discipline. I will study how agents should believe, as opposed to what they do believe. More precisely, I will study how agents should believe conditional on what they do believe – and given they have certain ends such as holding true and informative beliefs. In the way I will engage with it, metaphysics is subordinate to epistemology in the sense that metaphysical theses are necessary conditions for the satisfiability of epistemological norms. Given an instrumentalist understanding of normativity, this in turn means that metaphysical theses are necessary conditions for the possibility of attaining certain ends. How exactly this “transcendental” metaphysics is supposed to work is a complicated matter. I will try to explain it below.

Broadly speaking we can distinguish four approaches to understanding belief. Representationalists such as Fodor (1975) and Millikan (1984) claim that holding a belief consists in the possession of a mental representation with a propositional content. Dispositionalists and their predecessors, behaviorists, claim that holding a belief consists in certain acts or dispositions to act, such as the disposition to act as if the propositional content of the belief was true. A particular version thereof is interpretationism (Davidson 1984). According to it an agent holds a belief just in case a charitable interpretation of the agent’s acts or dispositions to act requires attributing this belief to the agent.
CHAPTER 1. INTRODUCTION

Each of these views can be combined with a functionalist understanding of belief, but functionalism can also be defended on its own. It says that holding a belief consists in the agent’s mental state standing in the right causal relations to sensory stimuli, acts or dispositions to act, and other mental states such as desires. While considering beliefs to have propositional contents is not essential to all of these positions, they all are compatible with doing so. The same is true of primitivism, which postulates belief as a primitive mental state that does not admit of further analysis. These propositional contents have a certain “conceptual structure.” The details of this structure will depend on one’s view of propositions (Stalnaker 1984), but minimally it will comprise the structure of set theory and logic. It is this minimal conceptual structure that the present project requires.

An approach to understanding belief this book rejects is eliminativism (Churchland 1981). The latter flatly denies the existence of beliefs. On this view beliefs are a convenient fiction we ascribe to ourselves and others in our everyday conception of the mind, our “folk psychology.” However, so this view says, the mature science that will eventually replace this pre-scientific view of the mind will hold that there are no beliefs. Every other position that gets rid of belief – for instance, by replacing it with probabilistic degree of belief (Jeffrey 1970) – is also rejected by this book. It is an assumption of the present project that there are beliefs, and that an agent can have ends that are attained if and only if she holds certain beliefs (whose propositional contents may, but need not be required to be true).

There are at least two sorts of conditionals (Adams 1970, Briggs forthcoming): indicative conditionals and counterfactual, or subjunctive, conditionals, (although not everybody is willing to identify counterfactual and subjunctive conditionals; see Bennett 2003). Each can be approached in at least two ways. On one view conditionals express propositions that are true or false (Stalnaker 1968, Lewis 1973a). On another view conditionals do not express propositions that are true or false, and do not have truth values. Instead, they express the uttering or thinking agent’s internal state of mind rather than a proposition about the external world. The latter “expressivist” view is prominent for indicative conditionals (Adams 1975, Edgington 1995), but some also adopt it for counterfactual conditionals, or counterfactuals (Edgington 2008 as well as Spohn 2013; 2015 whose account has it that counterfactuals express propositions relative to the agent’s conditional beliefs and a partition). In this book I will be concerned with counterfactuals. Indicative conditionals will play a role only insofar as they express conditional beliefs. While I think this is the right view of indicative conditionals, and I will assume it throughout this book, not much hinges on this assumption: if it was false, I just would not say much about indicative conditionals.
In contrast to this, much hinges on the assumption that counterfactuals express propositions that are true or false. The present project rejects the view that counterfactuals do not have have truth values. However, I hasten to add that this assumption does not bring with it what is known as “modal realism” (Lewis 1986a).

A factual or non-modal language allows one to say that something is the case: it is raining; the streets are wet. A modal language or system of representation presupposes a factual language and additionally allows one to say how something is the case: it could be raining; if it was, the streets would be wet. Of course, what is a factual claim in one language may be a modal claim in another language. For each factual language there is at most one linguistic, conceptual, or representational entity that accurately and maximally specifically – i.e., as completely as the factual language allows – describes, conceptualizes, or represents reality. I will call this entity the actual factual “world” for the factual language under consideration.

The actual factual world is not real. To use a dangerously loaded term, it is an idea, a mind-dependent construct that is similar to a state description in Carnap (1947a)’s sense. Besides the actual factual world there are many merely possible factual worlds, i.e. descriptions or representations that maximally specifically, but inaccurately represent reality. These factual worlds include every description or representation the factual language or system of representation allows for. (Which these are is itself something that needs to be learned, and we will see in chapter 6 how to do so.) Factual worlds give rise to factual propositions which we can formally represent as sets of factual worlds. So much for a factual language.

In a modal language we can say more than in a factual language. In addition to being able to say that it is not raining, we can say that it could have been raining, and that, if it had been, the streets would have been wet. Modal propositions can be formally represented as sets of modal worlds. The latter consist of two components: a factual component, which is a factual world, and a modal component that specifies what could and would have been in this factual world. Importantly, the modal component does not specify what could and would have been in reality. Instead, it specifies what could and would have been in some factual world, which may or may not be the actual factual world. This means that modal propositions are not about reality, but about ideas. Furthermore, for any factual world there is at most one modal component that accurately and maximally specifically describes what could and would have been in this factual world. This modal component is determined by the factual world and the language or system of representation. Reality has a say in this only insofar as she has a say in which factual world is actual, and which language one is speaking.
CHAPTER 1. INTRODUCTION

On this view factual and modal worlds are relative to a language or some other form of representation. They are not real in any language, thought, or representation-independent sense. We talk and think about, and conceptualize and represent reality in terms of what is and what is not, in terms of what could be, and in terms of what would have been. And we do so because we find it useful. Yet these *nons* and *coulde*s and *woulde*s are not part of reality. They belong to the language or system of representation we use to describe or represent reality, and thus to the realm of the mental. As her name suggests, the only thing that is really real is reality.

Suppose it is neither raining nor snowing, but that it could have been raining, and that the streets would have been wet if it had been. On the present view there is nothing in reality that is described by, corresponds to, or makes true these *nons* and *ors* and *ands* and *coulde*s and *woulde*s. They are tied to our representation of reality in thought and language – and needless to say, we cannot think, let alone talk about reality without some representation. Just as thinking and talking about reality are dependent on a language or system of representation, so is truth. This is why these claims and thoughts can have truth values without there being anything in reality that makes them true. What is true and what is not true, and what could be true and what would have been true depend on reality because the actual factual world depends on reality. Yet what is true also depends on the language or system of representation that these propositions belong to. Counterfactuals and other modal claims can express truths, and not merely states of mind such as conditional beliefs, without there being any modal reality that makes them true. The reason is that they can be understood as claims about ideas.

In sum, the present view which I will call *idealism about alethic modality*, i.e. the modality that pertains to what is true, is a third option between the realist and expressivist views. Like the expressivist, the modal idealist does not locate the (alethic) modalities in reality, but in the mind. Like the realist, the modal idealist does not interpret the modalities as expressing states of mind, but as expressing propositions that are true or false. Of course, what counts as a proposition is radically different on the realist and the idealist account. Modalities are *ideas*. However, they are not ideas with reality. They are *our* ideas. We conceptualize reality in terms of *coulde*s and *woulde*s, and we do so for the exact same reason that we conceptualize reality in terms of *nons* and *ors* and *ands*: because we find it useful. Different beings who also think or talk about reality may conceptualize or represent it in different terms or ideas – say, because they have different abilities and limitations, or because they have different ends. There is no right or wrong here, just a more or less useful for various purposes.
I will argue in chapter 5 that the same is true of deontic modality that expresses what should or ought to be. Rather than corresponding to some deontic reality it reduces to what we may desire and which means-end relationships obtain. What is real are our desires and beliefs. More precisely, what is real is our having desires and our holding beliefs. Their propositional contents are not.

Counterfactuals are about what would be, or would have been, the case if certain conditions – conditions that may well be contrary to fact – obtained, or had obtained. An agent believes a proposition just in case she holds it to be true (which does not imply that she has, or ought to have, the end of holding true beliefs). An agent believes a proposition conditional on another (or the same) proposition just in case she holds the former proposition to be true conditional on the assumption that the latter is. Such a conditional belief must be distinguished from the agent’s belief in the corresponding counterfactual (Leitgeb 2007). There are, however, circumstances in which a conditional belief is related to the belief in the corresponding counterfactual. This is so because some counterfactuals imply singular default conditionals, and because conditional beliefs are closely related to the beliefs in the corresponding default conditionals.

A default conditional is an if-then claim about what is usually, normally, or typically the case if certain conditions obtain, where normality is understood in a purely descriptive sense (Bear & Knobe 2017). If Tweety is a bird, it can normally fly. This means Tweety can fly in the most normal worlds in which it is a bird. It is more normal for Tweety to be a bird and be able to fly than for it to be a bird and not be able to fly. Often default conditionals are inferred from generic default rules such as that birds can normally fly, but penguins cannot. In this case the default conditional that Tweety can normally fly if it is a bird can only be inferred if one has no information about Tweety that contradicts the claim that it can fly. For instance, one cannot infer that it is more normal for Tweety-the-penguin to be a bird and be able to fly than for it to be a bird and not be able to fly.

Similarly, the generic default rule that presidents normally do not tweet official announcements does not allow one to conclude that it is more normal for Donald Trump to be president and not tweet than it is for him to be president and tweet – just as the generic information that, statistically speaking, presidents are likely male does not allow one to conclude that the first female president is likely male. Generic default rules and statistical information are formulated in terms of generic variables that are defined on some population of individuals; whereas, default conditionals and claims about single-case probabilities are formulated in terms of singular variables (Huber 2018: ch. 10). The question under what conditions the former license inferences to the latter is a variant of the reference class problem.
A default conditional $\alpha \Rightarrow \gamma$ is true in the actual world if and only if $\gamma$ is true in all possible worlds in which $\alpha$ is true and which are most normal from the point of view of the actual world. We will see below that there are several counterfactuals. Some of them (but not all, as these counterfactuals may conflict with each other) imply the corresponding default conditionals. To see why, let $\alpha \rightarrow \gamma$ be a counterfactual of this kind. $\alpha \rightarrow \gamma$ is true in the actual world if and only if $\gamma$ is true in all possible worlds in which $\alpha$ is true and which are most normal or – if the actual world is itself less normal than the most normal worlds in which $\alpha$ is true – if $\gamma$ is true in all possible worlds in which $\alpha$ is true and which are at least as normal as the actual world. The counterfactual $\alpha \rightarrow \gamma$ thus says that, normally – and even if things are not normal, as long as they are not less normal than the way things actually are – if $\alpha$ is true, then so is $\gamma$.

The actual world may, but need not be among the possible worlds which are most normal from the point of view of the actual world. If it is, the counterfactual $\alpha \rightarrow \gamma$ is true in the actual world if and only if the default conditional $\alpha \Rightarrow \gamma$ is. Otherwise the counterfactual may be false while the default conditional is true. However, the converse case cannot occur for a counterfactual of this kind.

Ida is certain that it is more normal for it to be sunny on Wednesday and her to have lunch in the park than for it to be sunny that day and her not to have lunch in the park. This may be because she is certain that she would have lunch in the park if it was sunny on Wednesday. In this situation Ida should believe that she will have lunch in the park conditional on the assumption that it is sunny on Wednesday. When an agent is certain of a default conditional, but no other information, she should believe its consequent, or then-part, conditional on the assumption that its antecedent, or if-part, is true. This normative principle relating belief in a default conditional and conditional belief is an instance of the royal rule. Since some counterfactuals imply default conditionals, and one may believe a default conditional because one believes the corresponding counterfactual, the royal rule also connects belief in some counterfactuals and conditional belief. This connection will be explored and exploited in chapters 7 and 10.

The plan for the remainder of the book is follows. In chapter 2 I will first discuss which agents I am focusing on, and which cognitive ends I am assuming them to have. Then I will describe how this relates to conditional beliefs and belief revision. Chapter 3 will first present the AGM theory of belief revision (Alchourrón & Gärdenfors & Makinson 1985). Then I will focus on the problem of iterated belief revision. In chapter 4 I will show how this problem finds a solution in ranking theory, which was introduced in Spohn (1988) and is most comprehensively discussed in Spohn (2012).
Chapter 5 will first answer the question why conditional beliefs should obey the ranking calculus. Then I will discuss the underlying view of normativity in epistemology. I will conclude with a note on the logic of conditional obligations, which is identical to the logic of conditional beliefs. In chapter 6 I will consider two small applications of ranking theory to problems in epistemology and the philosophy of science. The first explains how concepts can be learned in ranking theory. This includes logical learning as a prominent special case. The second explains how conditional information as conveyed by indicative conditionals can be learned. These applications illustrate how ranking theory can be fruitfully applied to tackle philosophical problems that have proven difficult for Bayesianism. I will conclude this chapter by dissolving a worry raised by Weisberg (2015).

In chapter 7 I will turn to the logic of counterfactuals and try to explain how I think we can and should engage with some philosophical problems. Central to this view of how to philosophize are an instrumental understanding of normativity, or rationality, according to which one ought to take the means to one’s ends, and the use of formal methods in establishing means-end relationships. The view is motivated by a deep mistrust of intuitions. While I may not always be able to hide my frustrations with some of the more speculative versions of intuition-based philosophy, my aim is a constructive one: to point out one way of engaging with some philosophical problems that does not overly rely on intuitions.

Philosophers interested in counterfactuals often consider it decisive which counterfactuals intuitively seem to be true. However, subjective intuitions vary across philosophers and within philosophers across time (Knobe & Nichols 2008). I want to supplement this intuition-based methodology with what may be called a “principled” account of the logic of counterfactuals. More specifically, I will propose a normative principle, viz. the royal rule, that relates default conditionals – and, thus, some counterfactuals – to conditional beliefs. The general idea behind this principle is that, in the absence of further information, alethic modalities constrain or guide doxastic modalities which pertain to belief. In the present context this principle says that an agent should believe a proposition $C$ given that she is certain of the proposition $A$ and the default conditional that, if $A$, then normally $C$, but, importantly, no other information. The idea is that, absent further information, default conditionals – and, thus, some counterfactuals – constrain or guide conditional beliefs. When not restricted to the present context where I assume the truth conditions of counterfactuals and default conditionals to be as stated above, the royal rule says that one ought to disbelieve a proposition given the assumption that it is, in a purely descriptive sense, abnormal for this proposition to be true, but no further information.
The royal rule is a qualitative version of Lewis’ (1980) “principal principle” which relates chances and degrees of belief. The latter principle says that an agent’s initial degree of belief in a proposition \( C \) ought to be equal to \( x \) given that the chance equals \( x \) that \( C \) is true and, perhaps, further “admissible” information, but no inadmissible information. Given some assumptions about what information is admissible, the principal principle entails that chances are what an agent’s initial conditional degrees of belief should be. Now, initial degrees of belief and, hence, initial conditional degrees of belief ought to obey the probability calculus – witness, for instance, the Dutch Book argument due to de Finetti (1937) and Ramsey (1926). Therefore chances do so as well.

Thus probabilism, i.e. the thesis that degrees of belief ought to obey the probability calculus, and the principal principle have a consequence that is about chances – namely that chances are probabilities. While this claim is presumably also in agreement with our subjective intuitions, there is no need to appeal to the latter in order to defend this claim. Probabilism and the principal principle do this for us. This illustrates how two normative principles from epistemology can entail a metaphysical thesis.

Suppose that, in addition, we can justify both probabilism and the principal principle. According to instrumentalism, what one ought to do is take the means to one’s ends. Thus, to justify a normative principle is to show it to be a means to attaining some end one may have. Probabilism can perhaps be justified by the Dutch Book argument, and the principal principle can perhaps be justified in some other way (Pettigrew 2013). In this case the thesis that chances are probabilities is a consequence of probabilism and the principal principle which in turn can be justified by being shown to be means to attaining ends one may have. Intuitions are certainly useful as a heuristics in arriving at these normative principles, and in considering various metaphysical theses. However, at no point does one have to appeal to intuitions in order to defend the metaphysical thesis.

The upshot of this way of engaging with some philosophical problems – of means-end philosophy – is the following. The metaphysical thesis that chances are probabilities is a necessary condition for the possibility of attaining certain ends one may have. Given that one has these ends, one ought to satisfy those norms. Yet one can satisfy those norms only if things are a certain way. Means-end philosophy thus tells one what metaphysical theses one is committed to by pursuing various ends.

\(^1\)Here it is tricky to rely on Joyce (1998; 2009)’s gradational accuracy argument, as Joyce (2009: 279) appeals to the principal principle in defense of his assumptions about inaccuracy.
In the same way I want to use the royal rule and the thesis that beliefs ought to obey the ranking calculus to derive some properties of descriptive normality. Given the truth conditions stated above, these properties will determine some of the logical principles satisfied by default conditionals and counterfactuals. These principles are expected to approximate the logical principles philosophers have proposed on the basis of subjective intuitions. However, we do not have to rely on those subjective intuitions in order to support these logical principles. Instead, we obtain them as consequences of two normative principles from epistemology and an assumption about the truth conditions of default conditionals and counterfactuals. This is the sense in which the account of the logic of counterfactuals will be principled. Of course, to be convinced by this means-end argument one needs to accept the assumptions made as well as pursue the ends the normative principles are means to attaining. Those who do not are given information about a means-end relationship for which they may have little use.

In order to carry out this argument in detail the following ingredients are needed. First we need a theory of conditional beliefs. This will be Spohn (1988; 2012)’s theory of ranking functions presented in chapters 3 and 4. Second we need a precise formulation of the normative principle relating some counterfactuals via default conditionals to conditional beliefs. We will get this in chapter 7. Third we need an argument analogous to the Dutch Book argument establishing the thesis that conditional beliefs should obey the ranking calculus. This will be the consistency argument from section 5.1. In addition the very principle relating some counterfactuals via default conditionals to conditional beliefs, the royal rule, needs to be justified as well by being shown to be a means to attaining an end one may have. This will be attempted in section 8.1.

In the remaining sections of chapter 8 I will consider two small applications of the resulting theory to problems in metaphysics and the philosophy of science. The first application concerns Lewis’ (1973b) definition of causation in terms of counterfactuals, which turns out to be a special case of Spohn (2006a)’s definition of causation in terms of ranking functions modulo the interpretation of the latter. The second application concerns a problem for Lewis’ (1979) theory of counterfactuals that arises from an application of Arrow (1951)’s impossibility theorem from social choice theory. I have first learned of this problem from Thomas Kroedel in Konstanz in July 2009. It is also raised in Morreau (2010). The relevant section 8.3 relies on joint work with Thomas Kroedel (Kroedel & Huber 2013). These applications illustrate how ranking theory can be fruitfully applied to tackle philosophical problems that haven proven difficult for the similarity account of counterfactuals (Stalnaker 1968, Lewis 1973a).
Chapter 9 attempts to show that the rank-theoretic normality account of counterfactuals is better suited for theorizing about causality than both the similarity account of counterfactuals as well as the structural equations framework (Spirtes & Glymour & Scheines 2000, Pearl 2009). As mentioned above, counterfactuals are claims such as the following: if Ida had not had coffee in the morning, she would have been tired at noon. They are about what would have been the case (Ida would have been tired at noon), if certain conditions had obtained (if Ida had not had coffee in the morning). As suggested by the term *counterfactuals*, these conditions may well be contrary to fact – in fact Ida had coffee in the morning. Causal claims are claims about one property or event being an effect of another property or event. Ida’s being alert at noon is an effect of her having coffee in the morning. In other words, her having coffee in the morning causes, or brings about, that she is alert at noon.

Counterfactuals are closely related to causality (Collins & Hall & Paul 2004, Paul & Hall 2013). A causal claim is often (e.g. in Lewis 1973b; 1979; 2000) said to be shorthand for a more complicated claim involving a specific counterfactual. The causal claim that Ida’s being alert at noon is an effect of her having coffee in the morning is often said to be shorthand for, or at least closely related to, the following three claims essentially involving a specific counterfactual. First, Ida had coffee in the morning. Second, Ida was alert at noon. After all, only properties that are instantiated or events that take place can be causes or effects. Third, the “causal” counterfactual: if Ida had not had coffee in the morning, she would have been tired at noon.

The causal counterfactual is “forward-looking.” Its antecedent is about the potential cause: whether or not Ida had coffee. Its consequent is about one of its alleged effects: whether or not Ida was alert. These forward-looking, causal counterfactuals are to be used in the study of causality. However, there are other counterfactuals that relate to causality in a different and sometimes opposing way, and still others that do not relate to causality at all. For instance, you may wonder whether Ida had coffee. You figure: if Ida had not been alert at noon, she would not have had coffee in the morning. That is, you reason from the absence of one of the alleged effects back to the absence of the potential cause. Causal counterfactuals reason in the opposite direction from the potential cause forward to one of its alleged effects. They also hold fixed what is actually the case: even if Ida had not been alert at noon, she would still have had coffee in the morning. This causal counterfactual flatly contradicts the previous counterfactual. Moreover, it is also true that if Ida had not drunk anything, she would, trivially, also not have had coffee. However, this counterfactual is not related to causality at all.
In the literature the former non-causal counterfactuals are known as “back-tracking” counterfactuals, and the latter are known as “spurious” counterfactuals (Menzies 2008). The question is where to draw the line between causal counterfactuals and backtracking and other non-causal counterfactuals (Woodward 2003).

The state of the art representation of causal counterfactuals are structural equations (Haavelmo 1943, Halpern & Pearl 2005a; b, Pearl 2009, Spirtes & Glymour & Scheines 2000). Importantly, though, structural equations presuppose rather than provide an answer to the question where to draw the line between causal and non-causal counterfactuals. They also do not capture all aspects of causation (Hiddleston 2005). The latter problem has led to the development of so-called “extended causal models” (Halpern 2008; 2016 and Halpern & Hitchcock 2010; 2013). These contain two elements representing two seemingly distinct modalities. The first element are structural equations which represent the “(causal) laws” of the model. The second element is a ranking function (or, in later versions, an ordering relation) which represents normality.

One goal of Chapter 9 is to show that these two modalities can be unified into one modality by adopting the theory of counterfactuals from chapter 7. It is to be noted, though, that normality in extended causal models is understood to include non-descriptive, normative elements (Hitchcock & Knobe 2009) which are explicitly excluded from the way normality is understood here. The unification will be achieved by formulating two constraints under which extended causal models with their two modalities can be subsumed under “counterfactual models” which contain just the one modality of descriptive normality.

The two constraints turn out to be formally precise versions of Lewis (1979)’s “system of weights or priorities” that governs overall similarity between possible worlds. This system is Lewis (1979)’s answer to the question where to draw the line between causal and non-causal counterfactuals. It appeals to the model-independent notion of a “law of nature.” The two constraints appeal to the model-relative notion of a “necessarily true default conditional.” The latter differs from the notion of a law of nature (Woodward 2003: ch. 6) and corresponds to what is represented by a structural equation. Menzies (2004) argues that such model-relativity is unavoidable. If so, the two model-relative constraints might be viewed as an answer to the question where to draw the line between causal and non-causal counterfactuals. However, without a means-end argument for this claim I can only refrain from making it. Both Lewis (1979)’s answer as well as this one locate the difference between causal and non-causal counterfactuals not in their truth-conditions, but in what is held fixed in determining overall similarity and descriptive normality, respectively. I will assume that this much is correct.
Chapter 10 will bring together the view of conditional beliefs developed in chapters 3-6 and the view of counterfactuals developed in chapters 7-9 by answering the question what one should believe about what would have been the case. This will be done by considering under what conditions default conditionals and counterfactuals can be tested “empirically.”

Once I tell you what Ida had for breakfast Monday through Thursday, and whether she was tired at noon on these days, you will have no difficulty inferring that Ida would have been tired at noon on Friday, if she had not had coffee on Friday morning. My interest lies in finding out exactly when one ought to make these inferences by stating the conditions under which one can reliably infer the truth values of counterfactuals from “empirically accessible” information. I will do so by stating a very simple theorem. A philosophically important implication of this theorem is that it allows us to justify inferences from empirically accessible information to counterfactuals by showing these inferences to be means to attaining ends one may have. A technologically important implication of this theorem is that it allows us to program computers to possess this inferential ability.

The fundamental theorem underlying statistical inference is the Strong Law of Large Numbers, SLLN. This “law” relates relative frequencies – which may, in general, be empirically accessible or “observable” – and probabilities. As a theorem the SLLN holds for every interpretation of probability. Therefore we can interpret the probabilities as chances. The latter go beyond what is observable in much the same way that counterfactuals do: both are alethically modal notions. Given this interpretation of the probabilities, the SLLN says that the chance is maximal that the observable relative frequencies converge to the chances which are, in general, empirically inaccessible or unobservable; and that the observable means (i.e. averages) converge to the unobservable “expected” values (these “expected” values are defined in terms of chances, so the epistemological connotations of this notion should be ignored).

Relative frequencies and means are the notions scientists use to describe data in a probabilistic manner in order to make statistical inferences based on the SLLN. Scientists also use the notion of the modes of a sample to describe data. The modes of a sample are defined as the most frequent outcomes in the sample. Modes have a wider applicability than means because the former, in contrast to the latter, make sense even when the data cannot be described numerically. While modes play an important role in descriptive statistics, modes do not play a role anymore in inferential statistics. The reason is that modes, in contrast to relative frequencies and means, do not behave like probabilities. Therefore the SLLN does not apply.
The thesis of chapter 10 is that modes cannot only be used to describe data, but can equally well be used in inference. Modes provide the empirically accessible information that entitles us to infer the truth values of empirically inaccessible default conditionals and counterfactuals. They do so in much the same way that relative frequencies and means provide the empirically accessible information that entitles us to infer the empirically inaccessible chances and “expected” values.

In order to support this thesis I will present a qualitative version of the SLLN. The first challenge is to find a precise formulation of the theorem to be proven and, in particular, of the qualitative notion of a probability measure. By now it will clear that this is a ranking function. Just as probabilities can be interpreted, among others, in a doxastic sense as conditional degrees of belief, in an alethic or metaphysical sense as chances, and in an empirical sense as relative frequencies, ranking functions can also be interpreted in at least three ways. Subjective ranking functions represent conditional beliefs. Alethic ranking functions represent descriptive normality and – via the assumption of their truth conditions – default conditionals and counterfactuals. In addition there are the empirical notions of absolute and relative failures that generalize the notion of the modes of a sample.

In the probabilistic case the observable information reported in the form of relative frequencies in principle allows one to “find out” (in a loose sense made precise below) exactly what the chances are. In the present rank-theoretic case the situation is different. The empirically accessible information is not rich enough to allow one to find out exactly what the alethic ranks are, not even in principle (and not even in a loose sense). One reason for this is that ranks are at best measured on a ratio scale, whereas probabilities are measured on an absolute scale. Something slightly weaker is true, though. The empirically accessible information reported in the form of absolute failures allows one to “find out” (in a slightly stricter sense made precise below as well) the features of alethic ranks that determine the truth values of non-iterated default conditionals and counterfactuals. (This works by, among others, initially believing all contingent default conditionals to be false.)

The situation in the rank-theoretic case is different in two further respects. First, the assumptions needed in order for the theorem to apply are weaker than the corresponding assumptions in the probabilistic case. In the latter case the experiments need to be designed such that the individual trials are independent and identically distributed in the sense of the chance measure. The assumptions in the present case are weaker given minimal assumptions about the relationship between chances and default conditionals. Second, the sense in which the new theorem allows one to find out the truth values of default conditionals and counterfactuals is stronger than the corresponding sense in the probabilistic case.
In the probabilistic case the sense in which the observable information allows one to infer what the unobservable chances and “expected” values are is that the former converge to the latter with maximal chance. In the rank-theoretic case the empirically accessible information allows one to infer the truth values of empirically inaccessible default conditionals and counterfactuals in the sense that the former stabilize on the latter with alethic or metaphysical necessity. Convergence means that the conjectured values get closer and closer to the correct value, which is compatible with them never actually reaching the correct value. Stabilization means that the conjectured values get it exactly right after finitely many steps, and continue to do so forever after. Stabilization implies, but is not implied by, convergence. Furthermore, given minimal assumptions about the relationship between chances and default conditionals, stabilization with alethic or metaphysical necessity implies, but is not implied by, convergence with maximal chance.

In the concluding chapter 11 I will offer a different perspective on some of the results of this book. Chapters 3 and 4 discuss the doxastic modalities of belief and conditional belief. Chapter 5 discusses the deontic modalities of obligation and conditional obligation. Chapters 7-9 turn to the alethic or metaphysical modalities of descriptive normality, necessity, and counterfactuality. Finally, chapter 10 adds the empirical modalities of absolute and relative failures. Combining all these modalities allows me to study what one should believe about what would have been the case given information about how often things have failed to occur.

As explained, except for our having wants and needs and our holding beliefs and conditional beliefs, the first three modalities have no reality in any substantial sense that is independent of our talking and thinking about reality, our ways of conceptualizing reality. Perhaps surprisingly, the same is true for the empirical modalities of absolute and relative failures. The latter report how often something occurs, and fails to occur, and this depends on some representation that allows us to identify something as an event (or object or individual) as well as to count several events (or objects or individuals) as instantiating one type.

Yet just as there aren’t any coulds and shoulds and woulds in reality, or any nots and ands and ors, there isn’t any many there either. Whether something exists may not depend on the choice of a language, or system of representation. However, what and who exists depends on a language, or system of representation. You may see four people entering a room, I just see one group. You may say that I had four dinners so far this week, I say that I had pizza on Monday and pasta on Tuesday and fajitas on Wednesday and curry on Thursday. I see four modalities of the same kind, you may see one big mess. Fortunately there is no right or wrong here, just a more or less useful for various purposes.
Finally, throughout the book we will be accompanied by the *ideal doxastic agent* Ida, named after my first philosophy teacher. Ida likes to travel the capitals of the world and cares a lot about the weather. Ida usually has coffee in the morning, and is tired at noon if she does not. She has lunch in the park if the weather is nice, and enjoys wine in the evening whenever possible.
Chapter 2

Belief First

In this chapter I will first discuss which agents I am focusing on, and which ends I am assuming them to have. Then I will briefly describe how this relates to conditional belief and belief revision. I rely on Huber (2013a).

2.1 Ideal doxastic agents

There are at least two senses of ‘should,’ or ‘ought.’ In general, what we can call the wide sense of ‘ought’ expresses what one values. Accordingly, Schroeder (2011) calls it the evaluative ‘ought.’ On the instrumentalist view adopted in this book, the wide sense of ‘ought’ expresses someone’s ends. On this view, the difference between ‘ought’ and ‘must’ in turn is the difference between a mere want and a genuine need. Importantly, every proposition can be in the scope of this wide sense of ‘ought.’ On the instrumentalist view this is so because there are no constraints on what ends one may have.

The difference between the wide and narrow sense can perhaps be explained by reference to classical decision theory (Savage 1954). The latter distinguishes between states of the world, outcomes, and actions which are characterized as functions from states of the world to outcomes. The states of the world are the objects of the decision maker’s beliefs, and these beliefs of hers are represented by a subjective probability measure. The outcomes are the objects of her desires, and these desires of hers are represented by a utility or value function. The actions are the alternatives from which the decision maker can and has to choose. In this framework, the propositions in the scope of the wide sense of ‘ought’ are the outcomes that form the domain of the decision maker’s utility or value function.
The wide sense of ‘ought’ is to be distinguished from what we can call its narrow sense, Schroeder (2011)’s deliberative ‘ought.’ The latter expresses what one should, or ought to, do and applies only to actions. More specifically, the narrow sense of ‘ought’ applies only to the intention to take an action. On the instrumentalist view this makes sense because the intention to take an action is a means to attaining the end of actually taking the action. In classical decision theory, what is in the scope of the narrow sense of ‘ought’ are the actions that form the domain of the decision maker’s expected utility or value function. The latter is formed by combining the decision maker’s subjective probability measure and her utility or value function.

The wide and narrow sense of ‘ought’ can come into seeming conflict with each other. One may have ends and beliefs to the effect that, in the wide sense, the poorest member of society should win the lottery, but that, in the narrow sense, she should not buy a lottery ticket. This may be so because one may not hold the conditional belief that the poorest member will win the lottery if she buys a ticket. Furthermore, the principle that ‘ought’ implies ‘can’ – that one cannot be required to do something that is not within one’s reach – applies only to the narrow sense of ‘ought,’ but not also to its wide sense. This principle holds for intentions to take actions, but in general fails to hold for actions themselves.

I am interested in how an ideal doxastic agent should organize her beliefs, and how she should revise her beliefs when she receives new information. Here ‘should’ is understood in its narrow sense. Thus I am interested in how an ideal doxastic agent should intend to organize and revise her beliefs. These intentions to believe are what I am assuming to be under the ideal doxastic agent’s control, as required by the principle that ‘ought’ implies ‘can.’ As a consequence, issues of doxastic involuntarism, i.e. the alleged inability to form beliefs at will, do not arise. For intentions – or willings – to believe can be formed at will, even if beliefs themselves cannot be so formed.

I call the agent a doxastic agent, as I am focusing on what she should do qua believer. What the agent should do qua believer is to hold certain beliefs, to refrain from holding other beliefs, and to revise her beliefs in certain ways if she receives new information. All of these requirements to believe make sense for the narrow sense of ‘ought’ because believing is an action. More specifically, believing is a cognitive action that is located in the agent’s internal world or mind. For this reason we can also call her a cognitive agent. In contrast to this it does not make sense to demand of an agent that she should, in the narrow sense of this term, hold certain knowledge – other than by demanding of her to hold certain beliefs. Therefore I am not calling the agent an epistemic agent.
Knowledge does not figure in norms, or requirements, in the way belief does. This is so because knowing is not an action, cognitive or otherwise, but only the result of other actions and events. Some beliefs may also be held solely because they are the result of other actions or events – say, by being caused by experience. However, how to revise one’s beliefs once those experientially or otherwise caused beliefs are held is a question of which cognitive action to take. Thus, to the extent that epistemology is a normative discipline that studies what cognitive actions to take, it studies what and how to believe, not what or how to know.

In the narrow sense of ‘ought’ we can require agents only to (intend to) take actions. Believing is, but knowing is not an action in this sense. We can require Ida to believe that Vienna is the capital of Austria, and that Athens is not the capital of Greece. However, we cannot require Ida to know that Vienna is the capital of Austria, let alone that Athens is not the capital of Greece. Similarly, looking and listening are actions, but seeing and hearing are not. We can require Ida to look whether it is sunny, and to listen whether the birds are singing. We cannot require her to see that it is sunny, or to hear that the birds are singing.

For the same reason knowledge cannot figure in the main argument place – the consequent – of a conditional obligation: we cannot require Ida to know that Vienna is the capital of Austria given that she believes it to be, nor can we require her to not know that Athens is the capital of Greece given that she does not believe it to be. However, knowledge may figure as a condition in a conditional obligation: we can require Ida to believe that Vienna is the capital of Austria given that she knows it to be, and we can require her to not believe that Athens is the capital of Greece given that she does not know it to be. I hasten to add that I do not subscribe to these requirements, though, especially not the latter. This is so because I am convinced that, whatever knowledge is, strictly speaking we never have any of it (although nothing in this book depends on this conviction). What we have are true beliefs, and these are all we need: our curiosity is satisfied if we hold sufficiently informative beliefs, and we get the remaining benefits from the truth of the beliefs.

The agent I am considering is not only doxastic, but also ideal in the sense that she does not suffer from any computational or other cognitive limitations, and can always identify all logical and conceptual truths. She also gets to voluntarily decide what to believe, and never forgets any of her beliefs. For such an ideal doxastic agent the actions she intends to take are the actions that she takes. Indeed, we may define a (cognitive) agent to be ideal just in case any (cognitive) action that she intends to take is an action that she takes. This means the restriction to agents that are ideal allows me to ignore the otherwise important distinction between actions an agent should intend to take, and actions she should take.
2.2 Belief and ends

Ideal doxastic agents sometimes believe that some things are the case. Sometimes they believe that some things are not the case. Then we speak of “disbelief.” Sometimes they suspend judgment about whether or not something is the case and neither believe nor disbelieve that it is the case. Suspensions of judgment are relative to a question, but otherwise differ from beliefs not in kind, but in degree: they are the beliefs that are held with the least firmness, namely none at all. In a sense, then, to suspend judgment with respect to a question is to have an opinion on the matter, namely the neutral one of not believing any answer to the question. This is different for questions for which one lacks the relevant conceptual resources to fully understand them. These are not questions about which one suspends judgment. Instead, one has literally no opinion on the matter, not even the neutral one of suspending judgment.

Ida believes that Vienna is the capital of her native Austria. Ida disbelieves that Athens is the capital of Greece. Ida suspends judgment with about whether or not Beijing is the capital of China. Suppose Ida’s only end is to have true beliefs. In this case a belief of hers is successful if, and only if, the content of this belief is true. It is unsuccessful if, and only if, the content of this belief is false. For the time being I am ignoring that it is often entire systems of beliefs, rather than individual beliefs, that are evaluated as successful or unsuccessful. This allows me to momentarily bracket questions such as whether the end of having true beliefs includes the end of having informative beliefs, and how the contents of beliefs are individuated. Given this end of hers, Ida’s belief that Vienna is the capital of her native Austria is successful because the content of this belief is true. On the other hand, her disbelief that Athens is the capital of Greece is unsuccessful. This is so because the content of her belief that Athens is not the capital of Greece is false. Finally, Ida’s suspending judgment about whether or not Beijing is the capital of China is neither successful nor unsuccessful.

A normative theory of belief, that is, a theory of rational belief, tries to capture how an ideal doxastic agent can attain her ends. I will mainly focus on the end of believing truths, of holding beliefs whose contents are true. I will mainly assume that these contents are sufficiently informative to satisfy the agent’s curiosity – say, by answering all her questions, although I do not assume that curiosity must come in the form of questions. In the way I have formulated this end, it comprises the end of disbelieving falsehoods because disbeliefs have been formulated as beliefs that something is not the case. Later, in chapter 5, we will add a clause to take into account that beliefs include suspensions of judgment.
2.2. BELIEF AND ENDS

There is a formulation of this end that says the following: an ideal doxastic agent has the end of believing the truth and of avoiding error or falsehood. In this formulation, the first part really says that an ideal doxastic agent has the end of holding beliefs whose contents are informative. The second part really says that she has the end of holding beliefs whose contents are true. This is particularly clear in the formulation given by James (1896: sct. VII; italics in the original):

*We must know the truth; and we must avoid error.*

Knowledge is generally assumed to be factive (Williamson 2000): we can know only true propositions. Suppose this is so. Then the first part of “our first and great commandments as would-be knowers” requires us to know many or informative propositions, rather than to know true propositions. It is the second part that requires us to hold true beliefs. And it is true beliefs we are required to hold, not true knowledge, as the latter is redundant if knowledge is factive. In the paragraph following this quote James substitutes – correctly, I think – belief for knowledge:

*Believe truth! Shun error!*

Suppose we do not only substitute belief for knowledge, but additionally drop the imperative formulation in favor of a declarative one. Then we arrive at the formulation of the end I am focusing on: to hold beliefs whose contents are true and sufficiently informative to satisfy the ideal doxastic agent’s curiosity. My project is to formulate the means she has to take in order to attain the end we are assuming her to have.

Two points are worth noting. First, whether it is transparent to an agent that she has a particular end does not matter for the question which means she should take to attain this end. Second, I am assuming, but not requiring the ideal doxastic agent to have the end of holding true and sufficiently informative beliefs. In other words, I am not claiming that the ideal doxastic agent should have this end. On the instrumentalist view adopted in this book, to say that, in the narrow sense of this term, the ideal doxastic agent should have this – or, for that matter, any other – end is to say that having this end is a means to attaining some other end of hers. She can be required to have the latter end in turn only if having it is a means to attaining yet another end of hers. And so on until we arrive at the ideal doxastic agent’s ultimate ends. These are the ends she has not as means to attaining other ends of hers, but as ends in themselves. For this reason it is meaningless to ask whether the agent should have the ultimate ends she has. It is a factual, not a normative question which ultimate ends an agent has. In combination with her abilities and limitations, their totality characterizes the agent if she is rational.
2.3 Conditional belief and belief revision

On a synchronic level we consider how an ideal doxastic agent should organize her beliefs at a given moment in time under the assumption that her only end is to hold true beliefs that are sufficiently informative. On this level there is little, perhaps surprisingly little, to be said about how an ideal doxastic agent should organize her beliefs: she should not simultaneously believe and disbelieve that something is the case, nor should she simultaneously believe and refrain from believing that something is the case. These constraints are known as “consistency.” They are just about all there is to be said about how an ideal doxastic agent should organize her beliefs at a given moment in time.

Ida may be lucky and start out with many true beliefs, but that is luck, not rationality. Ida may be unlucky and start out with many false beliefs, but that is bad luck, not irrationality. An agent is rational if, and only if, she does what she ought to do. On the instrumentalist view adopted in this book, she is rational if, and only if, she takes the means to her ends. Unless Ida’s beliefs are not consistent, whether or not she is rational is not determined by what or how she believes at a given moment in time. It is determined by how she revises her beliefs across time when she receives new information.

New information sometimes just pops up, as when people look out of the window and cannot help but form the belief that it is sunny or when computer programs are fed new data by a user. New information sometimes is deliberately and actively sought by the ideal doxastic agent, say, because her ends are not met. The latter is the case if Ida is so curious as to desire an answer to the question whether or not Beijing is the capital of China. In order to satisfy her curiosity she needs to believe one of the answers to this question (it does not have to be the true answer). She does so by seeking new information in order to subsequently revise her beliefs. For instance, she may look it up on the internet.

Clifford (1877: part I) boldly claims that

it is wrong always, everywhere, and for anyone, to believe anything upon insufficient evidence.

And he continues by adding the following evaluation:

If a [hu]man, holding a belief [...], keeps down and pushes away any doubts [and] purposely avoids the reading of books [...] – the life of that [hu]man is one long sin against [hu]mankind.
2.3. CONDITIONAL BELIEF AND BELIEF REVISION

Clifford’s second claim seems to be that it is wrong (always, everywhere, and for anyone) to ignore new information or “evidence,” and maybe even to not actively seek new information. Clifford’s first claim may well be true if, and only if, it is restricted to ideal doxastic agents with the appropriate ends. His second claim, however, addresses a question I want to bracket. I want to do so in much the same way Hume (1748: sect. 10, part 1) does when he writes:

[a] wise [hu]man [...] proportions [her or] his belief to the evidence.

For the time being, let us ignore the numbers alluded to in Hume’s “proportions.” Then we note that Hume does not require a wise human to gather “evidence” if she does not already have it. Hume merely requires her to believe according to the information she happens to have. This means he restricts the rationality of beliefs to their revision, and remains silent on the information gathering process.

Adjusted to the present context Hume’s claim seems to be the following: an ideal doxastic agent whose only end is to hold true and sufficiently informative beliefs, and who accepts or believes the new information she receives, is rational only if she revises her beliefs according to this new information. She is irrational if she revises her beliefs contrary to this new information. I subscribe to these two claims.

In the previous section I have replaced James (1896)’s imperative formulation with a declarative one by assuming rather than requiring the ideal doxastic agent to have the end of holding beliefs that are true and sufficiently informative. Similarly, I have replaced Clifford (1877)’s evaluative wording with a descriptive one by assuming the ideal doxastic agent to accept the information she receives rather than calling her names for not doing so. Much like Hume (1748) I will leave unspecified how the ideal doxastic agent has arrived at her information, and how she comes to have the end of holding true and sufficiently informative beliefs. In contrast to my focus on doxastic agents that are ideal these two restrictions really do restrict: I have next to nothing to say about what real or ideal doxastic agents should do if their ends do not relate to the one I am focusing on. And I say next to nothing about when real or ideal doxastic agents should gain new information or how they should go about the information gathering process.

This restriction to the organization and revision of beliefs is due to my view that all normative aspects of the information gathering process pertain to non-cognitive actions such as carrying out an experiment or opening one’s eyes (a few remarks about these non-cognitive actions can be found in Brössel & Huber 2014). I assume the cognitive aspects of the information gathering process such as forming the belief that one has a certain experience to be factual, not normative.
What I will say about how an ideal doxastic agent should revise her beliefs if she receives new information essentially depends on the notion of a conditional belief. Initially Ida suspends judgment about whether or not it will be sunny on Wednesday, and whether or not she will have lunch in the park. In addition she firmly holds the conditional belief that she will have lunch in the park if it is sunny on Wednesday. Subsequently Ida comes to believe that it will be sunny on Wednesday. In this case she should stop suspending judgment and form the belief that she will have lunch in the park.

Why should she do so? She should do so because, by assumption, she accepts the new information that it will be sunny on Wednesday. And, per request, she should hold on to her conditional belief that she will have lunch in the park if it is sunny on Wednesday. If Ida met this assumption and satisfied this request, but did not revise her beliefs as indicated, she would violate another pair of consistency constraints: to not simultaneously believe and disbelieve that something is the case given that something else is; nor to simultaneously believe and refrain from believing that something is the case given that something else is. Just as Ida’s beliefs should be consistent, her conditional beliefs should be conditionally consistent.

Consistency, in its conditional as well as non-conditional variants, still is just about all there is to be said about how an ideal doxastic agent should organize her beliefs at a given moment in time. However, now that the ideal doxastic agent has accepted the new information, consistency has an impact on how she should revise her beliefs across time: Ida must do something in order to restore consistency. What she must do, and has done, may easily have gone unnoticed, though. As little as there is to be said about how an ideal doxastic agent should organize her beliefs at a given moment in time, as little, perhaps surprisingly little, there is to be said about how she should revise her beliefs across time if she receives new information: she should do so by not doing something, namely by not giving up those conditional beliefs of hers whose conditions are “directly affected by the new information in their entirety” (chapter 4 will explain this notion in detail).

Be consistent in your beliefs and conditional beliefs at any moment in time! Hold on to your conditional beliefs whose conditions are directly affected by the new information in their entirety across time! This is just about all there is to be said about how an ideal doxastic agent should organize her beliefs synchronically, as well as how she should revise her beliefs diachronically if she receives new information. Here the ideal doxastic agent is assumed to have the end of holding beliefs that are true and sufficiently informative, and to accept the new information she receives. It remains to be said what it is for her to hold a conditional belief.
To be sure, things get a little bit more complicated once we take into account Hume’s “proportions.” Real and ideal doxastic agents alike hold some beliefs more firmly than others: Ida’s belief that it will or will not be sunny on Wednesday is held more firmly than her conditional belief that she will have lunch in the park if it is sunny on Wednesday. This conditional belief of hers in turn is held more firmly than her “belief” that Beijing is the capital of China, a belief she holds with firmness zero.

More to Hume’s point, new information is not always of the same quality in the sense that different sources of information are not always deemed to be equally reliable by the ideal doxastic agent. Ida’s belief that it will be sunny on Wednesday will be held more firmly when she looks out of the window and seems to see that it is sunny than when a friend she trusts tells her so than when the weather forecast she deems unreliable predicts so. However, the basic idea remains the same: when she receives new information that she accepts, an ideal doxastic agent should revise her beliefs by holding on to those conditional beliefs of hers whose conditions are directly affected by the new information in their entirety. Consistency does the rest of the work.

A variant of the reverse of this idea has been suggested by Ramsey (1929: 15a):

If two people are arguing ‘If $p$ will $q$?’ and are both in doubt as to $p$, they are adding $p$ hypothetically to their stock of knowledge and arguing on that basis about $q$.

This quote by Ramsey has inspired the idea that belief revision is a guide to the acceptability of indicative conditionals. Identifying indicative conditionals with conditional beliefs and reversing this idea delivers the first part of my proposal: conditional beliefs guide the revision of beliefs.

The second part of my proposal can also be found in the quote by Ramsey, which continues thus:

so that in a sense ‘If $p$, $q$’ and ‘If $p$, $\neg q$’ are contradictories.

The second part of my proposal is the requirement that conditional beliefs should be non-contradictory or consistent “in a sense.” This sense is consistency extended to include its conditional variant.

What Ramsey and his contemporaries and predecessors did not have is the right theory of conditional belief. This theory was introduced only in Spohn (1988).
Spohn (1988)’s theory of conditional belief includes the requirement that, when she receives new information, the ideal doxastic agent should hold on to those conditional beliefs of hers whose conditions are directly affected by the new information in their entirety. Thus it comprises the first part of my proposal: that conditional beliefs guide the revision of beliefs. It also comprises the second part of my proposal: consistency in its non-conditional and conditional form.

 Needless to say, I do, of course, not claim credit for Spohn (1988)’s theory of conditional belief. All good things come in threes, though.

 Consider again the quote by Ramsey. I cannot say whether he distinguished between indicative conditionals, default conditionals, and counterfactuals when he wrote this passage. However, he added the precautionary clause “and are both in doubt as to $p$.” Perhaps his intention was to restrict the discussion to indicative conditionals, and he added the precautionary clause because he took it to imply that both people assign a subjective probability smaller than one, but greater than zero to $p$. Perhaps he intended the discussion to include indicative conditionals as well as default conditionals and counterfactuals, and he added the precautionary clause because he was aware that the three go together in this case. Perhaps there was another reason, or none at all.

 Either way, the third part of my proposal turns Ramsey’s precautionary clause into a positive requirement. Suppose that an ideal doxastic agent is in doubt as to the antecedent of a default conditional in the sense that she neither believes nor disbelieves this antecedent. Suppose further that she is certain of the default conditional itself. Then she should hold the conditional belief in its consequent given its antecedent. In particular, she should hold this conditional belief if she is certain of the default conditional because she is certain of the corresponding counterfactual.

 As before, things get more complicated once we take into account Hume’s “proportions.” The basic idea remains the same, though, and it will prove to be a powerful one: together with Spohn (1988)’s theory of conditional belief and the assumption about the truth conditions of default conditionals and counterfactuals from the previous chapter, the third part of my proposal will deliver the logic of default conditionals and counterfactuals. This illustrates how a metaphysical thesis can be thought of as a necessary condition for the possibility of satisfying norms from epistemology, and, hence, the possibility of attaining certain ends. Pursuing the end of holding true and sufficiently informative beliefs and other ends to be introduced later commits an ideal doxastic agent to the metaphysical thesis that descriptive normality behaves a certain way, and, given their truth conditions, that default conditionals and counterfactuals satisfy certain logical principles.
But I am way ahead. Let us begin at the beginning.
Chapter 3

Belief Revision

In this chapter I will first introduce the AGM theory of belief revision. Then I will focus on the problem of iterated belief revisions. This chapter heavily relies on Huber (2013b; invited).

3.1 The AGM theory of belief revision

Belief revision theory is the study of how an ideal doxastic agent should organize her beliefs and conditional beliefs at a given moment in time, and how she should revise these beliefs across time when she receives new information. Among other things, Ida believes that it will rain on Tuesday, that it will be sunny on Wednesday, and that weather forecasts are always right. Belief revision theory tells Ida how to revise her beliefs when she receives the information that the weather forecast for Tuesday and Wednesday predicts rain. As we will see, this depends on the details of her doxastic state. Under one way of filling in these details, she should keep her belief that it will rain on Tuesday and give up her belief that it will be sunny on Wednesday. To state how Ida should revise her beliefs when she receives new information in a general and precise manner, we need a representation of her doxastic state and the new information she receives.

On one reading, belief revision theory models belief as qualitative attitude of the agent towards the sentences or propositions she understands and has an opinion on, i.e. that are in her language: she believes a proposition, she disbelieves the proposition by believing its negation, or she suspends judgment about the proposition by neither believing nor disbelieving it. On another reading, it models belief as comparative attitude: she believes a proposition at least as much as another.
CHAPTER 3. BELIEF REVISION

This is different in the theory of subjective probabilities which is sometimes referred to as Bayesianism (Easwaran 2011a; b, Weisberg 2011). Here belief is modeled as quantitative attitude of the agent towards sentences or propositions in her language: she believes a proposition to a specific degree, namely her degree of belief in the proposition. In other words, the Bayesian takes into account Hume’s “proportions.” As we will see, to adequately model iterated belief revisions, belief revision theory also has to model the agent’s beliefs in a quantitative or numerical way. Thus the Bayesian is right to follow Hume in this regard. In addition the Bayesian holds that Hume’s “proportions” should obey the probability calculus. However, for the end of holding true and sufficiently informative beliefs that we are assuming Hume’s human to have, it will turn out to be wise to proportion her beliefs according to the rules of a different calculus.

The AGM theory of belief revision derives its name from the seminal paper Alchourrón & Gärdenfors & Makinson (1985). Comprehensive overviews can be found in Gärdenfors (1988), Gärdenfors & Rott (1995), and Rott (2001). One version of the AGM theory represents the ideal agent’s doxastic state by a set of sentences from a formal language, her belief set, together with an entrenchment ordering over these sentences. The latter represents the details of her doxastic state: it represents – in a comparative way – how firmly the agent holds the beliefs in her belief set. The new information is represented by a single sentence.

The AGM theory distinguishes between the easy case – called expansion – and the general case – called revision. In expansion the new information does not contradict the agent’s old belief set and is simply added. In revision the new information may contradict the agent’s old belief set. The general case of revision is difficult because the agent has to turn her old belief set, which is required to be consistent, into a new belief set that contains the new information and is also required to be consistent. One way to deal with the general case is in two steps. In a first step – called contraction – the old belief set is cleared of everything that contradicts the new information. In a second step the agent simply expands by adding the new information. This means that the difficult doxastic task is handled by contraction, which turns the general case of revision into the easy case of expansion.

A formal language $L$ is defined recursively as follows. First, $L$ contains the contradictory sentence ⊥ and all elements of a given countable set of propositional variables $PV = \{p, q, r, \ldots\}$. Second, whenever $\alpha$ and $\beta$ are sentences of $L$, then so are the negations of $\alpha$ and of $\beta$, $\neg \alpha$ and $\neg \beta$, respectively, as well as the conjunction $\alpha$ and $\beta$, $\alpha \land \beta$. Third, nothing else is a sentence of $L$.

The new information is represented as a single sentence $\alpha$ from $L$. 
3.1. THE AGM THEORY OF BELIEF REVISION

The ideal agent’s doxastic state is represented as a set of sentences from $\mathcal{L}$, her belief set $\mathcal{B}$, plus an entrenchment ordering for $\mathcal{B}$. Importantly, the entrenchment ordering $\preceq$ is defined relative to the agent’s belief set. It does most of the work in a revision of her beliefs by ordering them according to how reluctant she is to give them up: the more entrenched a belief, the more reluctant she is to give it up.

Suppose the ideal doxastic agent receives new information that contradicts her belief set. Since she accepts the new information, and since the new belief set that results from the revision must be consistent, some of her old beliefs have to go. The entrenchment ordering determines which beliefs have to go when: the least entrenched beliefs have to go first. If giving up those is not enough to restore consistency, the beliefs next in the entrenchment ordering have to go next. And so on. The beliefs that would be given up last are the most entrenched ones. Maximaliy requires these to be the tautological sentences. They must always be believed and never be given up, as doing so cannot restore consistency and the agent wants to hold many beliefs. On the other end of the spectrum are the least entrenched sentences. According to Minimality these must be the sentences the agent does not believe to begin with. These sentences do not belong to her belief set, so are gone before the revision process has even begun.

Suppose one sentence logically implies another. According to Dominance the latter sentence must not be more entrenched than the former. This means giving up her belief in the latter sentence requires the agent to also give up her belief in the former. Dominance implies that the entrenchment ordering must be reflexive: every sentence must be at least as entrenched as itself. Conjunctivity says that two sentences must not both be more entrenched than their conjunction. This means the agent must not give up her belief in a conjunction without giving up her belief in at least one of the conjuncts. Conjunctivity and Dominance imply that the entrenchment ordering must be connected: any two sentences must be comparable to each other in terms of their comparative entrenchment. That is, either the first sentence is at least as entrenched as the second, or the second sentence is at least as entrenched as the first, or both. Finally, to ensure that the entrenchment ordering is a “well-behaved” ordering relation that gives rise to belief sets, it is required to be transitive by Transitivity: if one sentence is at least as entrenched as a second, and the second sentence is at least as entrenched as a third, then the first sentence must be at least as entrenched as the third.

We can state these requirements more precisely as follows. Let $\vdash$ be the logical consequence relation on $\mathcal{L}$, and let $\text{Cn} (\mathcal{B}) = \{ \alpha \in \mathcal{L} : \mathcal{B} \vdash \alpha \}$ be the set of logical consequences of $\mathcal{B}$. An entrenchment ordering $\preceq$ for $\mathcal{B}$ has to satisfy the following postulates for all sentences $\alpha, \beta$, and $\gamma$ from the agent’s formal language $\mathcal{L}$. 

\begin{itemize}
  \item \textbf{Reflexivity:} $\alpha \preceq \alpha$
  \item \textbf{Conjunctivity:} $\alpha \preceq \gamma \Rightarrow (\alpha \preceq \beta \land \gamma \preceq \beta)$
  \item \textbf{Dominance:} $\alpha \preceq \beta \Rightarrow \text{Cn} (\mathcal{B}) \preceq \beta$
  \item \textbf{Transitivity:} $\alpha \preceq \beta \land \beta \preceq \gamma \Rightarrow \alpha \preceq \gamma$
\end{itemize}
1. If $\alpha \leq \beta$ and $\beta \leq \gamma$, then $\alpha \leq \gamma$. \hspace{1cm} \text{Transitivity}

2. If $\beta \in Cn(\{\alpha\})$, i.e. if $\{\alpha\} \vdash \beta$, then $\alpha \leq \beta$. \hspace{1cm} \text{Dominance}

3. $\alpha \leq \alpha \land \beta$ or $\beta \leq \alpha \land \beta$. \hspace{1cm} \text{Conjunctivity}

4. Suppose $\bot \notin Cn(\mathcal{B})$, i.e. $\mathcal{B} \not\vdash \bot$, then $\alpha \leq \delta$. \hspace{1cm} \text{Minimality}

5. If $\delta \leq \alpha$ for all $\delta$ from $\mathcal{L}$, then $\alpha \in Cn(\emptyset)$, i.e. $\emptyset \vdash \alpha$. \hspace{1cm} \text{Maximality}

The work that is done by the entrenchment ordering in a revision of the agent’s beliefs can also be described differently in terms of expansion $\dot{+}$, revision $\ast$, and contraction $\dot{-}$ operators. These turn belief sets and new information into new belief sets. Formally, they are functions from $\wp(\mathcal{L}) \times \mathcal{L}$ into $\wp(\mathcal{L})$. We will primarily be interested in the restrictions of these operators to a fixed belief set $\mathcal{B}$. This way we can think of them as we think of entrenchment orderings, namely as representations of the ideal agent’s doxastic state at a given moment in time.

Expansion $\dot{+}$ turns each old belief set $\mathcal{B} \subseteq \mathcal{L}$ and sentence $\alpha \in \mathcal{L}$ into a new belief set $\mathcal{B} + \alpha = Cn(\mathcal{B} \cup \{\alpha\})$. This is the easy case described earlier about which there is little more to be said. The difficult and more interesting case is revision $\ast$. It turns each old belief set $\mathcal{B} \subseteq \mathcal{L}$ and sentence $\alpha \in \mathcal{L}$ into a new belief set $\mathcal{B} \ast \alpha$ and is required to satisfy a number of postulates.

Closure requires revised belief sets to be closed under the logical consequence relation: after the revision, the agent ought to believe all (and only) the logical consequences of the revised belief set. After all, she wants to hold many beliefs (that are true). Given Closure, the assumption that belief sets are sets formulates part two of our non-conditional consistency requirement from section 2.3: the agent should not simultaneously believe and refrain from believing that something is the case. Congruence is similar in spirit to Closure and requires that it is the content of the new information received, and not its particular formulation, that determines what is added to, and removed from, the agent’s belief set in a revision.

Although stated as a requirement, Success formulates our assumption from section 2.3 that the agent accepts the new information she receives. It requires her to add the new information to the revised belief set – and, given Closure, all sentences the new information logically implies. Consistency formulates part one of our non-conditional consistency requirement from section 2.3: the agent should not simultaneously believe and disbelieve that something is the case. It requires the revised belief set to be consistent as long as the new information is consistent.
As an aside, the formulation of our conditional consistency requirement from section 2.3 will have to wait until the next chapter.

The remaining postulates all formulate different aspects of the idea that, when revising her belief set by new information, the ideal doxastic agent should add and remove as few beliefs as possible, subject to the constraints that the resulting belief set is consistent and that the new information has been added successfully. For adding new beliefs contains the risk of not shunning error; and removing beliefs contains the risk of not believing the truth.

Inclusion says that, when revising her belief set, the agent should not form any new beliefs that she does not also form when she simply adds the new information. Preservation says that, when revising her belief set by new information that does not contradict it, the agent should hold on to, or preserve, all beliefs in her belief set. Conjunction 1 requires that, when revising her belief set by a conjunction, the agent add only beliefs that she also adds when first revising her belief set by one of the two conjuncts, and then adding the second conjunct. Finally, Conjunction 2 requires that, when revising her belief set by a conjunction, the agent add all beliefs that she adds when first revising her belief set by one of the two conjuncts, and then adding the second conjunct – provided the second conjunct is consistent with the result of revising her belief set by the first conjunct.

We can state these requirements more precisely. A revision operator $\ast$ has to satisfy the following postulates for all sets of sentences $B$ of the ideal doxastic agent’s formal language $L$ and all sentences $\alpha$ and $\beta$ from $L$.

$\ast 1. B \ast \alpha = Cn (B \ast \alpha)$. 

Closure

$\ast 2. \alpha \in B \ast \alpha$. 

Success

$\ast 3. B \ast \alpha \subseteq Cn (B \cup \{\alpha\})$. 

Inclusion

$\ast 4. \text{If } \neg \alpha \notin Cn (B), \text{ i.e. if } B \not\vdash \neg \alpha, \text{ then } B \subseteq B \ast \alpha$. 

Preservation

$\ast 5. \text{If } Cn (\{\alpha\}) = Cn (\{\beta\}), \text{ i.e. if } \{\alpha\} \vdash \beta \text{ and } \{\beta\} \vdash \alpha, \text{ then } B \ast \alpha = B \ast \beta$. 

Congruence

$\ast 6. \text{If } \neg \alpha \notin Cn (\emptyset), \text{ i.e. if } \{\alpha\} \not\vdash \bot, \text{ then } \bot \notin B \ast \alpha$. 

Consistency

$\ast 7. B \ast (\alpha \land \beta) \subseteq Cn ((B \ast \alpha) \cup \{\beta\})$. 

Conjunction 1

$\ast 8. \text{If } \neg \beta \notin B \ast \alpha, \text{ i.e. (given } \ast 1\text{) if } B \ast \alpha \not\vdash \neg \beta, \text{ then } Cn ((B \ast \alpha) \cup \{\beta\}) \subseteq B \ast (\alpha \land \beta)$. 

Conjunction 2
The two-step view of revision described earlier is known as the *Levi identity* (Levi 1977). It has the ideal doxastic agent first contract her old belief set $B$ by the negation of the new information, $\neg \alpha$, thus making it consistent with the new information (as well as everything logically implied by the new information). Then it has her expand the result $B \neg \neg \alpha$ by adding the new information $\alpha$:

$$B \ast \alpha = Cn ((B \neg \alpha) \cup \{\alpha\})$$

The Levi identity puts contraction $\neg$ center stage of the belief revision process. Contraction turns each old belief set $B \subseteq \mathcal{L}$ and sentence $\alpha \in \mathcal{L}$ into a contracted belief set $B \neg \alpha$ that is cleared of $\alpha$ (as well as everything logically implying $\alpha$). A contraction operator $\neg$ has to satisfy the following postulates for all sets of sentences $B$ of the agent’s language $\mathcal{L}$ and all sentences $\alpha$ and $\beta$ from $\mathcal{L}$.

\begin{align*}
\neg 1. & \quad B \neg \alpha = Cn (B \neg \alpha). \quad \text{Closure} \\
\neg 2. & \quad \text{If } \alpha \notin Cn (\emptyset), \text{ i.e. if } \emptyset \not\vDash \alpha, \text{ then } \alpha \notin Cn (B \neg \alpha). \quad \text{Success} \\
\neg 3. & \quad B \neg \alpha \subseteq Cn (B). \quad \text{Inclusion} \\
\neg 4. & \quad \text{If } \alpha \notin Cn (B), \text{ i.e. if } B \not\vDash \alpha, \text{ then } B \neg \alpha = B. \quad \text{Vacuity} \\
\neg 5. & \quad \text{If } Cn (\{\alpha\}) = Cn (\{\beta\}), \text{ i.e. if } \{\alpha\} \vDash \beta \text{ and } \{\beta\} \vDash \alpha, \text{ then } B \neg \alpha = B \neg \beta. \quad \text{Congruence} \\
\neg 6. & \quad Cn (B) \subseteq Cn ((B \neg \alpha) \cup \{\alpha\}). \quad \text{Recovery} \\
\neg 7. & \quad (B \neg \alpha) \cap (B \neg \beta) \subseteq B \neg (\alpha \land \beta). \quad \text{Conjunction 1} \\
\neg 8. & \quad \text{If } \alpha \notin B \neg (\alpha \land \beta), \text{ i.e. (given $\neg 1$) if } B \neg (\alpha \land \beta) \not\vDash \alpha, \text{ then } B \neg (\alpha \land \beta) \subseteq B \neg \alpha. \quad \text{Conjunction 2}
\end{align*}

As before, Closure is a necessary means to attain the end of holding many beliefs (that are true). This time it requires contracted belief sets to be closed under the logical consequence relation: after the contraction, the agent should believe all (and only) the logical consequences of the contracted belief set. Given Closure the assumption that belief sets are sets again formulates part two of our non-conditional consistency requirement from section 2.3. Congruence is still similar in spirit to Closure and now requires that it is the content of the information to be removed, and not its particular formulation, that determines what is removed from the agent’s belief set in a contraction.
3.1. THE AGM THEORY OF BELIEF REVISION

Success is a negative formulation, in terms of contraction rather than revision, of our assumption from section 2.3 that the agent accepts the new information. It is again stated as a requirement and demands that, in contracting her belief set by a non-tautological sentence, the agent remove this sentence from her belief set – and, given Closure, all sentences logically implying it. Inclusion requires that, in contracting her belief set, the agent not add any new beliefs.

The remaining postulates all formulate different aspects of the idea that, when contracting her belief set by a sentence, the agent should remove as few beliefs as possible, subject to the constraints that the resulting belief set is consistent and that the sentence to be removed, together with all sentences logically implying it, is successfully removed. The reason is that she has the end of holding true beliefs that are sufficiently informative. Vacuity requires the agent to leave her belief set unchanged when she contracts it by a sentence that is not even part of her belief set to begin with. Recovery requires that, when she contracts her belief set by a sentence, she remove so few beliefs that adding the removed sentence again afterwards allows her to recover all previously removed beliefs.

Conjunction 1 says that, when contracting her belief set by a conjunction, the agent should not remove any beliefs that she does not also remove when she contracts her belief set by one or the other of the two conjuncts alone. Finally, Conjunction 2 requires the agent to contract as follows: if she removes a conjunct in contracting her belief set by a conjunction, then she must not remove any belief in contracting her belief set by this conjunct that she does not also remove in contracting her belief set by the entire conjunction. The idea behind the last two postulates is that giving up her belief in one of its conjuncts is all the agent needs to do in order to give up her belief in an entire conjunction.

The Levi identity turns each contraction operator $\hat{-}$ satisfying $\hat{-}1, \hat{-}8$ into a revision operator $\ast$ satisfying $\ast1, \ast8$. The converse is true of the Harper identity (Harper 1976a). The latter has the ideal doxastic agent first revise her old belief set $B$ by the negation of the new information, $\neg\alpha$. Then it has her remove everything from the result $B \ast \neg\alpha$ that was not already also a logical consequence of the old belief set $B$:

$$B \hat{-}\alpha = (B \ast \neg\alpha) \cap Cn (B)$$

If we have the ideal doxastic agent’s belief set $B$, we can – with Rott (1991) – use her entrenchment ordering $\preceq$ for $B$ to define her revision operator $\ast$ restricted to $B$ as follows: for all sentences $\alpha$ in $L$,

$$B \ast \alpha = Cn (\{ \beta \in L : \neg\alpha \prec \beta \} \cup \{ \alpha \}) .$$

Here $\alpha \prec \beta$ holds if, and only if, $\alpha \leq \beta$, but not also $\beta \leq \alpha$ (i.e. $\beta \not\leq \alpha$).
The idea behind this equation is this. When the agent wants to revise her old belief set $B$ by the new information $\alpha$ in accordance with postulates $*1-8*$, she first has to clear $B$ of $\neg \alpha$ as well as everything that is as entrenched as, or less entrenched than, $\neg \alpha$. For instance, $B$ also has to be cleared of everything that logically implies $\neg \alpha$. However, if her entrenchment ordering $\preceq$ for $B$ satisfies postulates $1 \leq 5$, all sentences $\beta$ that are more entrenched than $\neg \alpha$ can and, hence, should be preserved. This gives us the “reduced belief set”

$$\{\beta \in L : \neg \alpha < \beta\}.$$

Next the agent has to add the new information $\alpha$. This gives us the set of sentences

$$\{\beta \in L : \neg \alpha < \beta\} \cup \{\alpha\}.$$

Finally she must add all sentences that are logically implied by the reduced belief set together with the new information. As shown by Gärdenfors (1988) as well as Gärdenfors & Makinson (1988), and using a result by Rott (1991), one can then prove the following “representation theorem.”

**Theorem 1 (Gärdenfors 1988, Gärdenfors & Makinson 1988)** Let $L$ be some formal language.

For each set of sentences $B \subseteq L$ and each entrenchment ordering $\preceq$ for $B$ satisfying $1 \leq 5$ there is a revision operator $\ast$ restricted to $B$– i.e., a function from $\{B\} \times L$ into $\wp(L)$ satisfying $*1-8*$ restricted to $B$– such that for all sentences $\alpha \in L$:

$$B \ast \alpha = \text{Cn}\left(\{\beta \in L : \neg \alpha < \beta\} \cup \{\alpha\}\right).$$

For each revision operator $\ast$ from $\wp(L) \times L$ into $\wp(L)$ satisfying $*1-8*$ and each set of sentences $B \subseteq L$ there is an entrenchment ordering $\preceq$ for $B$ satisfying $1 \leq 5$ such that for all sentences $\alpha \in L$:

$$B \ast \alpha = \text{Cn}\left(\{\beta \in L : \neg \alpha < \beta\} \cup \{\alpha\}\right).$$

This theorem states that its equation translates the postulates for entrenchment orderings into the postulates for revision operators, and conversely. This means revision operators can be represented by entrenchment orderings, and conversely. A different interpretation of this theorem says that its equation renders the agent’s entrenchment ordering’s obeying of postulates $1 \leq 5$ a means to attain the end of her revision operator’s satisfying postulates $*1-8*$; and that it renders her revision operator’s satisfying postulates $*1-8*$ a means to attain the end of her entrenchment ordering’s obeying of postulates $1 \leq 5$.

An analogous theorem holds for the relationship between the postulates for entrenchment orderings and the postulates for contraction operators.
3.2 Systems of spheres

There is a different way of representing postulates *1-*8 for revision operators * due to Grove (1988). Similar to Lewis’ (1973a) semantics for counterfactuals, which we will discuss in chapter 7, it uses “systems of spheres” that are defined on a set of possible worlds instead of entrenchment orderings that are defined on a formal language.

A set of possible worlds can be thought of as a set of complete, or maximally specific, descriptions of the way reality could be. One approach, used by Grove (1988), is to identify possible worlds with maximally consistent sets of sentences from the agent’s language $L$. These are consistent sets of sentences that become inconsistent if only a single new sentence is added. Another approach is to take possible worlds as primitive. For present purposes we do not have to take a stance on what possible worlds are. I will assume that we are given a non-empty set of possible worlds $W_L$ relative to which we interpret the sentences from $L$.

In order to state Grove (1988)’s approach it will be useful to have the following notation. $\left[\alpha\right] = \{w \in W_L : w \models \alpha\}$ is the proposition expressed by the sentence $\alpha$, i.e. the set of possible worlds $w$ in which $\alpha$ is true, $w \models \alpha$. If $W_L$ is the set of maximally consistent sets of sentences in $L$, then $w \models \alpha$ holds just in case $\alpha \in w$. If possible worlds are taken as primitive, the truth relation $\models$ has to be analyzed differently. Either way, there may be sets of possible worlds, or propositions, $B \subseteq W_L$ that are not expressed by any sentence $\alpha$ from $L$. (For the purposes of this chapter we can assume every set of possible worlds to be a proposition. The next chapter will be more precise.)

$\left[B\right] = \{w \in W_L : w \models \alpha\text{ for all }\alpha \in B\}$ is the proposition expressed by the set of sentences $B$. If $W_L$ is the set of maximally consistent sets of sentences in $L$, then for each proposition $B \subseteq W_L$ there is a set of sentences in $L$, a “theory,” $t(B)$, such that $B = \left[t(B)\right]$ ($t(B)$ is the intersection of $B$). This means that every proposition $B \subseteq W_L$ can be expressed by a set of sentences in $L$. If possible worlds are taken as primitive, this has to be assumed. I will make this assumption.

Let $B \subseteq W_L$ be a proposition, and let $S \subseteq \emptyset (W_L)$ be a set of propositions. $S$ is a system of spheres in $W_L$ that is centered on $B$ if, and only if, the following four conditions hold for all propositions $A, C \subseteq W_L$ and all sentences $\alpha$ from $L$.

\[\text{S1. If } A \in S \text{ and } C \in S, \text{ then } A \subseteq C \text{ or } C \subseteq A.\]

\[\text{S2. } B \in S; \text{ and, if } A \in S, \text{ then } B \subseteq A.\]

\[\text{S3. } W_L \in S.\]
S4. If $\llbracket\alpha\rrbracket \cap D \neq \emptyset$ for some $D \in S$, then there is $D' \in S$ such that: $\llbracket\alpha\rrbracket \cap D' \neq \emptyset$, and $D' \subseteq E$ for all $E \in S$ with $\llbracket\alpha\rrbracket \cap E \neq \emptyset$.

S1 says that systems of spheres are nested: any two spheres are such that one is contained in the other, or they are the same sphere. S2 says that the center of a system of spheres is itself a sphere in this system, and that every other sphere in the system contains the center as a sub-sphere. S3 says that the set of all possible worlds is a sphere in every system of spheres. Given S1, this implies that the set of all possible worlds contains every other sphere in any given system of spheres as a sub-sphere. Finally, in combination with S3, S4 says the following: for each logically consistent sentence $\alpha$ there is a smallest sphere $D' \in S$ that properly overlaps (has a non-empty intersection) with the proposition expressed by $\alpha$, $\llbracket\alpha\rrbracket$.

We will assume that the center (and, given S2, any other sphere) of a system of spheres is not empty unless $W_L$ is the only non-empty sphere.

For any sentence $\alpha$ from $L$, let $c_S(\alpha) = \llbracket\alpha\rrbracket \cap D'$ (= $\emptyset$ if $\alpha$ is logically inconsistent). $c_S(\alpha)$ is the set of possible worlds in $\llbracket\alpha\rrbracket$ that are "closest" to the center $B$, where the meaning of ‘closeness’ is determined by the system of spheres $S$. If $\alpha$ is logically consistent with (a set of sentences expressing) the center $B$, then $c_S(\alpha)$ is just the intersection of the center with the proposition $\llbracket\alpha\rrbracket$, $\llbracket\alpha\rrbracket \cap B$. This corresponds to the easy case of an expansion of the belief set $t(B)$.

The difficult case of a revision of the belief set $t(B)$ arises when $\alpha$ is logically inconsistent with (every set of sentences expressing) the center $B$. In this case the agent has to leave the center and move to the nearest or closest sphere $D'$ that properly overlaps with the proposition expressed by $\alpha$ and adopt their intersection, $\llbracket\alpha\rrbracket \cap D'$, as $c_S(\alpha)$ (unless, of course, $\alpha$ is itself inconsistent).

\footnote{In one respect Grove (1988)'s notion of a system of spheres is more general than Lewis (1973a)'s. Grove (1988) allows it to be centered on arbitrary propositions $B \subseteq W_L$; whereas, Lewis (1973a: 14f) assumes the center to contain the actual world, and it alone. These last two assumptions are known as the principles of weak and strong centering, respectively.

In another respect Grove (1988)'s notion is less general than Lewis (1973a)'s. S4 is a version of the “limit assumption” for overall similarity, which Lewis (1973a: 19f) rejects. In the next chapter a doxastic version of the limit assumption will be shown to be a consequence of the consistency requirement for conditional beliefs. In chapter 7 a version of the limit assumption for descriptive normality will be shown to be a consequence of the consistency requirement for conditional beliefs and the royal rule which relates descriptive normality and conditional beliefs.

Finally, while Lewis (1973a) does not assume S3, he assumes systems of spheres to be closed under arbitrary unions and non-empty, but otherwise arbitrary intersections. The first condition implies that the empty set is a sphere in every system of spheres (and that systems of spheres are also closed under empty intersections). Therefore the center has to be defined differently. Other than that these two conditions only make a difference if there are infinitely many spheres.}
3.2. SYSTEMS OF SPHERES

Here is a picture representing this situation:

If the ideal agent’s doxastic state is represented by a system of spheres $S$ that is centered on the proposition $[B]$ which is expressed by her belief set $\mathcal{B}$, we can define her revision operator $\ast$ restricted to $\mathcal{B}$ as follows:

$$\mathcal{B} \ast \alpha = t(c_S(\alpha))$$

The idea behind this equation is this. What the agent should believe after revising $\ast$ her old belief set $\mathcal{B}$ by the new information $\alpha$ is a set of sentences, or theory, expressing the proposition $c_S(\alpha)$. The latter contains the possible worlds in $[\alpha]$ that are closest, in the sense of $S$, when the center is the proposition expressed by her old belief set, $[B]$.

Expansion is the special case in which the proposition expressed by the new information properly overlaps with the proposition expressed by her old belief set, $[\alpha] \cap [\mathcal{B}] \neq \emptyset$. In this special case the agent does not have to leave the old center $[\mathcal{B}]$ of her doxastic state; it suffices if she narrows it down to the possible worlds also contained in $[\alpha]$. However, in the general case of revision this intersection may be empty. In this general case she may have to leave the old center $[\mathcal{B}]$ of her doxastic state and move to the smallest sphere $D'$ that properly overlaps with $[\alpha]$ and adopt their intersection, $D' \cap [\alpha]$, as the new center of her doxastic state.

As before we can picture the system of spheres centered on $[\mathcal{B}]$ as an “onion” around $[\mathcal{B}]$. The grey area $[\mathcal{B} \ast \alpha] = c_S(\alpha) = D' \cap [\alpha]$ is the logically strongest proposition the agent should believe after revising her old belief set $\mathcal{B}$ by the new information $\alpha$. It should be the new center of her doxastic state.
Grove (1988) proves the following representation theorem.

**Theorem 2 (Grove 1988)** Let $\mathcal{L}$ be a formal language, and let $W_\mathcal{L}$ be a non-empty set of possible worlds that meets our assumption for $\mathcal{L}$.

For each set of sentences $\mathcal{B} \subseteq \mathcal{L}$ and each system of spheres $S$ in $W_\mathcal{L}$ that is centered on $\mathcal{B}$ and satisfies $S_1$-$S_4$ there is a revision operator $\ast$ restricted to $\mathcal{B}$ such that for all sentences $\alpha$ from $\mathcal{L}$:

$$\mathcal{B} \ast \alpha = t\left(\mathcal{C}_S(\alpha)\right).$$

For each revision operator $\ast$ from $\wp(\mathcal{L}) \times \mathcal{L}$ into $\wp(\mathcal{L})$ satisfying $\ast_1$-$\ast_8$ and each set of sentences $\mathcal{B} \subseteq \mathcal{L}$ there is a system of spheres $S$ in $W_\mathcal{L}$ that is centered on $\mathcal{B}$ and satisfies $S_1$-$S_4$ such that for all sentences $\alpha$ from $\mathcal{L}$:

$$\mathcal{B} \ast \alpha = t\left(\mathcal{C}_S(\alpha)\right).$$

This theorem states that its equation translates the conditions on systems of spheres into the postulates for revision operators, and conversely. This means that revision operators can be represented by systems of spheres. It also means that the equation of the theorem renders the agent’s revising her beliefs by relying on a system of spheres satisfying conditions $S_1$-$S_4$ a means to attain the end of her revision operator’s obeying of postulates $\ast_1$-$\ast_8$. The same holds for the converse of these two claims.

In a sense, then, we can say that entrenchment orderings, revision operators, contraction operators, and reliance on a system of spheres all impose the same requirements on the revision of an ideal doxastic agent’s beliefs.
3.3 Iterated belief revision

In the AGM theory the ideal agent’s old doxastic state is represented by her belief set $B$ together with her entrenchment ordering $\preceq$ for $B$. The latter ordering guides the revision process: it specifies which elements of the old belief set are given up, and which are kept, when new information $\delta$ is received. The result of revising the old belief set $B$ by the new information $\delta$ is a new belief set $B \ast \delta$.

Ida’s old belief set $B$ includes the beliefs that it will rain on Tuesday, that it will be sunny on Wednesday, and that weather forecasts are always right. Suppose her belief $\alpha$ that it will be sunny on Wednesday is less entrenched than her belief $\beta$ that it will rain on Tuesday. Suppose further the latter in turn is less entrenched than her belief $\gamma$ that weather forecasts are always right so that $\bot \prec \alpha \prec \beta \prec \gamma$.

On Monday Ida receives the information that the weather forecast for Tuesday and Wednesday predicts rain, $\delta$. We assume that she accepts this new information; the AGM postulate of Success requires her to do so. Either way, now she has to give up her belief $\alpha$ that it will be sunny on Wednesday or her belief $\gamma$ that weather forecasts are always right. The reason is consistency: it follows from $\delta$ that at least one of these two beliefs is false, $\{\delta\} \vdash \neg(\alpha \land \gamma)$, which implies $\alpha \land \gamma \preceq \neg \delta$. Since $\alpha$ is less entrenched than $\gamma$, $\alpha \prec \gamma$, $\alpha$ has to go. On the other hand, $\gamma$ can and – given her end of holding many or informative beliefs: should – stay.

Furthermore, since $\{\gamma, \delta\} \not\vdash \neg \beta$, Ida need not and – for the same reason as above: should not – give up her belief $\beta$ that it will rain on Tuesday. This is so even if she holds on to her belief $\gamma$ that weather forecasts are always right, and adds the belief $\delta$ that the weather forecast for Tuesday and Wednesday predicts rain. In addition let us assume that $\neg \delta \prec \beta$ so that Ida’s entrenchment ordering and new belief set $B \ast \delta$ look as follows: where $\mu \sim \nu$ is shorthand for $\mu \preceq \nu$ and $\nu \preceq \mu$,

$$\bot \sim \neg \alpha \prec \alpha \sim \alpha \land \gamma \preceq \neg \delta \prec \beta \prec \gamma \prec \alpha \lor \neg \alpha$$

$$B \ast \delta = Cn(\{\epsilon \in \mathcal{L} : \neg \delta \prec \epsilon\} \cup \{\delta\}) \supset \{\beta, \gamma, \delta, \neg \alpha\}$$

To Ida’s surprise on Tuesday she receives the information that it does not rain after all. Therefore she has to revise her newly acquired belief set $B \ast \delta$ a second time by $\neg \beta$ to correct her belief $\beta$ that it rains on Tuesday. In addition, Ida has to give up her belief $\delta$ that the weather forecast for Tuesday and Wednesday predicts rain (this might be because she has misheard the weather forecast) or her belief $\gamma$ that weather forecasts are always right (this might be because she has been too gullible). Again, the reason is consistency: it follows from $\neg \beta$ that at least one of these two beliefs is false, $\{\neg \beta\} \vdash \neg(\delta \land \gamma)$, which implies $\delta \land \gamma \preceq \neg \beta$. 

Unfortunately the AGM theory is of no help here. While Ida could use her entrenchment ordering to revise her old belief set \( B \) to obtain a new belief set \( B * \delta \), the entrenchment ordering itself has not been revised. The AGM theory is silent as to whether \( \delta \) is now more entrenched than, as entrenched as, or less entrenched than \( \gamma \). However, the latter is exactly the kind of information that Ida needs to revise her beliefs a second time. More generally, the problem is that Ida’s doxastic state is represented as a belief set plus an entrenchment ordering before the revision process, but as a belief set without an entrenchment ordering after the revision process (or with an unspecified entrenchment ordering that is not determined by her previous one and the new information received). To handle \textit{iterated} belief revisions the ideal agent’s doxastic state has to be represented in the same way before and after the revision process. Gärdenfors & Rott (1995: 37) call this the “principle of categorical matching.”

Put differently, the problem is that *1-*8 only guide one-step revisions of the form \( \mathcal{B} * \alpha \). To handle \textit{iterated} belief revisions, additional postulates must be added that guide two-step revisions of the form \( (\mathcal{B} * \alpha) * \beta \). Otherwise the entrenchment ordering for \( \mathcal{B} \) that represents \( * \) restricted to \( \mathcal{B} \) (see theorem 1) does not sufficiently constrain the entrenchment ordering for \( \mathcal{B} * \alpha \) that represents \( * \) restricted to \( \mathcal{B} * \alpha \).

Nayak (1994), Boutilier (1996), Darwiche & Pearl (1997), Segerberg (1998), Fermé (2000), Rott (2003; 2006), and others do exactly this. They augment the AGM postulates by additional ones that indirectly guide how the agent should revise her entrenchment ordering in addition to her belief set when she receives new information. On their accounts the ideal agent’s doxastic state is represented as a belief set plus an entrenchment ordering before and after the revision process. Importantly, both of these elements are revised when new information is received.

Let us consider the postulates from Darwiche & Pearl (1997). Like *1-*8, they center around the idea of adding and removing as few beliefs as possible, subject to the constraints that belief sets are consistent and that new information is added successfully. The reason still is that adding new beliefs contains the risk of not shunning error; and removing beliefs contains the risk of not believing the truth.

*9 says that revising her old belief set by new information should result in the same new belief set as first revising her old belief set by a logical consequence of the new information, and subsequently revising the resulting belief set by the new information in its entirety. That is, revision by a more specific piece of information – say, that her friend Bay had caffeine-free coffee (so is not full of caffeine) – should override all changes that result from first revising Ida’s old belief set by a less specific piece of information – say, that Bay had coffee (which suggests that she is full of caffeine given Ida’s belief that coffee normally contains caffeine).
3.3. ITERATED BELIEF REVISION

∗10 is similar in spirit. It says that revising her old belief set consecutively by two pieces of information that are logically inconsistent should result in the same new belief set as revising her old belief set by the second piece of information alone. That is, revision by the second piece of information – say, that Ida had red wine from Burgundy – should override all changes that result from first revising her old belief set by the first piece of information that is logically incompatible with the second piece of information – say, that Ida had no wine.

Suppose the agent holds a belief after revising her old belief set by a piece of information. This may, but need not be new belief, i.e. a belief not held previously. ∗11 says that the she should also hold this belief if she first revises her old belief set by this very belief, and subsequently revises the resulting belief set by said piece of information. Finally, suppose there is a sentence that is logically compatible with the result of revising the agent’s old belief set by a piece of information. ∗12 says that this sentence should also be logically compatible with what she ends up believing if she first revises her old belief set by this very sentence, and subsequently revises the resulting belief set by said piece of information.

More precisely, a revision operator ∗ has to satisfy the following postulates for all sets of sentences \( B \) of the agent’s language \( L \) and all sentences \( \alpha \) and \( \beta \) from \( L \).

∗9. If \( \beta \in Cn(\{\alpha\}) \), i.e. if \( \{\alpha\} \vdash \beta \), then \((B ∗ \beta) ∗ \alpha = B ∗ \alpha\).

∗10. If \( \neg \beta \in Cn(\{\alpha\}) \), i.e. if \( \{\alpha\} \vdash \neg \beta \), then \((B ∗ \beta) ∗ \alpha = B ∗ \alpha\).

∗11. If \( \beta \in B ∗ \alpha \), i.e. (given ∗1) if \( B ∗ \alpha \vdash \beta \), then \( \beta \in (B ∗ \beta) ∗ \alpha \).

∗12. If \( \neg \beta \not\in B ∗ \alpha \), i.e. (given ∗1) if \( B ∗ \alpha \not\vdash \neg \beta \), then \( \neg \beta \not\in (B ∗ \beta) ∗ \alpha \).

To better understand what these four new postulates require, it will be helpful to consider the following reformulation of a system of spheres \( S \) in \( W_L \) centered on \( B \). Let \( B \subseteq W_L \) be a proposition, and let \( \leq \) be a binary relation on \( W_L \). \( \leq \) is an implausibility ordering on \( W_L \) with center \( B \) if, and only if, the following holds for all possible worlds \( w, w', \) and \( w'' \) from \( W_L \) and all propositions \( A \subseteq W_L \).

\[
\begin{align*}
\leq 1. \hspace{1cm} w \leq w' \text{ or } w' \leq w. \hspace{1cm} \leq \text{ is connected} \\
\leq 2. \hspace{1cm} \text{If } w \leq w' \text{ and } w' \leq w'', \text{ then } w \leq w''. \hspace{1cm} \leq \text{ is transitive} \\
\leq 3. \hspace{1cm} w \in B \text{ if and only if for all } w' \in W_L : w \leq w'. \\
\leq 4. \hspace{1cm} \text{If } A \neq \emptyset, \text{ then } \{v \in A : v \leq w^+ \text{ for all } w^+ \in A\} \neq \emptyset.
\end{align*}
\]
An implausibility ordering on $W_L$ with center $B$ orders the possible worlds in $W_L$ according to their implausibility. This time it follows that the center $B$ is not empty if, as we assume, the set of all possible worlds $W_L$ is not empty.

$\leq 1$ requires that any two possible worlds can be compared with respect to their implausibility: either the first possible world is at least as implausible as the second, or the second possible world is at least as implausible as the first, or both.

$\leq 2$ requires that the ordering is transitive: if one possible world is at least as implausible as a second, and the second possible world is at least as implausible as a third, then the first possible world is at least as implausible as the third.

$\leq 3$ requires that the possible worlds in the center are no more implausible than all other possible worlds. That is, the center is the proposition that contains all and only the least implausible possible worlds. Finally, $\leq 4$ requires that each proposition that contains a possible world also contains a possible world that is no more implausible than any possible world in this proposition. That is, each non-empty or logically consistent proposition contains a least implausible possible world.

A system of spheres in $W_L$ that is centered on $B$ can be understood as an implausibility ordering on $W_L$ whose center $B$ comprises the least implausible possible worlds. The problem with the AGM approach can now be described as follows. Before the revision process the ideal agent’s doxastic state is represented as a belief set $B$ plus an implausibility ordering $\leq_B$ with center $\llbracket B \rrbracket$. Yet after a revision by the new information $\alpha$ the ideal agent’s doxastic state is represented as a belief set $B \ast \alpha$ without an implausibility ordering $\leq_{B \ast \alpha}$ (or with an unspecified implausibility ordering that is not determined by $\leq_B$ and $\alpha$). Darwiche & Pearl (1997)’s postulates $\ast 9-\ast 12$ address this problem by indirectly guiding the revision of the implausibility ordering. For a fixed belief set $B$, and with $w < w'$ shorthand for $w \leq w'$, but not $w' \leq w$, postulates $\ast 9-\ast 12$ restricted to $B$ become the following requirements for all possible worlds $w$ and $w'$ from $W_L$ and all sentences $\alpha$ from the agent’s language $L$.

$\leq 5$. If $w, w' \subseteq \llbracket \alpha \rrbracket$, then $w \leq_B w'$ just in case $w \leq_{B \ast \alpha} w'$.

$\leq 6$. If $w, w' \not\subseteq \llbracket \alpha \rrbracket$, then $w \leq_B w'$ just in case $w \leq_{B \ast \alpha} w'$.

$\leq 7$. If $w \subseteq \llbracket \alpha \rrbracket$ and $w' \not\subseteq \llbracket \alpha \rrbracket$ and $w <_B w'$, then $w <_{B \ast \alpha} w'$.

$\leq 8$. If $w \subseteq \llbracket \alpha \rrbracket$ and $w' \not\subseteq \llbracket \alpha \rrbracket$ and $w \leq_B w'$, then $w \leq_{B \ast \alpha} w'$.
3.3. ITERATED BELIEF REVISION

≤ 1 says that the implausibility ordering among the possible worlds within the proposition expressed by the new information should be the same before and after a revision by the new information. ≤ 2 says that the implausibility ordering among the possible worlds outside of the proposition expressed by the new information should also be the same before and after a revision by the new information. As to ≤ 3, suppose a possible world within the proposition expressed by the new information is less implausible than a possible world outside of this proposition before a revision by the new information. According to ≤ 3 this should remain so after a revision by the new information. ≤ 4 says something similar. Suppose a possible world within the proposition expressed by the new information is at least as implausible as a possible world outside of this proposition before a revision by the new information. According to ≤ 4 this should also remain so after a revision by the new information.

Before we turn to a representation theorem for iterated belief revisions let us consider a third representation theorem for belief revision. As mentioned, in a sense, entrenchment orderings, revision operators, contraction operators, and reliance on a system of spheres all impose the same requirements on the revision of an ideal doxastic agent’s beliefs. According to the following theorem due to Grove (1988) these are also the requirements imposed by implausibility orderings.

**Theorem 3 (Grove 1988)** Let \( \mathcal{L} \) be a formal language, and let \( W_\mathcal{L} \) be a non-empty set of possible worlds that meets our assumption for \( \mathcal{L} \).

For each set of sentences \( B \subseteq \mathcal{L} \) and each implausibility ordering \( \leq \) on \( W_\mathcal{L} \) with center \([B]\) that satisfies \( \leq 1, \leq 4 \) there is a revision operator \( * \) restricted to \( B \) such that for all sentences \( \alpha \) from \( \mathcal{L} \):

\[
B * \alpha = t\left(\{\omega \in [\alpha] : \omega \leq \omega' \text{ for all } \omega' \in [\alpha]\}\right).
\]

For each revision operator \( * \) from \( \wp(\mathcal{L}) \times \mathcal{L} \) into \( \wp(\mathcal{L}) \) that satisfies \( *1,*8 \) and each set of sentences \( \mathcal{B} \subseteq \mathcal{L} \) there is an implausibility ordering \( \leq_{\mathcal{B}} \) on \( W_\mathcal{L} \) with center \([\mathcal{B}]\) satisfying \( \leq 1, \leq 4 \) such that for all sentences \( \alpha \) from \( \mathcal{L} \):

\[
\mathcal{B} * \alpha = t\left(\{\omega \in [\alpha] : \omega \leq_{\mathcal{B}} \omega' \text{ for all } \omega' \in [\alpha]\}\right).
\]

(If two sets of sentences \( \mathcal{B} \) and \( \mathcal{B}' \) are logically equivalent, they induce the same center \([\mathcal{B}] = [\mathcal{B}']\). In this case one can assume that \( \leq_{\mathcal{B}} = \leq_{\mathcal{B}'} \).)

\( \{\omega \in [\alpha] : \omega \leq \omega' \text{ for all } \omega' \in [\alpha]\} \) is the set of least implausible possible worlds in which the new information \( \alpha \) is true. It is the proposition expressed by the belief set \( \mathcal{B} * \alpha \) that results from revising \( * \) the ideal doxastic agent’s old belief set \( \mathcal{B} \) by new information \( \alpha \).
Against this background we can now state the following theorem for iterated belief revisions. It follows from a more general result by Darwiche & Pearl (1997) who extend previous work by Katsuno & Mendelzon (1991) on non-iterated belief revision. The theorem backs the claim that postulates $\ast 9-\ast 12$ for revision operators become requirements $\leq 5 \leq 8$ for implausibility orderings.

**Theorem 4 (Darwiche & Pearl 1997)** Let $\mathcal{L}$ be a formal language, and let $W_\mathcal{L}$ be a non-empty set of possible worlds that meets our assumption for $\mathcal{L}$.

Suppose $\ast$ is a revision operator from $\wp(\mathcal{L}) \times \mathcal{L}$ into $\wp(\mathcal{L})$ that satisfies $\ast 1-\ast 8$. According to theorem 3, there exists a family of implausibility orderings $(\leq_B)_{B \subseteq \mathcal{L}}$ on $W_\mathcal{L}$ such that for each set of sentences $B \subseteq \mathcal{L}$: $\leq_B$ satisfies $1 \leq 4$ and is such that, for all sentences $\alpha$ from $\mathcal{L}$, $B \ast \alpha = t (\{\omega \in \llbracket \alpha \rrbracket : \omega \leq_B \omega' \text{ for all } \omega' \in \llbracket \alpha \rrbracket \})$.

For this $\ast$ and any one of these families $(\leq_B)_{B \subseteq \mathcal{L}}$: $\ast$ satisfies $\ast 9-\ast 12$ if, and only if, for every set of sentences $B \subseteq \mathcal{L}$, $\leq_B$ satisfies $5 \leq 8$.

The approaches to iterated belief revision mentioned above have in common that they satisfy Gärdnforss and Rott (1995)’s principal of categorial matching: they represent the ideal agent’s doxastic state as a belief set plus an entrenchment ordering / system of spheres / implausibility ordering both before and after the revision process. Furthermore these approaches have in common that the new information is represented as just a sentence or propositional content. The latter is also true for the approach by Jin & Thielscher (2007) discussed below, but not for what Rott (2009) calls “two-dimensional” belief revision operators (see also Cantwell 1997, Ferné & Rott 2004, and Rott 2007).

In “one-dimensional” belief revision the new information takes the form of an input sentence $\alpha$. It is then the job of the belief revision method, as opposed to the new information itself, to specify where in the new entrenchment ordering this sentence should be placed. Nayak (1994)’s lexicographic revision and Boutilier (1996)’s natural revision and Segerberg (1998)’s irrevocable revision and Rott (2006)’s irrefutable revision are examples of revision methods specifying this.

In two-dimensional belief revision it is the new information itself that carries at least part of this information. Here the new information does not merely say that the input sentence $\alpha$ is true, so should be accepted according to the Success postulate. Instead the new information now specifies, at least to some extent, how firmly $\alpha$ is accepted by specifying that, in the new entrenchment ordering $\leq ^+$, $\alpha$ is at least as entrenched as some “reference sentence” $\beta$. Thus the new information is now of the form: $\beta \leq ^+ \alpha$. In addition the new information now says, not that $\alpha$ is true, but how firmly $\alpha$ is accepted. Thus the new information is now about the ideal agent’s new doxastic state – her internal, not the external world.
3.3. **ITERATED BELIEF REVISION**

Importantly, the shift from one- to two-dimensional belief revision replaces the *requirement* that the ideal doxastic agent accept the propositional content of the new information by the *assumption* that she does: our assumption from section 2.3 is substituted for the Success postulate. In addition this shift begins to incorporate the insight that acceptance of, or belief in, the propositional content of the new information – like belief in the sentences in her belief set – is a matter of degree. Like her doxastic state, the new information is now represented comparatively. Eventually both will have to be represented quantitatively.

Let us return to our example. On Monday Ida receives the information that the weather forecast for Tuesday and Wednesday predicts rain, \( \delta \). In one-dimensional belief revision she picks one of the iterated belief revision methods mentioned above. Then she revises her old belief set \( B \) and entrenchment ordering \( \preceq_B \) for \( B \) to obtain a new belief set \( B^* \delta \) and entrenchment ordering \( \preceq_{B^*\delta} \) for \( B^* \delta \). Different methods return different outputs, but on all of them Ida ends up believing that it will rain on Tuesday, \( \beta \). On Tuesday Ida receives the information that it does not rain after all, \( \neg \beta \). In one-dimensional belief revision Ida proceeds as before.

In two-dimensional belief revision Ida does not merely receive the qualitative information \( \neg \beta \) about Tuesday’s weather. Instead she receives the comparative information \( \gamma \preceq^+ \neg \beta \) about her new doxastic state. This piece of new information says that, in her new entrenchment ordering \( \preceq^+ \), the claim that it does not rain on Tuesday is at least as entrenched as the claim that weather forecasts are always right, indicating that she trusts her sight at least as much as the weatherperson (we could, of course, take a reference sentence other than \( \gamma \)).

Now, there are still several belief revision methods to choose from (see Rott 2009). Among others, this reflects the fact that Ida can respect the constraint \( \gamma \preceq^+ \neg \beta \) by lowering the doxastic status of \( \gamma \), or by raising the doxastic status of \( \neg \beta \). However, the new information now is more specific and leaves less room to be filled by the revision method. It is then only a small step to equip Ida with the complete information exactly where \( \neg \beta \) is located in her new entrenchment ordering.

As mentioned, in two-dimensional belief revision the new information is of comparative form. Therefore, to specify exactly where \( \neg \beta \) is located in the agent’s new entrenchment ordering, the new information must specify the relative position (in the new entrenchment ordering) of \( \neg \beta \) with respect to *every* sentence of her language. That is, the new information must specify for every sentence \( \epsilon \) from \( L \) whether \( \neg \beta \preceq^+ \epsilon \) or \( \neg \beta^+ \sim \epsilon \) or \( \epsilon \preceq^+ \neg \beta \) in the new entrenchment ordering \( \preceq^+ \). In many cases this will require the new information to determine the entire new entrenchment ordering, thus rendering any belief revision method superfluous.
A different way to let the new information specify exactly where ¬β is located in the new entrenchment ordering is to take into account Hume’s “proportions.” Doing so allows us to take the small, but crucial step to equip the ideal doxastic agent with the quantitative information that ¬β is entrenched to a specific degree. In this case the new information determines exactly where ¬β is located in the new “entrenchment ordering” on its own, without the help of a revision method – and without requiring the new information to determine the entire new doxastic state. (Strictly speaking we are not dealing with an entrenchment ordering anymore, as the ideal agent’s old and new doxastic state are now represented quantitatively.)

The step of equipping Ida with the quantitative information that she accepts ¬β a specific degree is taken by Spohn’s (1988) theory of conditional beliefs. It is the same step the Bayesians take when formulating their belief revision methods of Jeffrey and Field conditionalisation (Jeffrey 1983, Field 1978). This step allows the new information to completely specify the ideal agent’s new doxastic attitude towards a sentence or propositional content. The revision method then merely has to incorporate this new information into the ideal agent’s old doxastic state in a consistent way. As indicated in section 2.3, it will do so by requiring her to hold on to those conditional beliefs whose conditions are directly affected by the new information in their entirety. Of course, once the new information is represented quantitatively, the ideal agent’s doxastic state needs to be so represented as well. Spohn (1988)’s ranking functions are such “quantitative entrenchment orderings.”

Before presenting ranking theory let us return to the qualitative approaches of one-dimensional belief revision. Postulates *1-*12 are still compatible with many conflicting belief revision methods. This means the agent’s old doxastic state – her implausibility ordering ≤B with center ⟦B⟧ that represents her revision operator ∗ restricted to her old belief set B (see theorems 3 and 4) – together with the new information α still does not determine her new doxastic state – her implausibility ordering ≤B∗α with center ⟦B*α⟧ that represents her revision operator ∗ restricted to her new belief set B*α. That is, Darwiche & Pearl (1997)’s additional postulates go in the right direction, but they do not go far enough.

Jin & Thielscher (2007) attempt to remedy this situation by employing the notion of doxastic independence. In addition to *1-*12, they require the agent to consider new information β to be independent of a sentence α after revision by β if she considers β to be independent of α before revision by β. In other words, revisions should preserve doxastic independencies. While the idea behind Jin & Thielscher (2007)’s proposal is correct, as we will see, their actual requirement is too strong. The reason is that their notion of doxastic dependence is too strong: too many beliefs are rendered independent of too many other beliefs.
3.3. **ITERATED BELIEF REVISION**

According to Jin & Thielscher (2007), a believed sentence $\alpha$ is independent of another sentence $\beta$ if the believed sentence $\alpha$ is still believed after revision by the negation of the other sentence, $\neg \beta$. However, Ida can receive new information $\neg \beta$ whose negation $\beta$ she considers to be positively relevant to, so not independent of, a belief $\alpha$ of hers without making her give up this belief. For instance, Ida can receive the information $\neg \beta$ that the best player of her team will not be fit for the match without giving up her belief $\alpha$ that her team will win the match – all while considering the information $\beta$ that the best player of her team will be fit for the match to be positively relevant to, so not independent of, her belief $\alpha$ that her team will win the match.

More generally, the ways in which beliefs can depend on each other are many and varied. The qualitative and comparative notions of the AGM theory of belief revision as well as its refinements are too coarse-grained to capture these doxastic dependencies. Ida can receive new information which lowers or raises the doxastic status of one of her beliefs without affecting her merely comparative entrenchment ordering. To illustrate, suppose there is a contingent belief she holds more firmly than any other contingent belief. All her sources of information have testified that this belief is true. Ida then receives the information that one of these sources cannot be trusted after all. Consequently she lowers the doxastic status of this belief. However, she lowers the doxastic status of this belief without bringing it down to the level of any of her other contingent beliefs: said belief is still the most firmly held of her contingent beliefs. The qualitative and comparative notions of belief revision theory are unable to capture this. In order to adequately represent all doxastic dependencies, and to handle iterated belief revisions, we have to go all the way from qualitative belief sets and comparative entrenchment orderings / systems of spheres / implausibility orderings to quantitative ranking functions. Only Hume’s “proportions” get the job done.
Chapter 4

Conditional Belief

In this chapter I will first present the static and dynamic rules of ranking theory which is first developed in Spohn (1988) and discussed at book-length in Spohn (2012). Then I will show how ranking theory solves the problem of iterated belief revisions. This chapter relies on Huber (2013c; invited).

4.1 Ranking theory: static rules

Ranking functions are introduced by Spohn (1988; 1990) to represent qualitative belief. Spohn (2012) is a comprehensive treatise. Ranking theory is quantitative or numerical in the sense that ranking functions assign numbers, so-called ranks, to sentences or propositions. Thus we now take into account Hume’s “proportions.” These numbers are used in the definition of conditional ranking functions which represent conditional belief. As we will see, once conditional ranking functions are defined, we can interpret the axioms of ranking theory in qualitative, albeit conditional terms. The numbers assigned by conditional ranking functions are called conditional ranks. They are defined as differences of non-conditional ranks. In contrast to this, conditional probabilities are defined as ratios of non-conditional probabilities.

Instead of taking the objects of belief to be sentences of a formal language it is both more general and more convenient to take them to be propositions of an algebra over a non-empty set of possible worlds. As we will see in section 6.1, this does not commit us to the assumption that the ideal doxastic agent is “logically omniscient” in the sense that she is certain of all sentences that are logically true, and has the same doxastic attitude towards sentences that are logically equivalent.
The notion of an algebra presupposes the “minimal conceptual structure” of propositional contents alluded to in chapter 1. It does so by presupposing that we can meaningfully speak of the negation or complement of a set of possible worlds with respect to another set of possible worlds; the conjunction or intersection of two or more sets of possible worlds; as well as the disjunction or union of two or more sets of possible worlds. We will assume that the ideal doxastic agent understands the set of all possible worlds, the complement (with respect to the set of all possible worlds) of any set of possible worlds she understands, as well as the intersections and unions of any sets of possible worlds she understands.

The algebra of propositions represents the ideal doxastic agent’s language, although we use this term in a slightly non-standard way: the agent’s language is a “language of thought” that consists of those sets of possible worlds that she understands and has an opinion on in the weak sense of section 2.2 that includes suspension of judgment. It is defined as follows. A set of subsets of a non-empty set $W$, $\mathcal{A}$, is an algebra over $W$ if, and only if,

(i) the set of all possible worlds $W$ is a proposition in $\mathcal{A}$,

(ii) if $A$ is a proposition in $\mathcal{A}$, then the complement or negation of $A$, $W \setminus A = \bar{A}$, is also a proposition in $\mathcal{A}$, and

(iii) if both $A$ and $B$ are propositions in $\mathcal{A}$, then the union or disjunction of $A$ and $B$, $A \cup B$, is also a proposition in $\mathcal{A}$.

We say that an algebra $\mathcal{A}$ over $W$ is a $\sigma$-algebra if, and only if, the following holds for every countable set $\mathcal{B}$ of subsets of $W$: if all elements of $\mathcal{B}$ are propositions in $\mathcal{A}$, i.e. if $\mathcal{B} \subseteq \mathcal{A}$, then the union or disjunction of the elements of $\mathcal{B}$, $\bigcup \mathcal{B}$, is also a proposition in $\mathcal{A}$. Finally, we say that an algebra $\mathcal{A}$ over $W$ is complete if, and only if, the following holds for every (countable or uncountable) set $\mathcal{B}$ of subsets of $W$: if all elements of $\mathcal{B}$ are propositions in $\mathcal{A}$, i.e. if $\mathcal{B} \subseteq \mathcal{A}$, then the union or disjunction of the elements of $\mathcal{B}$, $\bigcup \mathcal{B}$, is also a proposition in $\mathcal{A}$. The power-set of a non-empty set of possible worlds $W$, $\mathcal{P}(W)$, is a complete algebra over $W$.

The algebra is the set of sets of possible worlds the agent understands and has an opinion on. As a consequence, the three clauses in its definition become three normative requirements. First, the agent is required to have an opinion on the set of all possible worlds. Second, if she has an opinion on a set of possible worlds, then she is required to also have an opinion on its negation. Third, if she has an opinion on two or more sets of possible worlds, then she is required to also have an opinion on their disjunction (and, hence, conjunction). The three clauses are not requirements of understanding because of our assumption from above.
4.1. RANKING THEORY: STATIC RULES

There is one more notion that we need. A set of subsets of $W$, $\mathcal{P}$, is a partition of $W$ if, and only if,

(i) all elements of $\mathcal{P}$ are non-empty,

(ii) any two distinct elements $A$ and $B$ of $\mathcal{P}$ are mutually exclusive, i.e. $A \cap B = \emptyset$ if $A \neq B$, and

(iii) the elements of $\mathcal{P}$ cover $W$ in the sense that $W \subseteq \bigcup \mathcal{P}$.

The elements of a partition are called “cells.” Every partition of a non-empty set of possible worlds $W$ generates a unique complete algebra over $W$. Its elements are all unions of one or more cells, plus the empty set.

Suppose we have a formal language $\mathcal{L}$, and $W_{\mathcal{L}}$ is the set of all models or truth value assignments for $\mathcal{L}$. Then $\mathcal{A} = \{ [\alpha] \subseteq W_{\mathcal{L}} : \alpha \in \mathcal{L} \}$ is an algebra of propositions over $W_{\mathcal{L}}$ (if the truth values are assigned in the standard way), or it generates such an algebra (this is so even for the non-standard truth value assignment from section 6.1). This algebra in turn generates a unique smallest $\sigma$-algebra, and a unique smallest complete algebra. This means that for every formal language of sentences there is an algebra of propositions over the set of models of the formal language. Since the converse is not true, the semantic framework of propositions is more general than the syntactic framework of sentences.

As should become clear, it is also more convenient. Consider a non-empty set of possible worlds $W$ and an algebra of propositions $\mathcal{A}$ over $W$. A function $\varrho$ from $\mathcal{A}$ into the set of natural numbers $\mathbb{N}$ extended by $\infty$, $\mathbb{N} \cup \{\infty\}$, is a ranking function on $\mathcal{A}$ if, and only if, for all propositions $A$ and $B$ from $\mathcal{A}$:

\[
\varrho(W) = 0 \quad (4.1)
\]
\[
\varrho(\emptyset) = \infty \quad (4.2)
\]
\[
\varrho(A \cup B) = \min \{ \varrho(A), \varrho(B) \} \quad (4.3)
\]

As in probability theory, if $\mathcal{A}$ is a $\sigma$-algebra, the third axiom can be strengthened to countable unions. The resulting ranking function is called “countably minimitive.” In contrast to probability theory, if $\mathcal{A}$ is a complete algebra, the third axiom can even be strengthened to arbitrary, i.e. possibly uncountable, unions. The resulting ranking function is called “completely minimitive.”

For a non-empty or consistent proposition $A \neq \emptyset$ from $\mathcal{A}$ the conditional ranking function $\varrho(\cdot | A) : \mathcal{A} \to \mathbb{N} \cup \{\infty\}$ based on the (non-conditional) ranking function $\varrho(\cdot) : \mathcal{A} \to \mathbb{N} \cup \{\infty\}$ is defined by the difference formula:

if $\varrho(A) < \infty$, then $\varrho(\cdot | A) = \varrho(\cdot \cap A) - \varrho(A)$.
For the case of $\varrho(A) = \infty$, Goldszmidt & Pearl (1996: 63) suggest infinity – while Huber (2007a: 517) suggests zero – as value of $\varrho(B \mid A)$, for all propositions $B$ from $\mathcal{A}$. Still considering the case of $\varrho(A) = \infty$, Huber (2006: 464) suggests zero as value of $\varrho(B \mid A)$ for all non-empty propositions $B$ from $\mathcal{A}$, and then stipulates $\varrho(\emptyset \mid A) = \infty$ to ensure that conditional ranking functions are ranking functions. Yet another option for this case is to define $\varrho(B \mid A)$ as zero if $A \subseteq B$, and as infinity otherwise. Raidl (ms1) contains a careful discussion of these options and corrects mistakes by Huber (2014a; 2015a; 2016a).

Spohn (2012: 79) briefly considers a rank-theoretic version of Popper-Rényi functions (Popper 1955, Rényi 1955) that takes conditional ranks as primitive and axiomatizes them rather than non-conditional ranks, as we have done. However, ultimately he restricts the conditions to those with a finite rank, as we will do. This means conditional ranks are undefined when the condition has infinite rank – much like (classical) conditional probabilities are undefined when the condition has probability zero.

As the following makes clear, this restriction is not very restrictive. A ranking function $\varrho$ is regular if, and only if, for all consistent propositions $A$ from $\mathcal{A}$:

$$\varrho(A) < \varrho(\emptyset) = \infty. \quad (4.4)$$

In contrast to probability theory (Hájek ms2), it is always possible to define regular ranking functions, no matter how rich or fine-grained the underlying algebra of propositions. Thus conditional ranks can be defined for all consistent conditions.

Doxastically, ranks are interpreted as grades of disbelief. An agent disbelieves a proposition $A$ if, and only if, she assigns a positive rank to $A$, $\varrho(A) > 0$. The agent assigns rank zero to propositions she does not disbelieve. However, this does not mean she believes these propositions. Instead, belief in some proposition is characterized as disbelief in its negation: an agent believes a proposition $A$ if, and only if, she disbelieves its negation $\overline{A}$, $\varrho(\overline{A}) > 0$. An ideal doxastic agent suspends judgment with respect to a proposition (and its negation) if, and only if, both the proposition and its negation are assigned rank zero.

Conditional ranks are interpreted as grades of conditional disbelief. An agent disbelieves a proposition $A$ conditional on a proposition $C$ with finite rank if, and only if, she assigns a positive rank to $A$ conditional on $C$, $\varrho(A \mid C) > 0$. She believes $A$ conditional on $C$ if, and only if, she disbelieves $\overline{A}$ conditional on $C$, $\varrho(\overline{A} \mid C) > 0$. She suspends judgment with respect to $A$ (and its negation) conditional on $C$ if, and only if, she assigns conditional rank zero to both $A$ and $\overline{A}$.
4.1. RANKING THEORY: STATIC RULES

Finally, an agent is (conditionally) certain of a proposition if, and only if, she assigns infinite (conditional) rank to its negation. She (conditionally) deems a proposition possible if, and only if, she assigns a finite (conditional) rank to it.

An agent believes a proposition if, and only if, she believes it conditional on the tautological proposition $W$: $\varrho(\cdot) = \varrho(\cdot | W)$. Thus rank-theoretic conditional belief generalizes rank-theoretic belief. It does so in the same way as probabilistic conditional degree of belief generalizes probabilistic degree of belief. In contrast to probability theory, where the probability of $A$ is determined by the probability of $\overline{A}$, the rank of $\overline{A}$ is not, in general, determined by the rank of $A$.

We can reformulate the doxastic interpretation of the axioms of ranking theory in qualitative terms by assuming the definition of a conditional ranking function – that is, by assuming that conditional ranks are numbers from $\mathbb{N} \cup \{\infty\}$ that are defined as differences of non-conditional ranks as above. This implies almost half of the third axiom even in its strongest form: $\varrho(\bigcup B) \leq \min \{\varrho(A) : A \in B\}$ if $\varrho(\bigcup B) < \infty$. This inequality requires the agent to (conditionally) disbelieve all disjuncts of a disjunction she (conditionally) disbelieves provided she deems the condition possible. For $\varrho(\bigcup B) = \infty$ we get this inequality by requiring the agent to be certain of all its conjuncts $\overline{A}$ if she is certain of a conjunction $\bigcap \overline{\{A : A \in B\}}$. The latter implies that, conditional on any condition she deems possible, the agent should be conditionally certain of all its conjuncts if she is conditionally certain of a conjunction. Thus, unlike conditional belief, conditional certainty is not more demanding than its non-conditional counterpart.

Part, but not all, of what the other half or inequality of the third axiom says is that the agent should disbelieve a disjunction $A \cup B$ if she disbelieves both its disjuncts $A$ and $B$. Given the definition of a conditional ranking function, the third axiom extends this requirement to conditional beliefs. For any proposition $C$ she deems possible, the agent should disbelieve a disjunction $A \cup B$ conditional on $C$ if she disbelieves $A$ conditional on $C$ and she disbelieves $B$ conditional on $C$. In addition the third axiom requires the agent to be certain of a conjunction $\overline{A} \cap \overline{B}$ if she is certain of both its conjuncts $\overline{A}$ and $\overline{B}$. (As above, this implies that she should be certain of a conjunction $\overline{A} \cap \overline{B}$ conditional on $C$ if she is certain of $\overline{A}$ conditional on $C$ and she is certain of $\overline{B}$ conditional on $C$.) The strengthenings of the second half or inequality of the third axiom extend these requirements to countable and arbitrary disjunctions and conjunctions, respectively. For any proposition $C$ that she deems possible, the agent should disbelieve a disjunction $\bigcup B$ conditional on $C$ if she disbelieves, conditional on $C$, each disjunct $A$ from $B$; and she should be certain of a conjunction $\bigcap \overline{\{A : A \in B\}}$ if she is certain of all its conjuncts $\overline{A}$. 
The first axiom says that the ideal doxastic agent should not disbelieve the tautological proposition. The second axiom says that she should not deem the contradictory proposition possible. The fourth axiom, regularity, requires her to deem any consistent proposition possible so that she has conditional (dis)beliefs for all consistent conditions. Thus, given the definition of a conditional ranking function, we can formulate the axioms of ranking theory in qualitative terms.

According to the first axiom, the agent should not disbelieve $A \cup \overline{A}$. The third axiom then yields that she should not simultaneously believe and disbelieve $A$, for any proposition $A$. This is part one of our non-conditional consistency requirement from section 2.3. According to the definition of a conditional ranking function, the agent should not disbelieve $A \cup \overline{A}$ conditional on any condition she deems possible. The third axiom then yields that, for any proposition $A$, she should not simultaneously believe and disbelieve $A$ on any condition she deems possible. This is part one of our conditional consistency requirement from section 2.3 for conditions the agent deems possible. Part two of both requirements follows from the fact that a ranking function is a function.

In fact, given the definition of a conditional ranking function, part one of the conditional consistency requirement from section 2.3 is just about all that is required by the axioms of ranking theory. As noted above, the definition of a conditional ranking function implies that $\varrho(\bigcup B) \leq \min \{\varrho(A) : A \in B \}$ if $\varrho(\bigcup B) < \infty$. Part one of the conditional consistency requirement from section 2.3 implies that $\min \{\varrho(A | \bigcup B) : A \in \mathcal{A} \} = 0$ if $\varrho(\bigcup B) < \infty$. The definition of a conditional ranking function turns this into $\min \{\varrho(A) : A \in \mathcal{A} \} \leq \varrho(\bigcup B)$ if $\varrho(\bigcup B) < \infty$. This gives us the third axiom under the condition that $\varrho(\bigcup B) < \infty$. The additional requirement that the agent should be certain of a conjunction $\bigcap \{\overline{A} : A \in B \}$ if, and only if, she is certain of all its conjuncts $\overline{A}$, gives us the third axiom under the condition that $\varrho(\bigcup B) = \infty$. The remaining axioms follow from the additional requirements that the agent should not disbelieve the tautological proposition, and that she should be certain of it (and it alone).

The choice of the natural numbers cannot be obtained in this way. Instead, we have to point to the third axiom which, in its strongest version, requires the co-domain of a ranking function to be well-ordered. Since we also need at least infinitely many different numbers, the choice of the natural numbers plus infinity is the simplest means to attain this end. However, there may be other ends that require different choices (Kroedel & Huber 2013: fn. 12 mention one). Therefore, it is worth mentioning that Spohn (1988) develops ranking theory with the class of ordinals, and Spohn (2012: 72) with the set of real numbers plus infinity.
Finally, suppose a ranking function is not a function, i.e. a functional relation between propositions and numbers. Then some proposition is not related to any number – or more than one. In the latter case the agent simultaneously believes and refrains from believing this proposition conditional on itself (if at least two of the numbers are finite), or she simultaneously is certain of and refrains from being certain of its negation (if one of the numbers is infinite). The former case cannot arise because the algebra is the agent’s language, i.e. the set of propositions she (understands and) has an opinion on. Thus, given the definition of a conditional ranking function, the conditional consistency requirement from section 2.3 is just about all that the axioms of ranking theory require.

Ranks are numbers. However, unlike probabilities, which are measured on an absolute scale, ranks do not utilize all the information carried by these numbers. Instead, ranks are at best measured on a ratio scale (Hild & Spohn 2008) – at best, for even the choice of zero as threshold for disbelief is somewhat arbitrary, as Spohn (2015: 9) notes (but see Raidl ms2 for subtle differences for conditional belief). This is perhaps clearest if we consider what Spohn (2012: 76) calls $\varrho$’s two-sided ranking function $\tau$. This is a function from $\mathcal{A}$ into the set of integers $\mathbb{Z}$ extended by $\infty$ and $-\infty$, $\mathbb{Z} \cup \{\infty\} \cup \{-\infty\}$, that is defined in terms of $\varrho$ as follows: for all proposition $A$ from $\mathcal{A}$,

$$\tau(A) = \varrho(\overline{A}) - \varrho(A).$$

Ranking functions and two-sided ranking functions are interdefinable:

$$\varrho(\cdot) = -\min\{\tau(\cdot), 0\}.$$  

Two-sided ranking functions are more difficult to axiomatize. However, they may be more intuitive, as they characterize belief in positive terms: a proposition is (conditionally) believed if, and only if, its two-sided (conditional) rank is positive. Interestingly, any other non-negative, finite threshold equally gives rise to a notion of belief (that is consistent and deductively in the sense of the next chapter): a proposition $A$ is believed if, and only if, its two-sided rank is greater than some threshold $n$, $\tau(A) > n$. This means that ranking theory validates the “Lockean thesis” (Foley 2009, Hawthorne 2009) according to which an agent should believe a proposition if, and only if, her grade or degree of belief is sufficiently high. Furthermore, while it may appear unfair to reserve infinitely many numbers for belief (and for disbelief), and only the number zero for suspension of judgment, we now see that this may be changed by adopting a threshold other than zero. Of course, there are still only finitely many levels for suspension of judgment, and infinitely many levels for belief (and for disbelief).
4.2 Ranking theory: dynamic rules

Interpreted doxastically, the axioms of ranking theory are static norms for how the ideal doxastic agent should organize her beliefs and conditional beliefs at a given moment in time. These axioms are supplemented by three dynamic norms for how she should update or revise these beliefs across time if she receives new information of various formats. (One may argue that these norms are kinematic rather than dynamic. For we assume rather than require the agent to accept the new information she receives, and I do not say anything about what causes the agent to receive new information. The reason I call the norms dynamic is that they depend on the end we assume the agent to have.) These update rules are not competitors, though. Instead, the first update rule is just a special case of the second, and the third update rule can be defined in terms of the second.

The first update rule is defined for the case where the new information comes in the form a certainty. It mirrors the update rule of strict conditionalization from probability theory (Vineberg 2000).

**Update Rule 1 (Plain Conditionalization, Spohn 1988)** Suppose $\varrho(\cdot) : \mathcal{A} \to \mathbb{N} \cup \{\infty\}$ is the ideal doxastic agent’s ranking function at time $t$, and she deems $E$ and $\overline{E}$ from $\mathcal{A}$ possible at $t$. Suppose further between $t$ and $t'$ her ranks for $E$ and $\overline{E}$ are directly affected and she becomes certain of $E$, but no logically stronger proposition. Finally, suppose her doxastic state is not directly affected in any other way such as forgetting, a change in her ends, etc. Then her ranking function at time $t'$ should be $\varrho_E(\cdot) = \varrho(\cdot | E)$.

Plain conditionalization asks the ideal doxastic agent to revise her old ranking function by holding on to those conditional beliefs whose condition is the most specific, i.e. the logically strongest, proposition she becomes certain of, subject to the constraint that her new doxastic state is a ranking function. Therefore plain conditionalization satisfies the principle of categorical matching. In other terms one could perhaps say that plain conditionalization has the agent revise her beliefs by holding on to every “inferential belief” whose premise is the logically strongest proposition she becomes certain of as a result of some experiential event that is not under her doxastic control.

The second update rule is defined for the case where the new information comes in the form of new ranks for the elements of a partition. It mirrors the update rule of Jeffrey conditionalization from probability theory (Jeffrey 1983) and generalizes plain conditionalization in the same way as the latter generalizes strict conditionalization.
Update Rule 2 (Spohn Conditionalization, Spohn 1988) Suppose $\varrho(\cdot) : \mathcal{A} \to \mathbb{N} \cup \{\infty\}$ is the ideal doxastic agent’s ranking function at time $t$, and she deems all cells of the experiential partition $\{E_i \in \mathcal{A} : i \in I\}$ possible at $t$. Suppose further between $t$ and $t'$ her ranks on this partition are directly affected and change to $n_i \in \mathbb{N} \cup \{\infty\}$, where $\min \{n_i : i \in I\} = 0$. Finally, suppose her doxastic state is not directly affected on any finer partition, or in any other way such as forgetting, a change in her ends, etc. Then her ranking function at time $t'$ should be

$$
\varrho_{E_i \to n_i}(\cdot) = \min_{i \in I} \{\varrho(\cdot \mid E_i) + n_i\}.
$$

Spohn conditionalization asks the ideal doxastic agent to revise her old ranking function by holding on to those conditional beliefs whose condition is one of the most specific propositions, or cells, whose doxastic standing has changed as a result of an experiential event that is not under her doxastic control (subject to the constraint that her new doxastic state is a ranking function). The restriction to hold fixed only those conditional beliefs whose condition is one of these most specific propositions whose doxastic standing has been directly affected is important.

Ida prefers red wine to white wine to no wine. She believes that there is red wine left if there is wine left at all, and that she will have red (but not white) wine tonight if she has wine at all. She also believes that she has white (but not red) wine tonight given that there is white wine left, but no red wine. When she enters the cellar she seems to perceive that there is white wine left, but no red wine. The propositional content of the new information Ida receives is that there is white wine left, but no red wine. How firmly she comes to believe this propositional content depends on how reliable or trustworthy she deems her perception to be on this particular occasion. That there is white wine left, but no red wine logically implies that there is wine left. Ida also comes to believe this second propositional content (as she should according to Spohn conditionalization).

In this case Ida should hold on to her conditional belief that she will have red (but not white) wine tonight if there is red wine left, but no white wine. She should not also hold on to her conditional belief that she will have white (but not red) wine tonight if there is wine left. Otherwise she ends up having inconsistent beliefs! The same is true if Ida does not merely come to believe, but becomes certain that there is red wine left, but no white wine. This is the reason for the restriction in plain conditionalization to hold fixed only those conditional beliefs whose condition is the logically strongest proposition the agent becomes certain of. It is also the reason for the restriction in Spohn conditionalization to hold fixed only those conditional beliefs whose condition belongs to the least fine-grained or comprehensive experiential partition on which her ranks are directly affected.
The above illustrates that plain conditionalization is the special case of Spohn conditionalization where the experiential partition consists of $E$ and $\overline{E}$ and the new ranks are zero and infinity, respectively. Thus, really we are dealing with one update rule so far.

The third update rule is defined for the case where the new information reports the differences between the old and new ranks for the elements of a partition. It mirrors the update rule of Field conditionalization from probability theory (Field 1978) and is further developed in Bewersdorf (2013).

**Update Rule 3 (Shenoy Conditionalization, Shenoy 1991)** Suppose $\varrho(\cdot): \mathcal{A} \to \mathbb{N} \cup \{\infty\}$ is the ideal doxastic agent’s ranking function at time $t$, and she deems all cells of the experiential partition $\{E_i \in \mathcal{A} : i \in I\}$ possible at $t$. Suppose further between $t$ and $t'$ her ranks on this partition are directly affected and change by $z_i \in \mathbb{N} \cup \{\infty\}$, where $\min \{z_i : i \in I\} < \infty$. Finally, suppose her doxastic state is not directly affected on any finer partition, or in any other way such as forgetting, a change in her ends, etc. Then her ranking function at time $t'$ should be

$$
\varrho_{E_i \uparrow z_i}(\cdot) = \min_{i \in I} \{\varrho(\cdot \cap E_i) + z_i - m\}, \quad \text{where} \quad m = \min_{i \in I} \{z_i + \varrho(E_i)\}.
$$

In contrast to Field conditionalization, whose input parameters do not behave like probabilities, the input parameters of Shenoy conditionalization behave like ranks. It is required that at least one of the $z_i$ be finite. Hence they can be “normalized” (by defining $z_i' = z_i - \min_{i \in I} \{z_i\}$) so that at least one of them is zero. Then they can be interpreted as the values of a ranking function (on the algebra generated by the experiential partition), and updating as the combination of two ranking functions.

Assume that the algebra $\mathcal{A}$ is the power-set $\varrho(W)$ of the non-empty set of all possible worlds $W$ so that each possible world $w$ in $W$ has its own rank, namely the rank assigned to the singleton proposition $\{w\}$. Spohn conditionalizing $E$ and $\overline{E}$ to 0 and $n$, respectively, keeps the relative positions of all possible worlds in $E$ and all possible worlds in $\overline{E}$ fixed. It improves the rank of $E$ to 0 – remember: low numbers represent low grades of disbelief – and it changes the rank of $\overline{E}$ to $n$. In contrast to this, Shenoy conditionalizing $E$ and $\overline{E}$ by $0$ and $z$, respectively, improves the possible worlds within $E$ by $z$, as compared to the possible worlds in $\overline{E}$. $m$ is a normalization parameter. It ensures that the tautological proposition $W$ is assigned rank zero so that Shenoy conditionalization yields a ranking function satisfying the first axiom.

Shenoy conditionalization is definable in terms of Spohn conditionalization:

$$
\varrho_{E_i \uparrow z_i}(\cdot) = \varrho_{E_i \rightarrow n_i}(\cdot), \quad \text{where} \quad n_i = z_i + \varrho(E_i) - m.
$$

Thus, really we are still dealing with one update rule. I will call it the update rule.
Spohn conditionalization is result-oriented: the numbers \( n_i \) characterize the result of the experiential event on the agent’s ranks for the cells \( E_i \): \( q_{E_i \rightarrow n_i} (E_i) = n_i \). These new ranks depend in part on the agent’s initial ranks, which is why the numbers \( n_i \) do not characterize the impact of the experiential event independently of the agent’s initial beliefs.

In contrast to this, the numbers \( z_i \) in Shenoy conditionalization characterize the impact of the experiential event independently of the agent’s initial beliefs (except for the scale of the latter). They do so in the sense that the rank of \( E_i \) is deteriorated by \( z_i - \min_{i \in I} \{ z_i \} \) relative to the rank \( \min_{i \in I} \{ z_i \} \) of the “best” cell. However, the best cell is not, in general, the cell with the lowest initial rank. It is so only in the special case of two cells \( E \) and \( \bar{E} \) and parameters 0 and \( z \), respectively. In this special case it holds that \( \tau_{E \uparrow 0, E \uparrow z} (E) - \tau (E) = z \) and \( \tau_{E \uparrow 0, \bar{E} \uparrow z} (\bar{E}) - \tau (\bar{E}) = -z \). The latter need not be true anymore when the evidential partition contains more than two propositions. That is, the grade to which the experiential event confirms the cell \( E_i \) is not, in general, determined by the parameters \( z_i \).

Like probabilistic degrees of incremental confirmation (Huber 2005a, 2008a), rank-theoretic grades of confirmation are defined as differences between prior and posterior grades of belief: \( \tau_{E_i \uparrow z_i} (A) - \tau (A) \) is the grade to which the experiential event confirms the proposition \( A \). Except for the special case where the evidential partition consists of two cells and we consider the grade of confirmation of one of them, these differences depend on the parameters \( z_i \) and the agent’s prior grades of disbelief. Indeed, they do so twice. Ranks are not measured on an absolute scale. Thus the parameters \( z_i \) (and, hence, grades of confirmation) are meaningful only relative to the scale on which the agent’s prior grades of disbelief are measured. This is a remarkable difference to the probabilistic case. The input parameter \( \alpha_i \) of Field conditionalization is positive if, and only if, the experiential event confirms, i.e. raises the probability of, the cell \( E_i \). Therefore these parameters, unlike those of Shenoy conditionalization, can be interpreted as degrees of confirmation.

Plain, Spohn, and Shenoy conditionalization characterize the new information numerically. This reflects the fact that the quality of new information varies with how reliable or trustworthy the agent deems its source: it makes a difference if the weatherperson Ida has never met predicts that it will rain, if a friend Ida trusts tells her so, or if Ida seems to see for herself that it is raining. Simplifying somewhat, in each case the propositional content Ida comes to believe is that it is raining, but the effect of the new information on her old beliefs will be a different one in each case. The difference in how reliable or trustworthy Ida deems the sources of information is reflected in the numbers accompanying this propositional content.
CHAPTER 4. CONDITIONAL BELIEF

In reality things are more complicated. The propositions the experiential event causes the agent to come to believe to various grades may not be about reality, but about what reality appears or seems to be. Then, the source of information the agent deems reliable or trustworthy to some degree may not the weatherperson, but the audio-visual apparatus that is exposed to the weather report on a particular occasion. Furthermore, these experiences may, in general, include all her senses, and not just her auditory and visual systems. Finally, the complete “contents” of these experiences may not be expressible by a proposition in her language.

Fortunately none of these simplifying assumptions are essential. All that is necessary is that the agent can identify all propositional contents whose doxastic standing is directly affected by the experience she undergoes. Since any such set of propositions generates a unique least fine-grained or comprehensive partition, each experience generates exactly one experiential partition. Its elements are the propositions that have been directly affected by the experience in their entirety. Of course, we humans find it difficult to determine by introspection what we have experienced directly, and what we have subsequently inferred from experience (and we rarely, if ever, report the contents of our experiences in undigested form). However, this anti-luminosity or intransparency is our problem, not the theory’s.

The axioms of ranking theory ask of the agent that her conditional beliefs be conditionally consistent. The update rule of ranking theory asks of the agent that they remain so. It does so by asking her to hold on to those conditional beliefs whose conditions are directly affected by the new information in their entirety.

Now that Hume’s “proportions” have been taken into account, and we have made explicit the dependence on the agent’s language or algebra of propositions, we can be a bit more precise. The update rule asks the agent to hold on to those conditional grades of disbelief whose conditions are elements of the experiential partition: \( \varrho(\cdot \mid E_i) = \varrho_{E_i \rightarrow n_i}(\cdot \mid E_i) \) for all cells \( E_i \) in the experiential partition. In fact, together with the requirement that \( \varrho_{E_i \rightarrow n_i}(E_i) = n_i \), this is an alternative, but equivalent formulation of Spohn conditionalization.

The experiential partition has to be the least fine-grained or comprehensive partition. It needs to list all conditions that are directly affected in their entirety by the experiential event. This is, of course, but a variant of Carnap (1947b)’s “principle of total evidence.”

Ida believes that there is red wine left if there is wine left at all, and that she will have red (but not white) wine tonight if she has wine at all. She also believes that she has white (but not red) wine tonight given that there is white wine left, but no red wine. She receives the information \( E \) that there is white wine left, but no red wine.
The update rule asks Ida to hold onto her grades of belief conditional on $E$ – that is, conditional on there being white wine left, but no red wine. It also asks her to hold onto her grades of belief conditional on the negation of $E$, $\overline{E}$ – that is, conditional on the disjunctive assumption of there being red wine left, or there being no white wine left. In particular, Ida should hold onto her grade of belief that she has white (but not red) wine tonight given that there is white wine left, but no red wine. However, Ida is not required to hold onto her grade of belief that she has red (but not white) wine tonight if she has wine at all. Indeed, depending on the details of Ida’s doxastic state, she may well be required to give up this conditional belief and adopt the new conditional belief that she has white (but not red) wine tonight given that she has wine at all.

It is worth noting that the update rule goes beyond the requirement that the agent should hold on to her qualitative conditional beliefs. It is possible for her to believe the same cells of the experiential partition (to the same grades), hold the same qualitative beliefs conditional on all cells of the experiential partition before and after a revision of her ranking function, yet violate Spohn conditionalization. In contrast to the axioms of ranking theory, its update rule cannot be stated in qualitative terms, not even conditional ones. Hume’s “proportions” matter.

Throughout the previous chapters I have stressed that we are assuming the agent to accept the new information she receives, rather than requiring her to do so. We have now seen what form the new information takes: it comes in the form of new ranks for the cells of a partition all of which are elements of the agent’s language that she has deemed possible initially. We will see in section 6.1 what the agent should do if these propositions do not belong to her language.

The experiential event determines which propositions are in the experiential partition. The new grades to which the agent disbelieves these propositions merely depend on the experiential event. They are not determined by it. This is so because these new grades of disbelief also depend on the agent’s initial grades of disbelief, as well as the grades to which she deems the source of information trustworthy. The latter deemings are determined by the experiential event – except for their scale, and except that the agent has to organize them as she has to organize her beliefs: consistently in the sense that $\min_{i \in I} \{z_i\} < \infty$. Otherwise they cannot be used as input. It is here, then, where causes turn into reasons. The deemings are caused by the experiential event, but to use them as reasons or input for the update rule, the agent has to organize them consistently. This means our assumption that the agent accepts the new information she receives contains an important element of normativity: to reason with them, the agent must organize the experientially caused deemings in a consistent manner.
It is also worth noting that the source of information is to be understood as token rather than type: what matters is the report of the source on a particular occasion, not some generic source. How trustworthy the agent deems the report of a source of information is determined by the collection of all parameters $z_i$.

Finally, the update rule satisfies the principle of categorical matching: it turns an old ranking function and new information into a new ranking function. Given regularity, we can assume the old ranking function to be regular, as well as all input parameters of the new information to be finite. In this case the update rule yields a new ranking function that is still regular. Hence the update rule satisfies the principle of categorical matching twice: it turns ranking functions and new information into ranking functions; and, it turns regular ranking functions and new information whose input parameters are finite into regular ranking functions.

Figure 1 for section 4.3:
4.3 Iterated belief revision revisited

In ranking theory, the ideal agent’s old doxastic state is represented by a ranking function. The latter numerically represents the strength of her beliefs. In addition it determines her belief set. The new information is also represented quantitatively. It completely specifies the agent’s new doxastic attitude towards the propositions in the experiential partition. Thus ranking theory fully takes into account Hume’s “proportions.”

Furthermore, the update rule of ranking theory satisfies Gärdenfors & Rott (1995)’s principle of categorical matching twice: the ideal agent’s doxastic state is represented as a (regular) ranking function before and after the revision process. It does so by requiring the agent to hold onto those conditional grades of disbelief whose conditions are directly affected in their entirety by some experiential event, subject to the constraint that the new doxastic state is another ranking function. Specifically, the agent’s old ranking function and the new information she receives determine her new ranking function. If we require regularity, as we do, the update rule is in a position to handle indefinitely iterated belief revisions. Let us see how.

Ida’s ranking function $R$ will assign a positive rank to the proposition $\neg A$ that it will not be sunny on Wednesday. Consequently her rank for $A$ will be zero. This is so because the first and third axiom imply that, for any proposition $X$, at least one of $X$ and $\neg X$ has rank zero. $R$ will assign a greater rank to the proposition $\neg B$ that it will not rain on Tuesday. Finally, $R$ will assign an even greater rank to the proposition $\neg C$ that weather forecasts are not always right. In other words, where we previously had $\bot \prec \alpha \prec \beta \prec \gamma$, we now have $0 < R(\neg A) < R(\neg B) < R(\neg C)$.

This holds true in general. For a regular ranking function $R$, the relation

$$
\alpha \preceq_R \beta \quad \text{if, and only if,} \quad R(\llbracket \alpha \rrbracket) \leq R(\llbracket \beta \rrbracket)
$$

is an entrenchment ordering for the belief set $B_R = \{ \gamma \in \mathcal{L} : R(\llbracket \gamma \rrbracket) > 0 \}$. This shows that, when the assumption of regularity is made, ranking theory satisfies the AGM postulates for belief revision. We can use the results mentioned in the previous chapter to reformulate this, where I continue to assume that $R$ is regular.

The set of propositions $S_R = \{ R^{-1}(n) \subseteq W : n \in \mathbb{N} \}$ is a system of spheres in $W$ centered on $R^{-1}(0)$, where $R^{-1}(n) = \{ w \in W : R(\llbracket w \rrbracket) = n \}$ is the set of possible worlds that are assigned rank $n$. Here I make the simplifying assumption that the algebra of propositions is the power set of $W$, $\wp(W)$. If this assumption is not made, the definitions of the system of spheres $S_R$, and of the implausibility ordering $\preceq_R$ below, are slightly more complicated.
An agent with ranking function $R$ believes a proposition $A$ if, and only if, $R^{-1}(0) \subseteq A$, where $R^{-1}(0) = \{w \in W : R(\{w\}) = 0\}$ is called her belief core. The relation

$$w \leq_R w' \text{ if, and only if, } R(\{w\}) \leq R(\{w'\})$$

is an implausibility ordering on $W$ whose center is the set of least implausible worlds, $R^{-1}(0) = \{v \in W : R(\{v\}) \leq R(\{w\}) \text{ for all } w \in W\}$. The latter is again the agent’s belief core. It is equal to the conjunction or intersection of all of her beliefs, $\bigcap \{C \subseteq W : R(C) > 0\}$.

The update rule of ranking theory also satisfies the four additional postulates for iterated belief revision proposed by Darwiche & Pearl (1997). This can be verified (see Spohn 2012: 91ff) by checking that the four postulates $\leq_5 \leq_8$ hold for $\leq_R$ and $\leq_{R^*}$, where $R^*$ is any ranking function that results from $R$ by what we may call a “Spohn shift” on the proposition $E$, i.e. an application of Spohn conditionalization to the experiential partition $\{E, \overline{E}\}$ with input parameters $R^*(E) = 0$ and $R^*(\overline{E}) > 0$.

On Monday Ida receives the information that the weather forecast for Tuesday and Wednesday predicts rain. In order for Spohn conditionalization to tell her how to revise her beliefs, she has to tell us how firmly she now, on Monday, disbelieves the proposition $D$ that it is not the case that the weather forecast for Tuesday and Wednesday predicts rain. As an approximation (more on this in the next chapter), it suffices if we determine how many information sources saying $D$ it would now, on Monday, take for her to give up her disbelief in $D$ – as compared to how many information sources saying $Y$ it would have taken earlier, before Monday, for her to give up her disbelief in $\overline{Y}$ for $Y = A, \overline{A}, B, \overline{B}, C, \overline{C}, D, \overline{D}$.

Suppose Ida’s old ranks are $R(A) = 1$, $R(D) = 2$, $R(B) = 5$, and $R(C) = 7$, and her new rank for $\overline{D}$ is $R^*(\overline{D}) = 13$. According to Spohn conditionalization, Ida’s other new ranks are:

$$R^*(Y) = \min \left\{ R(Y | D) + 0, R\left(Y | \overline{D}\right) + 13 \right\}$$

To calculate Ida’s new ranks $R^*(Y)$ we thus need to have her old conditional ranks $R(Y | D)$ and $R\left(Y | \overline{D}\right)$ as well as her new ranks for the conditions $D$ and $\overline{D}$. This in turn requires us to have her old ranks for various conjunctions. Suppose the numbers, Hume’s “proportions”, are as follows:
Then Ida’s new ranks are \( R^* (\overline{C}) = 6 \), \( R^* (\overline{B}) = 7 \), and \( R^* (A) = 7 \).

Note that \( C \) is a proposition Ida believes both before and after revision by \( D \), \( R (C) > 0 \) and \( R^* (C) > 0 \), although \( D \) is positively relevant to, so not independent of, \( C \) in the sense that \( R (C | D) = 7 > 6 = R (C | \overline{D}) \). In other words, Ida receives new information \( D \) whose negation is positively relevant to, and so not independent of, her belief that \( C \) without making her give up her belief that \( C \). On the other hand, if Ida considers \( D \) independent of a proposition \( Y \) before revision by \( D \), then she also does so after revision by \( D \).

This holds true in general. Suppose two propositions \( X \) and \( E \) are independent according to a ranking function \( R \) (Spohn 1999), i.e.

\[
R (X \cap E) + R (\overline{X} \cap \overline{E}) = R (\overline{X} \cap E) + R (X \cap \overline{E}).
\]

Then \( X \) and \( E \) are independent according to any ranking function \( R^* \) that results from \( R \) by a Spohn shift on the experiential partition \( \{E, \overline{E}\} \). This feature, which is known as rigidity, vindicates the idea behind Jin & Thielscher (2007)’s proposal that belief revision should preserve doxastic independencies. It does so by fixing their notion of doxastic independence. \( X \) is positively (negatively) relevant to \( E \) if, and only if, the left-hand side above is smaller (greater) than the right-hand side.
Note also that \( A \) is a proposition Ida believes conditional on the disjunctive assumption \( C \cup D \) before she revises her beliefs by \( D \), since \( R(\overline{A} \mid C \cup D) = 1 > 0 = R(A \mid C \cup D) \). However, after revision by \( D \) she believes the negation of \( A \), \( \overline{A} \), conditional on \( C \cup D \), since \( R^*(A \mid C \cup D) = 7 > 0 = R^*(\overline{A} \mid C \cup D) \). In other words, Ida does not hold on to her conditional belief in \( A \) given \( C \cup D \), even though \( C \cup D \) is a logical consequence of the new information \( D \) and also believed by her. She gives up this conditional belief and replaces it with the new conditional belief that \( A \) given \( C \cup D \). This illustrates that the agent is merely required to hold on to those conditional grades of disbelief whose conditions are directly affected in their entirety by the experience she undergoes, i.e. the conditions which are cells of the experiential partition. Ida is merely required to hold onto her grades of disbelief conditional on the proposition \( D \), and conditional on its negation \( \overline{D} \). This is why it is important that the experiential partition is maximally specific.

Spohn conditionalization gives Ida a complete new ranking function \( R^* \) that she can use to revise her newly acquired belief set \( B^* = \{ Y \in A : R^*(\overline{Y}) > 0 \} \) a second time when, on Tuesday, she receives the information that it is sunny after all. All she has to do is tell us how strongly she then disbelieves the proposition \( B \) that it will rain on Tuesday. Suppose Ida’s very new rank for \( B \) is \( R^{**}(B) = 13 \).

For \( R^* \):

\[
\begin{array}{c|cc}
 A & B & \overline{B} \\
\hline
 A & 8 & 13 & 7 & 18 \\
\overline{A} & 20 & 33 & 14 & 8 \\
\overline{C} & 6 & 33 & 8 & 18 \\
 C & 33 & \infty & \infty & \infty \\
 \end{array}
\]

For \( R^{**} \):

\[
\begin{array}{c|cc}
 A & B & \overline{B} \\
\hline
 A & 21 & 26 & 0 & 11 \\
\overline{A} & 33 & 26 & 19 & 27 \\
\overline{C} & 13 & \infty & 46 & 11 \\
 C & \infty & \infty & \infty & \infty \\
 \end{array}
\]
4.3. **ITERATED BELIEF REVISION REVISITED**

Then Ida’s very new ranks are \( R^\ast (\overline{A}) = 1 \), \( R^\ast (\overline{C}) = 11 \), and \( R^\ast (\overline{D}) = 11 \).

This means that Ida did not mishear the weather forecast, but was too gullible (or so we assume for purposes of illustration). Therefore she has to give up her belief \( C \) that weather forecasts are always right. In addition she has to regain her belief \( A \) that it will be sunny on Wednesday.

Figure 1 pictures Ida’s doxastic career as a sequence of “onions.” In contrast to the AGM theory, the layers now carry numbers which reflect how far apart they are from each other according to the ideal agent’s doxastic state. Figure 2 pictures this situation differently. It allows for empty layers and has one, possibly empty, layer \( R^{-1}(n) \) for each natural number \( n \). Ida’s old rank for \( D \) is two, \( R(D) = 2 \), and her old rank for \( \overline{D} \) is zero, \( R(\overline{D}) = 0 \). Ida’s new ranking function \( R^\ast \) arises from her old one \( R \) by first improving the \( D \)-worlds by two ranks so that the new rank of \( D \) is zero, \( R^\ast(D) = 0 \). Then the \( \overline{D} \)-worlds are deteriorated by thirteen ranks so that the new rank of \( \overline{D} \) is thirteen, \( R^\ast(\overline{D}) = 13 \). The relative positions of the \( D \)-worlds, and of the \( \overline{D} \)-worlds, expressed in the conditional ranking functions \( R(\cdot | D) = R^\ast(\cdot | D) \) and \( R(\cdot | \overline{D}) = R^\ast(\cdot | \overline{D}) \), respectively, are kept fixed.

We see that ranking theory can handle indefinitely iterated revisions of belief. It can do so in contrast to the AGM theory. It can do so in contrast maybe also to Bayesianism. Here is why. The Bayesian agent is sometimes forced to assign probability zero to a consistent proposition – and no probability to a set of possible worlds (Vitali 1905) – if her probability measure is countably additive. Neither of these problems can be overcome by the introduction of infinitesimal probabilities, unless one also drops countable additivity (but then they can be overcome: see Bernstein & Wattenberg 1969). The update rules for probabilistic degrees of belief are all defined in terms of conditional probabilities. The latter are only defined for conditions with positive probability. So, a Bayesian agent may find herself in the awkward position that she cannot revise her degrees of belief when she undergoes an experience that directly affects a proposition she initially assigned degree of belief zero out of necessity, but that she now assigns a positive degree of belief.

To enable the Bayesian agent to conditionalize on such propositions she is sometimes equipped with a Popper-Rényi (PR) function \( P \), which is defined for pairs of propositions. It is possible that \( 0 < P(A | C) < 1 \) even if \( P(C | W) = 0 \). However, as noted by Harper (1976b: 243), it does not help to identify the agent’s initial degree of belief function with \( P(\cdot | W) \), and her degree of belief function after becoming certain of \( E \) with \( P(\cdot | E) \). The reason is that \( P(\cdot | E) \) is merely a classical probability measure. We would violate the principle of categorical matching and be unable to handle iterated revisions of degrees of belief.
To do better we might represent the agent’s initial doxastic state with a PR function $P(\cdot | \cdot)$, and find an update rule that turns it into a new PR function $P^E(\cdot | \cdot)$. To this end Spohn (2012: 206ff) considers the PR analogue of Jeffrey conditionalization. The latter turns each $P(\cdot | \cdot)$ into a new $P_{E \rightarrow p_i}(\cdot | \cdot)$, but Spohn finds it too restrictive (for the same reason as Morgan ms finds the PR analogue of strict conditionalization too restrictive). To be fair, though, this is due to Spohn’s assumption that the input parameters for the cells of the evidential partition are the same for just about all conditions. If the agent’s doxastic state is represented by a two-place PR function, we must allow the input parameters $p_i$ for the cells $E_i$ to vary with the condition $C$ in the second argument place. That is, we must not assume $P_{E_i \rightarrow p_i}(E_i | C)$ to equal $p_i$ for just about every condition $C$. Instead, we need to extend the principle of categorical matching to the experiential input: the format of the experiential input must align with the format of the prior and posterior doxastic state.

Jeffrey and Field conditionalization satisfy this extension of the principle of categorical matching: the prior and posterior doxastic state are degrees of belief, and the experiential input are new degrees of belief, or the amounts by which these change, respectively. So do Spohn and Shenoy conditionalization: the prior and posterior doxastic state are grades of disbelief, and the experiential input are new grades of disbelief, or the amounts by which these change, respectively. If the agent’s doxastic state is represented by a PR function, then the extension of the principle of categorical matching requires that the experiential input are new conditional degrees of belief, or something that specifies these together with the agent’s prior doxastic state. For this is what the prior and posterior doxastic state are on this account.

Allowing the input parameters to vary with the condition solves the problem of restrictiveness: just as every probability measure can be obtained from every regular probability measure on the same algebra by Jeffrey conditionalization; and just as every ranking function can be obtained from every regular ranking function on the same algebra by the update rule; so every PR function can be obtained from every normal PR function on the same domain by the PR analogue of Jeffrey conditionalization with input parameters that may vary with every non-empty condition. Here a PR function is normal if, and only if, $Pr(\overline{C} | C) < 1$ for every non-empty proposition $C$ in the underlying algebra. Normality is for PR functions what regularity is for classical probability measures – except that it is always possible to define a normal PR function, no matter how rich or fine-grained the underlying algebra of propositions.
Unfortunately this “solution” is not a solution: if the input parameters can vary with every non-empty condition, there is no guarantee that the result is a PR function. Once again we violate the principle of categorical matching and cannot handle iterated revisions of conditional degrees of belief.

The Bayesian is, of course, free to propose constraints on the input parameters that guarantee the update rule results in a PR function. However, there is reason to be skeptical this will do. Boutilier (1995: 180) confirms this assessment: “We conclude that the revision of probabilistic belief states is not as well understood as we might have imagined [nor] as well behaved as we might hope.” Similarly Morgan (ms: 8), who notes this “requires combining [his alternative to the PR analogue of strict conditionalization] with a conditionalized version of Jeffrey” conditionalization – but then adds “space limitations prevent elaboration here.”

What is the problem? First, note that a PR function can be represented by an implausibility ordering with a classical probability measure for each position in the ordering, except for the last position which is equipped with the function that assigns 1 to all propositions (van Fraassen 1976, Spohn 1986). In other words, conditional degrees of belief can be represented as implausibility orderings with degrees of belief. So, the situation for iterated revisions of degrees of belief is parallel to that for iterated revisions of beliefs. Spohn (2012: 209ff) contends that the attempt to provide an update rule for PR functions is doomed to fail in the same way as the attempt to handle iterated revisions of belief in the AGM theory.

Now, Spohn’s argument is one by analogy, and the analogy has to be made with respect to two-dimensional belief revision theory, not the one-dimensional AGM theory. For two-dimensional belief revision is the qualitative counterpart to making the input parameters new positions in the implausibility ordering with new degrees of belief – which, by the above representation, is allowing the input parameters to vary with the condition, but constraining them so they result in a PR function. Still, in section 3.3 we have not found two-dimensional belief revision to be able to adequately handle iterated revisions of belief either. Therefore the burden of proof lies with the Bayesian.

If the problem is the same, so is the solution. Spohn (2012: 210f, relying on his 2006b) goes on to show that the Bayesian can do better by subscribing to ranking theory. A normal PR function can be represented by an implausibility ordering with a classical probability measure for each position in the ordering. If we strengthen the comparative implausibility ordering to a quantitative regular ranking function, just as we have done at the beginning of this section, we arrive at ranked probabilities. The latter are regular ranking functions with a classical probability measure for each rank that is taken on by some non-empty proposition.
There is also an update rule for ranked probabilities (Boutilier 1995: 169ff proposes it, too). It is the update rule for ranking functions modulo the addition of a classical probability measure for each rank that is taken on. This update rule satisfies the principle of categorical matching and its extension. It can handle iterated revisions of degrees of belief: the Spohn shift of a ranked probability on finite rank-theoretic, and positive probabilistic, input parameters is a ranked probability. Moreover, the update rule is not restrictive: every ranked probability can be obtained from every regular (in both the probabilistic and rank-theoretic sense) ranked probability that is defined on the same domain. The Bayesian can handle iterated revisions of degrees of belief by subscribing to ranking theory. Whether she can do so without taking this leap remains to be seen.
4.3. ITERATED BELIEF REVISION REVISITED

Figure 2 for section 4.3:
Bibliography


[62] Hájek, Alan (ms1), *Most Counterfactuals Are False*.

[63] Hájek, Alan (ms2), *Staying Regular?*


[74] Hartmann, Stephan & Rad, Soroush Rafiee (ms), Learning Conditionals. Unpublished manuscript.***


[154] Raidl, Eric (ms1), Ranked Semantics for Conditionals.


[183] Spohn, Wolfgang (manuscript), Temporal Self-Location, Forgetting, Biased Evidence, and Auto-Epistemology.


