The Handbook of Rationality

Edited by: Markus Knauff, Wolfgang Spohn

Citation:

The Handbook of Rationality
Edited by: Markus Knauff, Wolfgang Spohn
DOI: 10.7551/mitpress/11252.001.0001
ISBN (electronic): 9780262366175
Publisher: The MIT Press
Published: 2021

Funding for the open access edition was provided by the MIT Libraries Open Monograph Fund.

The MIT Press

5.2 Belief Revision

Hans Rott

Summary

This chapter gives an introduction to the problem of rational belief change and its formal modeling in qualitative logical theories. It first presents an outline of the approach based on rationality postulates and of the most important constructive approaches within the classical AGM model of belief revision. Then it gives the basic ideas of extensions of the classical model to iterated belief revision, to multiple revision and to two-dimensional belief change, to belief merging and to belief updates, as well as to default reasoning.

1. The Problem of Belief Revision

The term "belief revision" refers to the change of a system of beliefs or opinions in response to new information, particularly to information that is inconsistent with this system. In the 1960s and 1970s, the works of Isaac Levi and William Harper prepared the ground for systematic studies of belief change processes. In the 1980s, a seminal research paradigm of belief revision was established by Carlos Alchourrón, Peter Gärdenfors, and David Makinson (also known as "AGM"), who discovered a common structure in the logic of normative systems and of the logic of counterfactual conditionals (see chapter 6.1 by Starr, this handbook)—areas that turned out to be structurally related and were subsequently merged into the field of belief revision (also known as "theory change"). Belief revision theories are different from descriptive psychological investigations (see chapter 5.4 by Gazzo Castañeda & Knauff, this handbook) and investigations of theory dynamics as conducted in the philosophy of science in that they aim at providing postulates and construction recipes for *rational* processes of belief change.

The problem of belief revision is best illustrated by an example. Suppose you believe that

(a) The magician's swords are real swords. s

(b) Lisa, the magician's assistant, is in the box. b

(c) If Lisa is in the box, she cannot escape the swords. $b \rightarrow \neg e$

(d) If the magician's swords are real swords and Lisa cannot escape them, she will die. $s \wedge \neg e \rightarrow d$

Like any person capable of good reasoning, you conclude from these premises that

(e) Lisa will die. d

At the end of the show, however, Lisa climbs out of the box without a scratch. This is a surprise to you. You receive new information $\neg d$ that contradicts your initial premises (a)–(d). Thinking about the matter, you realize that given your initial premises, logic in a way compelled you to believe d. Now that you have found out, luckily, that d is not true after all, and you want to add $\neg d$ to your stock of beliefs, there are only two options: either your logical reasoning that delivered the conclusion d was faulty, or (at least) one of the premises (a)–(d) was false. Assuming that your logic has been flawless and you want to accept the new piece of information $\neg d$, you have to give up (at least) one of your premises. Logic, however, is completely silent about *which* of the premises to give up. Something else has to step in that provides you with guidance how to proceed. Belief revision theories specify what this "something else" is and how it helps you to resolve such problems in the transformation of belief states.

In general, you will not want to give up *all* of your beliefs in order to accommodate the new information $\neg d$ and avoid becoming inconsistent. This radical move would mean an unnecessary loss of valuable information. So when giving up the belief d, you have to decide which of your reasons for holding d to retain and which to retract. You have to *choose* between retracting either s or b or $b \rightarrow \neg e$ or $s \wedge \neg e \rightarrow d$. You may, for instance, decide that you were too rash in endorsing $b \rightarrow \neg e$. Lisa might twist her body into a corner of the box so as to avoid being impaled by the swords.

2. The Representation of Belief

Any theory of the rational formation and transformation of belief states consists of (1) a static part describing the states of belief that a person may be in and (2) a dynamic part describing how belief states should change in response to external input. The static picture constrains the dynamic picture. Assuming that there are such things as plain beliefs, belief revision theorists need to decide how to represent *beliefs* (what a person believes) and *belief states* (the mental states that believers are in).

This chapter considers qualitative models in which belief states don't carry any numerical information and reasoners accept new information without numerical qualification. There are no numbers representing degrees of "certainty," "security," "plausibility," and so on, and belief is always plain belief. In quantitative theories, on the other hand, reasoners do not, or in any case do not have to, accept new information *simpliciter* but can accept it with a certain degree of certainty or plausibility. The best-known quantitative modeling of belief is the probabilistic one, but there are alternative numerical methods of belief change using *ranking functions* (Spohn, 2012; similar ideas are part of possibility theory as summarized by Dubois & Prade, 2015; cf. chapter 4.7 by Dubois & Prade and chapter 5.3 by Kern-Isberner, Skovgaard-Olsen, & Spohn, both in this handbook). Quantitative approaches are more expressive than qualitative methods. Their additional expressive power is a substantial advantage (it is essential, for instance, for expressing significant independence relations between beliefs), but there is a price to be paid for it: often it is not clear where the numbers come from and what exactly they are supposed to mean. While in probability theory, numerical values may be explicated in terms of betting quotients, the meaning of the numbers in ranking functions is less clear (but see Spohn, 2012, chapter 8).

Belief revision theories have traditionally represented beliefs as sentences or propositions (i.e., roughly, contents of sentences). We will focus on sentential models that typically work with a language **L** closed under applications of the Boolean operators of negation, conjunction, disjunction, and material implication.

There is an important contrast in modern epistemology between foundationalist and coherentist theories of the justification of belief. With a grain of salt, this contrast is mirrored in belief revision theories. *Foundationalists* in belief revision assume that there is a set of basic beliefs that are somehow given and may be thought of as justified in a direct, noninferential way. All other beliefs are justified only insofar as they can be derived from the stock of basic beliefs. In more formal terms, such theories presume that there is a *belief base* H of distinguished sentences that have, epistemologically speaking, an independent standing. Another factor to be decided in a formal modeling of belief states is the logic governing the language **L**. A *logic* or *consequence operation* is a function Cn that assigns to any belief base H the set $Cn(H)$ of all logical consequences of H. It is commonly assumed that Cn is Tarskian and includes classical propositional logic.

We call H a *base for* the belief set K if and only if $K = Cn(H)$. *Belief sets* are sets K of sentences in **L** that are closed under Cn, that is, $K = Cn(K)$. If A is an element of K, then A is *believed* (*held to be true*) by a reasoner whose belief set is K. If the negation $\neg A$ is an element of K, then A is *rejected in* K. Of course, in general, there are also sentences A that are neither believed nor rejected but on which the reasoner suspends judgment. A belief set can be viewed as a theory that is a partial description of the world.

A belief derived from a belief base is justified by those elements of the base that are used in its derivation. Like other foundationalist theories in epistemology, the modeling using belief bases suggests that propositions should not be accepted as beliefs unless they are positively justified. Reasoners are supposed to keep track of the justifications for their beliefs. In contrast, *coherence theories* hold that reasoners need not keep track of the pedigrees of their beliefs. They should instead focus on the logical or inferential structure of the beliefs—what matters for the rationality of a belief is how it coheres with the other beliefs that are accepted in the current belief state. For the coherentist, there is no designated set of basic beliefs: each belief depends in some way or other on there being certain other beliefs that support it.

As models of explicit (active, occurrent) belief, *bases* are psychologically more realistic since they are usually finite entities. On the other hand, *belief set* dynamics offers a competence model that helps us to understand what people ought to do ideally if they were not bounded by limited logical or computational reasoning capabilities. There are also interpretations that make sense of the requirement of logical closure even for less-than-ideal reasoners: a belief set may be taken to represent the set of beliefs a reasoner is *committed* to (Levi, 1997) or the set of beliefs *ascribed* to a reasoner. More thoughts on logical closure can be found in chapter 3.1 by Steinberger (this handbook).

3. Kinds of Belief Change

If belief states are represented by sets of sentences, there can only be three doxastic attitudes: a sentence can be accepted, rejected, or neither. We distinguish two basic types of belief change: a sentence A can be *inserted* into a belief set (turned into a belief) or *deleted* from a belief set (turned into a nonbelief). The former is called a *revision* by A, denoted by $K*A$, and the latter a *contraction* with respect to A, denoted by $K\div A$.

An *expansion* $K+A$ of a belief set K by a new piece of information A is simply formed by set-theoretic addition and subsequent logical closure: $K+A = Cn(K\cup\{A\})$. An expansion thus defined is closed under logical consequence, and it is consistent as long as A is consistent with K. But what if A is inconsistent with K? According to the classical rule of *ex falso quodlibet*, a single inconsistency entails any arbitrary sentence, so there is only one inconsistent set closed under Cn, namely, the set L of *all* sentences in the language **L**. Thus, the expansion operation is useless in the belief-contravening case. If consistency is to be preserved, belief-contravening changes require choices in retractions that cannot be made on the basis of set theory and logic alone.

This specific problem of belief-contravening revision was illustrated by our introductory example. From a logical point of view, there were several ways of constructing the revision when accommodating $\neg d$. There is no purely logical reason for making one choice rather than another among the sentences eligible for retraction. The reasoner needs additional information concerning these sentences and a well-defined method for constructing revisions that takes into account a number of coherence constraints (see section 4.1). Such a method is represented formally by a *revision function* $*$ associated with the reasoner's belief set K that can take any sentence A and return as value the revised belief set $K*A$.[1]

The contraction process faces parallel problems. The belief set K of our initial example contains the premises s, b, $b\to\neg e$, and $s\wedge\neg e\to d$ together with all their logical consequences (among which is d). Suppose that you want to contract K with respect to d. Of course, the target sentence d itself must be deleted from K when forming the contracted belief set $K\div d$. But this is not enough. At least one of the four premises must be given up as well in order to prevent d from being rederived through Cn. Again, there is no purely logical reason for making one choice rather than another. Deleting d from your belief set poses quite the same problems as adding $\neg d$ to it.

We have now formed an idea that the problem of belief revision (by $\neg d$) is closely related to the problem of belief contraction (with respect to d). In parallel with revision functions, one can introduce the concept of a *contraction function* $\div$ associated with the reasoner's belief set K that takes sentences A and returns as values the contracted belief sets $K\div A$. While for belief-contravening revision, the problem of maintaining consistency is most palpable, in belief contraction, there is no problem with consistency: shrinking one's set of beliefs can never introduce an inconsistency. But instead, the problem of logical closure makes itself felt rather acutely. The logical interaction between an updated "data base" and its derived consequences may be seen as the ultimate source of the problem of belief change.

4. Two Strategies for Characterizing Rational Changes of Belief

When tackling the problem of belief revision in more concrete detail, two general strategies have been followed. First, one can write down a list of desiderata that an appropriate belief revision function should fulfill. That is, the standards for revision and contraction functions can be laid down in the form of *rationality postulates*. Second, one can present *explicit constructions* for rational changes of belief. The solution to the problem of belief revision will not be complete unless we know how to define and compute appropriate revision and contraction functions for a given belief state. We will see that constructions of belief changes make essential use of "doxastic preference relations" (or similar structures on which the necessary choices may be based).

4.1 The Strategy Based on Rationality Postulates

In their seminal study, Alchourrón, Gärdenfors, and Makinson (1985) proposed a set of general postulates for the rational revision of belief sets. They are now usually called the *AGM postulates*:

(∗1) $K*A$ is closed under Cn.
(∗2) $A\in K*A$.
(∗3) $K*A\subseteq K+A$.
(∗4) If $\neg A\notin K$, then $K+A\subseteq K*A$.
(∗5) If A is consistent, so is $K*A$.
(∗6) If $Cn(A) = Cn(B)$, then $K*A = K*B$.
(∗7) $K*(A\wedge B)\subseteq (K*A)+B$.
(∗8) If $\neg B\notin K*A$, then $(K*A)+B\subseteq K*(A\wedge B)$.

We can think of this set as coming in four pairs. Postulates (∗2) and (∗6) concern the *input*: the input should be accepted, and it is its content, not its syntactic form, that matters. The other three pairs encode three different ideas of coherence. Postulates (∗1) and (∗5) represent a *static* notion of coherence very much like the idea of a reflective equilibrium. They concern the beliefs held by a person at a certain point in time. Consistency is the major driving force behind belief revision theories, and it is a minimal notion of static coherence.[2] Closure is a more demanding idea of static coherence. Postulates (∗3) and (∗4) formulate a *dynamic* notion of coherence. They refer to a short sequence of belief states, namely, the transition from a state to its successor state—but only when the new piece of information A is consistent with the belief set K: in this case, the revision is formed by set-theoretically adding A to K and then taking the logical closure. Dynamic coherence in this sense instantiates a (restricted) maxim of conservatism or, in different terms, of informational economy, minimal change, minimum mutilation (Quine), or cognitive inertia. Prior beliefs should not be abandoned, and new beliefs should not be adopted, beyond necessity. Minimal change has frequently been advertised as one of the principal ideas motivating belief revision theories, but this claim is controversial (Rott, 2000). Postulates (∗7) and (∗8) encode a *dispositional* notion of coherence referring to potential transitions from a single belief state to various possible successor states. More specifically, (∗7) and (∗8) relate the change by a conjunction to the change by one of the conjoined sentences. Dispositional coherence in belief revision turned out to be closely related to the coherence of choices as studied in rational choice theory. While static coherence is essentially a logical notion, dynamic and dispositional coherence are notions relating to two different concepts of economic behavior.[3] But notice that all postulates except (∗2) refer to the underlying logic *Cn*.

Isaac Levi advanced the thesis that any rational belief revision occasioned by a new piece of information A is decomposable into two successive steps, namely, an elimination of $\neg A$, followed by an expansion by A (which produces no inconsistency, thanks to the antecedent contraction step). This suggestion, which is in accordance with the intuitive discussion of our introductory example, has become known as the *Levi identity*:

(L) $K * A = (K \div \neg A) + A$.

Many authors have endorsed (L) and sought to reduce the problem of belief revision to the problem of belief contraction. But there is also a converse equation defining contraction in terms of revision. This is the so-called *Harper identity*, which is, in a rather precise sense, the converse of the Levi identity:

(H) $K \div A = K \cap (K * \neg A)$.

According to (H), the contraction of a belief set K with respect to A is that part of K that will survive K's revision by $\neg A$. Using equations (L) and (H) as bridge principles, a set of eight *AGM postulates* for the rational *contraction* of belief sets is equivalent to the set of postulates for belief revision. Here are just the most interesting ones:

(÷7) $K \div A \cap K \div B \subseteq K \div (A \wedge B)$.

(÷8) If $A \notin K \div (A \wedge B)$, then $K \div (A \wedge B) \subseteq K \div A$.

The conjunction postulates (÷7) and (÷8) compare contractions of K with respect to a conjunction $A \wedge B$ with contractions with respect to the conjoined sentences A and B. (÷7) states that any sentence that is contained both in the contraction of K with respect to A and in the contraction of K with respect to B is contained in the contraction of K with respect to $A \wedge B$. As a kind of converse, if the contraction with respect to $A \wedge B$ eliminates A, then any sentence contained in the contraction with respect to $A \wedge B$ is also contained in the contraction with respect to A. This is natural if eliminating $A \wedge B$ is taken to mean eliminating A or eliminating B (or eliminating both, if A and B are equally entrenched).

AGM called the first six postulates for revisions and contractions the *basic* postulates and the last two ones *supplementary* postulates. The latter pair of postulates entails that belief change operations can be construed as being guided by some well-behaved doxastic preference relation (cf. section 4.2). In many contexts, however, postulates (∗7) and (∗8) have turned out to be too strong. Interesting weakenings, for example, are the following:

(∗7c) If $B \in K * A$, then $K * (A \wedge B) \subseteq K * A$.

(∗8c) If $B \in K * A$, then $K * A \subseteq K * (A \wedge B)$.

Taken together, these conditions say that if B is contained in $K * A$, then $K * A$ is equal to $K * (A \wedge B)$. The "c" in "(∗7c)" and "(∗8c)" means "cumulative," due to the fact that there are corresponding postulates for nonmonotonic logics that are characteristic for *cumulative reasoning* (cf. section 6.4).

4.2 The Constructive Strategy

We have already seen that logical postulates are not sufficient to determine a particular change of K with respect to some proposition A. The construction of belief changes has to draw on extra-logical factors. In some constructions, the syntactic structure of the base for K is itself taken to encode the relevant information (Hansson, 1991; Lewis, 1981; Nebel, 1992). Since the 1990s, most researchers have assumed that these extra-logical factors are *doxastic preference relations* (*plausibility relations*) over sentences or over possible worlds and that such relations are part of the reasoner's belief state. Others have suggested that external, objective criteria (like objective probability, informational content, truthlikeness, or objective similarity between possible worlds) should play the key role in fixing the revision-guiding structures.

AGM explored three methods for the construction of the contractions.[4] The first method defines $K \div A$ as the intersection of the best maximal subsets of K not implying A, where what is best is determined by a doxastic preference relation (*partial meet contraction*). This method was later shown to be intimately related to a semantic modeling using possible worlds (Grove, 1988; Katsuno & Mendelzon, 1991). Here, $K \div A$ is defined as the set of sentences in K that are true in all maximally "plausible" worlds that do not satisfy A, where what is most plausible is determined by a system of spheres of possible worlds centered on the worlds satisfying K. The second method keeps in $K \div A$ precisely those propositions of K that are "safe" in all minimal subsets of K that imply A, where again what is safe is determined by a doxastic preference relation (*safe contraction*). The third method keeps in $K \div A$, roughly speaking, those propositions that are doxastically "well entrenched," where entrenchment relations are yet another kind of doxastic preferences (*entrenchment-based contraction*). Importantly, the doxastic preference relations used in each of these constructions are independent of the particular sentence to be contracted.

It is natural to ask how the postulates strategy and the constructions strategy relate to one another. Assuming that doxastic preferences meet a number of formal requirements (like transitivity and completeness), AGM and their followers were able to prove various *representation theorems* to the effect that a contraction operation for a belief set K satisfies the rationality postulates for contraction if and only if it can be (re)constructed by any of the standard construction recipes. Revisions are taken care of with the help of the Levi identity. As a corollary, the three standard methods are thereby proven equivalent.[5]

While the idea of informational economy has played a rather modest role in AGM-style theorizing of belief revision, ideas from economics have always formed an essential part of belief revision theories through the systematic study of the structure and application of preference relations to cognitive choice problems. Varying (usually weakening) the formal requirements for the relevant preference relations results in variations (usually weakenings) of the postulates encoding dispositional coherence. Some natural relaxations of the requirements for preference relations (including the renunciation of completeness) lead to the postulates (∗7c) and (∗8c) mentioned above. Rationality constraints for choices (as studied by leading economists like P. Samuelson, K. Arrow, and A. Sen; see chapter 8.1 by Grüne-Yanoff, this handbook) can be applied in quite a straightforward way to choice problems involved in belief change processes—choices concerning least plausible beliefs or concerning most plausible possible worlds—and were thus shown to "translate" into rationality postulates for belief revision (Rott, 2001). Seen from this perspective, the problem of belief revision is just an instantiation of the general problem of rational choice, with a few additional constraints placed by the background logic *Cn*. While the formation and transformation of belief sets would usually be classed as a problem of theoretical reason, the making of choices belongs to the realm of practical reason (cf. chapter 2.2 by Wedgwood, this handbook). However, the interpretation of this connection is far from clear. Should we take the use of choice functions as indicating that rational believers are free to decide to believe or disbelieve a proposition according to their preferences? Are they free to adopt the doxastic preferences they wish?[6]

5. Iterated Changes of Belief

There is a broad consensus about how to use selection structures in one-shot belief revision: take the AGM solution. However, iterated belief change or, equivalently, the change of selection structures exhibits substantial ambiguities (Darwiche & Pearl, 1997). Even within the narrow confines of simple qualitative theories (no numbers, only propositional inputs), there are quite a few models that have been proposed. They typically adopt the AGM postulates for one-shot belief revision and add one rationality postulate taking care of iterations. We look at five important proposals.

The first model uses one and the same selection structure for all possible belief sets. This structure does not represent the reasoner's current belief state but is *external* to it (Areces & Becher, 2001). In contrast, the other four models identify the reasoner's belief state with her selection structure. The second model is maximally *conservative*: the selection structure is changed so that the input just gets accepted, but the changes made are absolutely minimal, and as a consequence, the newly acquired information is very easily lost in further belief changes (Boutilier, 1996). The third model, *restrained* revision, is close to the conservative one but gives the input somewhat more impact (Booth & Meyer, 2006). The fourth model is *moderate* in that it recommends giving high priority to the incoming information, but it does not wipe out all plausibility distinctions between worlds or situations that do not comply with it (Nayak, 1994). The fifth model is *radical* in that the input is given uncompromising priority over the previous beliefs: any world or situation that does not satisfy the input is considered maximally implausible after the revision (Segerberg, 1998). Each of these models is well motivated by a possible-worlds semantics and can be axiomatically characterized by a single axiom taking care of the iteration case: it equates $(K*A)*B$ with $K*(A \wedge B)$, if B is consistent with $K*A$ (external and conservative revision); if B is consistent with $K*A$ and A is consistent with $K*B$ (restrained revision); if B is consistent with A (moderate revision); or in any case (radical revision). And it equates $(K*A)*B$ with $K*B$ otherwise (exception: with $L*B$ for external revision[7]). The belief set resulting from any finite number of revision steps can be generated inductively by repeated applications of the respective axiom. We can see that according to each model, two successive revision steps can be reduced to a one-shot revision, mostly by means of some characteristic case distinction (but this kind of reducibility has recently been challenged by Booth & Chandler, 2017). Each of these competing models for iterated belief change is adequate for some cases, but it is easy to come up with examples showing that they have drawbacks in other cases. Unfortunately, no *methodology* has been put forward yet that would tell us when it is rational to use which model.

6. Further Developments

The AGM model is a classic by now that has in many ways been extended and revised.

6.1 Extensions of Belief Revision

Multiple revisions are occasioned by sets of sentences that have to be accepted or withdrawn simultaneously (Delgrande & Jin, 2012; Fuhrmann & Hansson, 1994). In *nonprioritized belief revision*, a reasoner does not invariably incorporate the new information but first decides in the light of her current belief state whether or not she wants to revise her beliefs at all (Hansson, 1997). *Two-dimensional belief revision* is a method that changes beliefs in response to imperatives of the form "Accept A with a degree of plausibility that at least equals that of B" (Cantwell, 1997; Fermé & Rott, 2004; Rott, 2012). Put differently, this belief change operation takes two arguments, an *input sentence* A and a *reference sentence* B, or the input is of the form "$B \leq A$." This method of changing belief states can be used for belief revision (with respect to the input sentence) as well as for belief contraction (with respect to the reference sentence). Two-dimensional belief revision is much more flexible than the methods of iterated belief revision mentioned in section 5. For instance, beliefs can be *lowered*, and nonbeliefs can be *raised*, in doxastic status while still remaining beliefs and nonbeliefs, respectively.

6.2 Belief Merging

The field of *belief merging* or *belief fusion* is concerned with the rational aggregation of two or more belief states (Baral, Kraus, Minker, & Subrahmanian, 1992; Konieczny & Pino Pérez, 2002; Nayak, 1994; Pigozzi, 2015). In many variants of merging, negotiation procedures or aggregation processes similar to those known from social choice theory find application. While belief merging should be thought of as primarily combining belief states of several agents, Konieczny and Pino Pérez (2002) also make room for propositional "inputs" in the original AGM sense (they call them "integrity constraints"). They show that the merging postulates can be modified or supplemented in a wide variety of interesting ways, provide a constructive approach, and give appropriate representation theorems for basic as well as for supplemented models.

6.3 Updates

Katsuno and Mendelzon (1992) discovered an important distinction between belief revisions and belief updates. *Revisions* are prompted by new information about an unchanging world, while *updates* are prompted by information about changes in the world. The difference is most conspicuous in the case where the new piece of information A is consistent with the belief set K. Revisions by A make a person eliminate from the set of worlds compatible with K those worlds that do not satisfy A. If you learn that Lisa is the owner of the theater, with the understanding that she has owned it for some time, then you simply discard all possible worlds in which she is not the owner (this is in accordance with dynamic coherence). In contrast, if you learn that Lisa is the owner of the theater with the understanding that she has just bought it, then you don't drop any of your possible worlds but modify each of them individually by "letting Lisa become" the owner of the theater. An update by A makes a person consider, for each world compatible with K, how this world would develop if A were brought about in it. A formal mark of such updates is that they violate postulate (∗4) and instead satisfy the following monotonicity condition:

(∗M) If $K_1 \subseteq K_2$, then $K_1 * A \subseteq K_2 * A$. (*∗-Monotonicity*)

6.4 Default Inferences as Expectation Revision

Given the information that Tweety is a bird, you will typically conclude that Tweety can fly. But given the information that Tweety is a penguin, which of course implies that Tweety is a bird, you will *not* infer that Tweety can fly. In commonsense reasoning, an extension of the set of premises not only *gains* information but often *loses* some conclusions that were drawn on the basis of the smaller premise set. Everyday inferences thus do not conform to the rules of classical logic but exhibit patterns of *defeasible* or *nonmonotonic* reasoning. Most conspicuously, they violate the following rule of ordinary (Tarskian) logics:

(M) If $H \subseteq H'$, then $Inf(H) \subseteq Inf(H')$. (*Inf-Monotonicity*)

Gärdenfors and Makinson (1994) suggested that the set of defeasible consequences $Inf(H)$ of a finite set of premises $H = \{A_1, \ldots, A_n\}$ can be regarded as the theory that results from the revision of some set of *background assumptions*, *defaults*, or *expectations E* by the conjunction of the premise set *H*. They proposed the following equation:

(I) $Inf(H) = E * (A_1 \wedge \ldots \wedge A_n)$.

For the inference operation *Inf* thus defined, adding information can defeat previously drawn conclusions, just as when we enlarge the singleton premise set containing only "Tweety is a bird" (A_1) with the premise "Tweety is a penguin" (A_2). Given our expectations, the input A_1 makes us conclude that Tweety can fly, whereas conjoining the input A_2 will make us draw the opposite conclusion (note that, given our background knowledge, $A_1 \wedge A_2$ is equivalent to A_2 alone).

Defeasible or nonmonotonic inference operations are quite irregular as compared to Tarskian logics. The failure of (M) is indeed incisive. However, commonsense reasoning nevertheless appears to leave intact quite a number of important classical inference patterns (Brewka, Marek, & Truszczyński, 2011; Kraus, Lehmann, & Magidor, 1990; Makinson, 1994, 2005). Gärdenfors and Makinson (1994) proved correspondences of the basic and full sets of rationality postulates for belief revision and appropriate sets of rationality postulates for default reasoning, and numerous other correspondences are detailed in Rott (2001). Results such as these suggest that belief change and defeasible reasoning are indeed just "two sides of the same coin" (Gärdenfors, 1991).

7. Conclusion

Research on rational belief change originally started out in philosophy and is now most actively pursued in computer science, but it is an essentially interdisciplinary undertaking. One of the tasks ahead is to develop a methodology for when to apply which of the many models on offer. Another task is to close the gap between idealized models of perfectly rational reasoners and the actual facts about the limited logical and computational capacities of human beings. But whether we view it from a normative or an empirical perspective, the formation and transformation of beliefs is a problem that raises its head just about everywhere.

Notes

1. Many authors present revision functions as two-place functions taking a pair of a belief set and a sentence and returning a new belief set. This way of putting things appears misleading to me, since one and the same belief set may well be coupled with different belief revision strategies, so $*$ as acting on pairs $\langle K, A \rangle$ cannot be functional in the *K* argument. For more details, see Rott (1999).

2. The consistency requirement can be seen as marking off the enterprise of belief revision from that of paraconsistent logic, which shares the aim of explicating rational deliberation in the face of contradictory information. A logic or inference operation is *paraconsistent* if it does not satisfy the classical rule of *ex falso quodlibet* (i.e., if not everything is derivable from a classically inconsistent set of premises). Wassermann (2011), Girard and Tanaka (2016), and Testa, Coniglio, and Ribeiro (2017) argue that one should *combine* belief revision and paraconsistent logics.

3. See Rott (2003). The thesis that recipes for belief revision ought to be derived from decision-theoretic principles has been championed by Isaac Levi at least since Levi (1991).

4. For more detailed overviews, see Gärdenfors and Rott (1995), Hansson (1999), and Fermé and Hansson (2018).

5. A great variety of representation theorems can be found in Gärdenfors (1988), Hansson (1999), Bochman (2001), Rott (2001), and Fermé and Hansson (2018).

6. This raises the question of doxastic voluntarism, which is linked with belief revision theory in Rott (2017).

7. "External" methods based on global (i.e., belief-state independent) similarities or distances between possible worlds behave differently (cf. Lehmann, Magidor, & Schlechta, 2001).

References

Alchourrón, C., Gärdenfors, P., & Makinson, D. (1985). On the logic of theory change: Partial meet contraction and revision functions. *Journal of Symbolic Logic*, *50*, 510–530.

Areces, C., & Becher, V. (2001). Iterable AGM functions. In M.-A. Williams & H. Rott (Eds.), *Frontiers in belief revision* (pp. 261–277). Dordrecht, Netherlands: Kluwer.

Baral, C., Kraus, S., Minker, J., & Subrahmanian, V. S. (1992). Combining multiple knowledge bases consisting of first-order theories. *Computational Intelligence*, *8*, 45–71.

Bochman, A. (2001). *A logical theory of nonmonotonic inference and belief change*. Berlin, Germany: Springer.

Booth, R., & Chandler, J. (2017). The irreducibility of iterated to single revision. *Journal of Philosophical Logic*, *46*, 405–418.

Booth, R., & Meyer, T. (2006). Admissible and restrained revision. *Journal of Artificial Intelligence Research*, *26*, 127–151.

Boutilier, C. (1996). Iterated revision and minimal change of conditional beliefs. *Journal of Philosophical Logic*, *25*, 263–305.

Brewka, G., Marek, V., & Truszczyński, M. (Eds.). (2011). *Nonmonotonic reasoning: Essays celebrating its 30th anniversary*. London, England: College Publications.

Cantwell, J. (1997). On the logic of small changes in hypertheories. *Theoria*, *63*, 54–89.

Darwiche, A., & Pearl, J. (1997). On the logic of iterated belief revision. *Artificial Intelligence*, *89*, 1–29.

Delgrande, J., & Jin, Y. (2012). Parallel belief revision: Revising by sets of formulas. *Artificial Intelligence*, *176*, 2223–2245.

Dubois, D., & Prade, H. (2015). Possibility theory and its applications: Where do we stand? In J. Kacprzyk & W. Pedrycz (Eds.), *Springer handbook of computational intelligence* (pp. 31–60). Berlin, Germany: Springer.

Fermé, E., & Hansson, S. O. (2018). *Belief change: Introduction and overview*. Berlin, Germany: Springer.

Fermé, E., & Rott, H. (2004). Revision by comparison. *Artificial Intelligence*, *157*, 5–47.

Fuhrmann, A., & Hansson, S. O. (1994). A survey of multiple contractions. *Journal of Logic, Language and Information*, *3*, 39–76.

Gärdenfors, P. (1988). *Knowledge in flux: Modeling the dynamics of epistemic states*. Cambridge, MA: MIT Press.

Gärdenfors, P. (1991). Belief revision and nonmonotonic logic: Two sides of the same coin? In J. van Eijck (Ed.), *Logics in AI: JELIA 1990* (Lecture Notes in Computer Science (Lecture Notes in Artificial Intelligence), vol. 478). Berlin, Germany: Springer.

Gärdenfors, P., & Makinson, D. (1994). Nonmonotonic inference based on expectations. *Artificial Intelligence*, *65*, 197–245.

Gärdenfors, P., & Rott, H. (1995). Belief revision. In D. M. Gabbay, C. J. Hogger, & J. A. Robinson (Eds.), *Handbook of logic in artificial intelligence and logic programming: Vol. IV. Epistemic and temporal reasoning* (pp. 35–132). Oxford, England: Oxford University Press.

Girard, P., & Tanaka, K. (2016). Paraconsistent dynamics. *Synthese*, *193*, 1–14.

Grove, A. (1988). Two modellings for theory change. *Journal of Philosophical Logic*, *17*, 157–70.

Hansson, S. O. (1991). *Belief base dynamics* (Unpublished doctoral dissertation). Uppsala University, Uppsala, Sweden.

Hansson, S. O. (1999). *A textbook of belief dynamics: Theory change and database updating*. Dordrecht, Netherlands: Kluwer.

Hansson, S. O. (Ed.). (1997). Non-prioritized belief revision [Special issue]. *Theoria*, *63*.

Katsuno, H., & Mendelzon, A. O. (1991). Propositional knowledge base revision and minimal change. *Artificial Intelligence*, *52*, 263–294.

Katsuno, H., & Mendelzon, A. O. (1992). On the difference between updating a knowledge base and revising it. In P. Gärdenfors (Ed.), *Belief revision* (pp. 183–203). Cambridge, England: Cambridge University Press.

Konieczny, S., & Pino Pérez, R. (2002). Merging information under constraints: A logical framework. *Journal of Logic and Computation*, *12*, 773–808.

Kraus, S., Lehmann, D., & Magidor, M. (1990). Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence*, *44*, 167–207.

Lehmann, D., Magidor, M., & Schlechta, K. (2001). Distance semantics for belief revision. *Journal of Symbolic Logic*, *66*, 295–317.

Levi, I. (1991). *The fixation of belief and its undoing*. Cambridge, England: Cambridge University Press.

Levi, I. (1997). The logic of full belief. In I. Levi, *The covenant of reason: Rationality and the commitments of thought* (pp. 40–69). Cambridge, England: Cambridge University Press.

Lewis, D. (1981). Ordering semantics and premise semantics for counterfactuals. *Journal of Philosophical Logic*, *19*, 217–234.

Makinson, D. (1994). General patterns in nonmonotonic reasoning. In D. M. Gabbay, C. J. Hogger, & J. A. Robinson (Eds.), *Handbook of logic in artificial intelligence and logic programming: Vol. III. Nonmonotonic reasoning and uncertain reasoning* (pp. 35–110). Oxford, England: Oxford University Press.

Makinson, D. (2005). *Bridges from classical to nonmonotonic logic*. London, England: King's College Publications.

Nayak, A. C. (1994). Iterated belief change based on epistemic entrenchment. *Erkenntnis*, *41*, 353–390.

Nebel, B. (1992). Syntax-based approaches to belief revision. In P. Gärdenfors (Ed.), *Belief revision* (pp. 52–88). Cambridge, England: Cambridge University Press.

Pigozzi, G. (2015). Belief merging and judgment aggregation. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. Retrieved from https://plato.stanford.edu/archives/win2016/entries/belief-merging/

Rott, H. (1999). Coherence and conservatism in the dynamics of belief. Part I: Finding the right framework. *Erkenntnis*, *50*(2–3), 387–412.

Rott, H. (2000). Two dogmas of belief revision. *Journal of Philosophy*, *97*, 503–522.

Rott, H. (2001). *Change, choice and inference: A study of belief revision and nonmonotonic reasoning*. Oxford, England: Oxford University Press.

Rott, H. (2003). Economics and economy in the theory of belief revision. In V. F. Hendricks, K. F. Jørgensen, & S. A. Pedersen (Eds.), *Knowledge contributors* (pp. 57–86). Dordrecht, Netherlands: Kluwer.

Rott, H. (2012). Bounded revision: Two-dimensional belief change between conservative and moderate revision. *Journal of Philosophical Logic*, *41*, 173–200.

Rott, H. (2017). Negative doxastic voluntarism and the concept of belief. *Synthese*, *194*, 2695–2720.

Segerberg, K. (1998). Irrevocable belief revision in dynamic doxastic logic. *Notre Dame Journal of Formal Logic*, *39*, 287–306.

Spohn, W. (2012). *The laws of belief: Ranking theory and its philosophical applications*. Oxford, England: Oxford University Press.

Testa, R. R., Coniglio, M. E., & Ribeiro, M. M. (2017). AGM-like paraconsistent belief change. *Logic Journal of the IGPL*, *25*, 632–672.

Wassermann, R. (2011). On AGM for non-classical logics. *Journal of Philosophical Logic*, *40*, 271–294.

The MIT Press would like to thank the anonymous peer reviewers who provided comments on drafts of this book. The generous work of academic experts is essential for establishing the authority and quality of our publications. We acknowledge with gratitude the contributions of these otherwise uncredited readers.

This book was set in Stone Serif and Stone Sans by Westchester Publishing Services.

Library of Congress Cataloging-in-Publication Data

Names: Knauff, Markus, editor. | Spohn, Wolfgang, editor.
Title: The handbook of rationality / edited by Markus Knauff and Wolfgang Spohn.
Description: Cambridge : The MIT Press, 2021. | Includes bibliographical references and index.
Identifiers: LCCN 2020048455 | ISBN 9780262045070 (hardcover)
Subjects: LCSH: Reasoning (Psychology) | Reason. | Cognitive psychology. | Logic. | Philosophy of mind.
Classification: LCC BF442 .H36 2021 | DDC 153.4/3—dc23
LC record available at https://lccn.loc.gov/2020048455