Chapter 1
Motivation

This is a brief introduction to the theory of belief change. It provides an example of a belief change problem, and lists some of the major issues that are investigated in this research area.

1.1 An Example of the Belief Change Problem

We consider the following set of sentences in natural language [94]: "Juan was born in Puerto Carreño" (α), "José was born in Puerto Ayacucho" (β), and "Two people are compatriots if they were born in the same country" (γ). We assume that this set represents all the currently available information about Juan and José. Suppose that we receive the following piece of new information: "Juan and José are compatriots" (δ). If we add the new information to our corpus of beliefs, then we obtain a new set of beliefs that contains the sentences α, β, γ and δ. We can define an operation of addition as one that takes a sentence and a set of previous beliefs and returns the minimal set that includes both the previous beliefs and the new sentence. This operation exemplifies the simplest way of changing a set of sentences. There are other types of change that are not that simple.

For example, suppose that upon consulting an atlas we discover to our surprise that Puerto Carreño is in Colombia (ϵ) and Puerto Ayacucho is in Venezuela (ϕ). If we add ϵ and ϕ to the set $\{\alpha,\beta,\gamma,\delta\}$, the result will be a set with contradictory information: Juan and José are compatriots but Puerto Carreño and Puerto Ayacucho do not belong to the same country. The addition does not satisfactorily reflect the notion of a *consistent revision*. If we wish to retain consistency, then some subset of the original set must be discarded or perhaps a part of the new information has to be rejected. In our example, there are several possible alternatives. The information about Juan's or José's birthplace could be wrong, and so could the atlas. Finally the claim that Juan and José are compatriots could be wrong. Any of these three options, either individually or combined, will allow us to solve the problem of the incompatibility among the original and the new information or beliefs. Consequently, we

E. Fermé and S. O. Hansson, *Belief Change*, SpringerBriefs in Intelligent Systems, https://doi.org/10.1007/978-3-319-60535-7_1

can specify an operation that takes a set and a sentence and returns a new consistent set. The new set includes parts (or all) of the beliefs in the original set and it also includes the new sentence (if we are willing to accept it). The outcome of a revision can be expressed as a consistent subset of the outcome of the addition. This operation is based on two notions: *consistency* and a *selection* among the possible ways to perform the change.

There are other ways to change a set of beliefs. Suppose that we discover that γ is incorrect, and therefore wish to discard it from our set. The result should be a new set where γ is absent. We may for instance want it to be undetermined whether Juan and José are compatriots. Note that this is different from accepting as a fact that Juan and José are not compatriots. We can ask if the process of discarding information should behave as the inverse of the process of adding information: If after discarding some information we proceed to add it again, will we obtain the original set or not? Like revision, the operation of discarding requires the selection of one out of several possible results.

1.2 Some Questions About the Belief Change Problem[1]

Any formalization of belief change requires the selection of a language in which the beliefs are represented. In our previous example the information about Juan and José is represented by a set of sentences in natural language. The use of a linguistic representation of beliefs implies the acceptance of important idealizations. Whatever language is chosen, the question emerges how to use the language to represent the epistemic state: should it be represented by a single sentence or by a set (perhaps an infinite set) of sentences? In the latter case, should the set be closed under some notion of logical consequence or should it only be a simple enumeration of sentences? The second option implies the need to obtain in some way the consequences of these sentences and to differentiate between implicit and explicit information.

Can the belief state be changed spontaneously or does change require an external stimulus? In other words, is the belief state internally stable? If the belief state is changed only in response to external stimuli, should the belief state and the information that provokes the change be represented by the same or different types of formal structures? Should both be sentences or both be sets of sentences? How should the sentences be interpreted? If an epistemic interpretation of the sentences is chosen, what are the possible statuses of the sentences? Acceptance, rejection, indeterminateness, or perhaps degrees of acceptability? What types of information can be represented in the belief state?

Generally speaking, it seems to be fundamental to define operations that answer to the notion of minimal change, or maximal preservation of the belief state. That is to say, it is required in some way to "calculate the value" of the information to

[1] Borrowed from [11].

be discarded. Does a preference order exist that represents the credibility or informational value of expressions in the language? Is this order included in the belief state or is it intrinsic to the change operation? Should minimal change be defined quantitatively or qualitatively?

In what ways can a belief state be modified? Are they independent or interrelated? What is the relationship between the original and the updated belief state? How should an operation to revise the original belief state be constructed? What are the parameters of this operation? The original belief state and the new information are obvious such parameters, but are there any other parameters? Should the change operation take into account the history of the produced changes, or is each new change performed independently of those performed earlier?

These kinds of questions have encouraged several authors to propose different belief change models and to assume some of the above options and discard others. By far the most influential of these models was proposed by Carlos Alchourrón (1931–1996), Peter Gärdenfors, and David Makinson in their paper "On the Logic of Theory Change: Partial Meet Contraction and Revision Functions". Many research papers have been called "seminal", but few deserve that designation as much as this article in the *Journal of Symbolic Logic* in 1985. It was the starting point of a large and rapidly growing literature that employs formal models in the investigation of changes in belief states and databases.

This book is an introduction on and an overview of the research that has been inspired by the AGM article.