Chapter 2
Inside the Black Box

As we saw in Chapter 1, there are two major formal approaches to rationality in belief change. One is the *constructive approach*. We can design various mechanisms for operations of belief change, and it is then a topic for discussion whether these constructions are built on plausible principles. In this approach we may ask for instance whether selection functions and relations of epistemic entrenchment correspond to credible ways for a cognitive agent to change its beliefs. The other alternative is the *axiomatic approach* in which we consider various properties that belief change operations can have, expressed with the AGM postulates and others of the same sort. The tenability of each of these postulates can be scrutinized with the help of examples of reasonable changes in belief.

Both these approaches have weaknesses. A major problem with the constructive approach is that if a mechanism is implausible or difficult to explain, it may nevertheless yield the right results. An unconvincing construction can be defended as a "black box", a gear that we should be happy with because it does what it is supposed to do, even if we do not fathom how it does so. On the other hand, postulates only provide a partial description of how beliefs are changed. That a change operation satisfies a set of plausible postulates does not prevent it from also satisfying other, quite implausible ones.

The best solution to this problem is to combine the constructive and the axiomatic approach, in other words to specify mechanisms that we consider to be plausible and characterize them completely in terms of axioms. That was the route taken by the AGM authors, and in doing so they set a standard for subsequent researchers in the field. The purpose of this and the following chapter is to uncover problems in the AGM framework that can justify the development of alternative frameworks for belief change. In the present chapter we will follow the constructive approach, and investigate the use of selection mechanisms in both partial meet contraction and sphere systems. In Section 2.1 the notion of epistemic choice is discussed, and choice functions are introduced. In Section 2.2 it is clarified how in both partial meet contraction and sphere models, the application of a selection function is followed by

S.O. Hansson, *Descriptor Revision*, Trends in Logic 46,
DOI 10.1007/978-3-319-53061-1_2

the intersection of the selected sets. The plausibility of this sequence of operations is scrutinized in Section 2.3, and the formal limits to its applicability are pointed out in Section 2.4. Finally, in Section 2.5 we discuss the crucial question whether the AGM selection mechanisms are applied to the right objects.

2.1 Epistemic Choice

Choice has a central role in the theory of belief change. Operations of change take the form of replacing a belief set by another that satisfies a given success condition (such as $p \in K * p$ for revision and $p \notin (K \div p) \setminus \mathrm{Cn}(\varnothing)$ for contraction). In typical cases, there are many belief sets satisfying this condition, and exactly one of them is the outcome. The process of identifying one of the alternatives as the outcome is usually conceived as a choice. However, it must be recognized that the notion of choice is far from unproblematic in an epistemic context. We do not normally choose what to believe in the same way that we choose between dishes in a restaurant. Most belief changes seem to be uncontrollable effects of external influences rather than the results of voluntary choices made by the subject ([119, pp. 143–145]. See also: [8, 22, 141, 142, 192, 197, 199, 223, 232, 251, 252].)

> Svetlana has two sisters, Olga and Aleksandra. Olga is severely ill. One day Svetlana came to Pavel and said, sobbing: "Now I have only one sister."
>
> "How terrible", he said. "I knew that Olga was approaching the inevitable but I had hoped that she would live to see her grandchild."

Logically speaking, what Svetlana said only gave Pavel reason to believe that either Olga or Aleksandra had died. Nevertheless, his belief that Olga had died came to him immediately, unpreceded by any choice or other premeditation. But presumably, if he had carefully compared the alternatives and chosen which of them to believe in, the outcome would have been the same. We can take such reconstructibility in terms of premeditated choice as a criterion of rational belief change. Spontaneous behaviour can be rational, but only if it coincides with what one could have done if guided by rational reflection.

This "as if" approach to rationality (that is also common in decision theory [140, pp. 381–382]) has important implications for the formal representation of belief change. If a process of belief change takes place *as if* it was an actual choice, then that provides us with a reasonable justification for representing it as a choice.

Choices have been extensively studied in economics and in particular in social choice theory [237]. The standard formal representation of choice used in these disciplines, namely choice functions, has been taken over by belief change theory.[1]

[1] Arrow introduced choice functions in economics. He said: "We do not want to prescribe that $C(S)$ contains only a single element; for example, S may contain two elements between which the chooser is indifferent." [7, p. 4]. At that time choice functions were already used in logic, but the standard definition in logic was different. A choice function for a set $\mathfrak{X}$ of non-empty sets was defined as a

A choice function is defined over a set $\mathcal{A}$ of alternatives. It can be used to make a selection among any subset of $\mathcal{A}$:

Definition 2.1 *C is a* choice function *for a set $\mathcal{A}$ if and only if for each subset $\mathcal{B}$ of $\mathcal{A}$:*

(1) $C(\mathcal{B}) \subseteq \mathcal{B}$, *and*

(2) $C(\mathcal{B}) \neq \varnothing$ *if* $\mathcal{B} \neq \varnothing$.

A choice function C for $\mathcal{A}$ is based on *a relation $\multimap$ if and only if for all $\mathcal{B} \subseteq \mathcal{A}$:* $X \in C(\mathcal{B})$ *if and only if* $X \in \mathcal{B}$ *and* $X \multimap Y$ *for all* $Y \in \mathcal{B}$.

According to this definition, $C(\mathcal{B})$ can have more than one element. In everyday talk about choice, choices sometimes have this property, sometimes not:

Example 1:
"I am going to throw away these old LP records unless you want some of them. Choose those you want, and then I will throw away the rest."

Example 2:
"Since you have done so much for me I want to give you an LP record from my collection. You are free to choose whichever you like."

Choice functions, as defined above, represent the type of choice instantiated in the first of these examples.

2.2 The Select-and-Intersect Method

In social choice theory, when a choice function delivers an outcome with more than one element, this means that all those elements are (considered to be) equally choiceworthy. It is then left to the decision-maker to further narrow down the choice to one single object. Which element of $C(\mathcal{B})$ she ends up with is presumed to be arbitrary from the viewpoint of rationality. Hence, if Alex, Bailey, and Casey are three willing candidates for marriage, then $C(\{\text{Alex, Bailey, Casey}\}) = \{\text{Alex, Bailey}\}$ does not indicate a wish for bigamy but rather vacillation between Alex and Bailey.

Therefore, strictly speaking, choice functions in social choice theory only cover the first of two stages in a choice process. The second stage that slims down the outcome to a single element is often described as a matter of picking rather than choosing [245]. We can call this the *select-and-pick* method.

(Footnote 1 continued)
function C such that $C(X) \in X$ for all $X \in \mathfrak{X}$. [147] – On the use of choice functions in logic, see also [138].

In belief change as well, choice functions with multiple outputs leave us with a need for a further process that takes us from several objects to a single one.[2] In belief change, however, the second stage is different.[3] It consists in forming the intersection of the sets chosen in the first stage. This intersection is taken to be the outcome of the operation. This has been called the *select-and-intersect method* [135]. It comes in two major versions, both of which were introduced in Chapter 1. In partial meet contraction, the first stage is a selection among remainders, and in sphere-based revision it is a selection among possible worlds. The second stage, intersection, is the same in both cases.

At first glance, the select-and-intersect method may seem to be an almost impeccable way to deal with ties. When we hesitate between two or more potential outcomes, then it would seem natural to use their intersection, i.e. what they all have in common, as the output. But closer inspection will reveal that the select-and-intersect method can be questioned on at least three accounts. First, we can dispute the *preservation of optimality under intersection*. In the first step of the select-and-intersect process, options are chosen that are in some sense optimal. In partial meet contraction the first step passes on the best or most choiceworthy remainders that satisfy the success condition to the second stage for intersection. But is that optimality retained after intersection? Or would perhaps the intersection of some other set of remainders be more choiceworthy, while still satisfying the success condition? If the latter is true, then the achievement of the first stage was lost in the second.

Secondly, the *preservation of success under intersection* cannot always be taken for granted. In partial meet contraction, the success condition is the elimination of some input sentence p. All the sets chosen in the first stage satisfy that condition (since elements of $K \perp p$ do not contain p). It follows that their intersection, the final outcome of the operation, does not contain p either. In other words, this success condition is preserved under intersection. But does that apply to all success conditions that we may wish to apply? If not, then that is a constraint on the applicability of the select-and-intersect method.

Thirdly and perhaps most importantly, the *adequacy of the options selected for intersection* is contestable. In the AGM approach, the primary selection is made among remainders or (in the sphere model) possible worlds. Are these plausible outcomes? As noted above, the use of intersection can be justified as a means to adjudicate between equally plausible outcomes. It would seem more difficult to justify the select-and-intersect method if the objects chosen for intersection are not plausible outcomes of the operation.

In the next three sections we are going to look more closely at each of these problems for the select-and-intersect method.

---

[2] A few studies have been devoted to indeterministic belief change operations. These are operations that deliver, for each input, a set that may contain more than one possible outcome [66, 169].

[3] This difference would seem to have implications for the view that the use of choice functions in both areas reveals an underlying unity between practical and theoretical reasoning. On that view, see [205, 215, 217].

2.3 Is the Intersection as Good as Its Origins?

Although the use of choice functions in belief change is largely modelled after social choice theory, the use of intersection among options is unknown in social choice. The reason for this is obvious: in a social choice context optimality is not preserved under intersection [118].

> GAME SHOW HOST: Congratulations! You have won the first prize. This means that you now have a choice between two options. One is a Porsche 991 and 50 litres of petrol. The other is a Lamborghini Huracán and 50 litres of petrol. Which of them do you choose?
> CONTESTANT: I am unable to choose between them. The two alternatives are exactly equally good.
> GAME SHOW HOST: Thanks for telling us. We will now follow our standard procedure for such cases of indecision, and give you the intersection between the two sets you could not choose between. One of the sets contains a Porsche 991 and 50 litres of petrol, and the other a Lamborghini Huracán and 50 litres of petrol. Let me congratulate you once more. You are now the happy owner of the intersection of those two sets, namely 50 litres of petrol, of the highest quality.

This absurdity would have no relevance for belief change if it could be shown that contrary to other collections of objects, logically closed sets of sentences do not lose in choiceworthiness by being intersected with other equally choiceworthy objects. However, no such argument seems to be forthcoming. This problem was first pointed out by Tor Sandqvist. He proposed that we consider two collections of beliefs sets, $\mathcal{A}$ and $\mathcal{B}$. Suppose that each belief set in $\mathcal{A}$ is preferable to each belief set in $\mathcal{B}$. From this, he says, it does not follow that the belief set $\bigcap \mathcal{A}$ is preferable to the belief set $\bigcap \mathcal{B}$. The reason for this is that the elements of $\mathcal{A}$ may be "each very valuable but such that their intersection is practically worthless – namely, if whatever makes each of them so valuable fails to be that which they all have in common." [227, p. 292].

This argument is in need of a supporting example, but the construction of such an example is made difficult by the fact that we may have different standards of choiceworthiness for belief sets. Belief change theory is in general neutral between such standards, but if we wish to illustrate how choiceworthiness can be lost in intersection, the standard of choiceworthiness has to be made explicit. The following example has been chosen because its standard of choiceworthiness is particularly susceptible to deterioration through intersection:

> Ibraham chooses between five sets of religious beliefs, namely the full set of Roman Catholic beliefs (C_1), that of Lutheran beliefs (C_2), that of Sunni beliefs (I_1), that of Shia beliefs (I_2), and finally the beliefs of Spinozan pantheism (P). Judging these belief systems according to their ability to give him guidance and peace of mind, he considers each of C_1, C_2, I_1, and I_2 to be equally choiceworthy, and each of them to be more choiceworthy than P. However, $C_1 \cap C_2 \cap I_1 \cap I_2$, the state of hesitation between the four belief systems he ranks highest, is much

worse than P. It gives him no peace of mind, and the guidance it provides on how to conduct his life is tantalizingly incomplete. For instance it tells him that there is only one road to salvation, but leaves him ignorant of which that road is.

This problem also has a reverse form that comes out most clearly in sphere-based possible world models of revision. In these models, the original belief set K is assumed to be the intersection of all the possible worlds that have maximal plausibility. As explained in Section 1.4, it follows that the possible worlds that have K as a subset are all equally plausible. To see why this is problematic, note that my present belief set K neither contains the statement that Proxima Centauri b, the closest known exoplanet, has intelligent life (p) nor the statement that it does not ($\neg p$). Consequently, there are possible worlds containing $K \cup \{\neg p\}$ and also possible worlds that contain $K \cup \{p\}$. It follows from the sphere-based construction that these worlds are all equally plausible. This is counter-intuitive since I hold $\neg p$ to be more plausible than p. On a more basic level, it is difficult to see what it means to apply a concept of plausibility – or any other property that correlates in the intended way with epistemic choiceworthiness – to a single possible world.

2.4 Do All Success Conditions Withstand Intersection?

Up to now we have only discussed two success conditions, namely those of contraction (absence of the input sentence) and revision (presence of the input sentence). Both these success conditions have the following characteristic:

A property on sets is *preserved under intersection* if and only if the following holds for all non-empty collections $\mathfrak{X}$ of sets:

If each element of $\mathfrak{X}$ has the property, then so does $\bigcap \mathfrak{X}$ [126].

It is the preservation under intersection of the respective success conditions that makes the select-and-intersect method operable for contraction and revision. If p is absent from all elements of $\mathfrak{X}$, then it is also absent from $\bigcap \mathfrak{X}$. Similarly, if p is present in all elements of $\mathfrak{X}$, then it is also present in $\bigcap \mathfrak{X}$. The following example shows that we may sometimes wish to perform an operation of change with a success condition that is not preserved under intersection.

According to the public prosecutor's indictment, the accused has committed either murder or voluntary manslaughter. Susan is the judge assigned to the case. According to procedural law, she has three options. She can find the accused guilty of murder, find him guilty of voluntary manslaughter, or acquit him. She is convinced that he has killed the victim, but finds it difficult to adjudicate whether it was murder or not. Although the procedural law admits disjunctive indictments, it does not allow disjunctive verdicts. She therefore has to make up her mind so that she can either conclude that the accused committed murder or that he is guilty of voluntary manslaughter.

The success condition for the shift in her beliefs that the situation requires can best be described as a requirement that she either comes to believe that the accused committed murder (m) or that he committed voluntary manslaughter (v). This is of course different from believing that he committed either murder or voluntary manslaughter ($m \vee v$), which she already does. The belief change called for can be formalized as an operation of *choice revision*, in which the input is a set of sentences rather than a single sentence [60, 64]. The success condition of choice revision by a set A of sentences is that the output should contain at least one element of A. To see that this condition is not preserved under intersection, we can use our example $A = \{m, v\}$ and consider the two potential outcomes $X_1 = \mathrm{Cn}(\{m\})$ and $X_2 = \mathrm{Cn}(\{v\})$. The success condition is satisfied by both X_1 and X_2 but not by their intersection $X_1 \cap X_2$.[4] Choice revision also defies the decomposition principle discussed in the previous chapter, i.e. it does not seem to be reconstructible in terms of expansion and contraction.

2.5 Do We Select Among the Right Objects?

When choice functions are used in social choice theory, they operate on sets containing objects available for choice, such as physical objects or social states of affairs. The standard properties of choice functions have been developed from our intuitions about their application to objects we can choose between. In belief change, we use choice functions to obtain a new belief set. To choose a belief set means to choose among potential belief sets, just as choosing a dessert means to choose among desserts. Therefore, we should expect the choice functions (selection functions) of belief change theory to be applied to potential belief sets.

However, as we saw in Chapter 1, selection functions are standardly applied to sets of remainders and possible worlds. It is not difficult to show that neither of these are plausible belief sets. Beginning with possible worlds, we have already noted that if W is a possible world, then it holds for each sentence q in the language that either $q \in W$ or $\neg q \in W$. The absurdity of belief sets with this property was noted by two of the AGM authors already in 1982 [3, p. 21]. An example of how the sphere model works can serve to illustrate the point: On one occasion I had a belief set K containing the sentence "There is milk in my fridge" (p). When opening my fridge I found this to be wrong and revised my belief set by $\neg p$. The sphere model (as in Fig. 1.3, substituting $\neg p$ for p) depicts this change as a process in which I first selected the most plausible possible worlds in which $\neg p$ is true, and then adopted the intersection of all those worlds as my new belief set. In each of the options selected in the first stage I would be a full-fledged *Besserwisser*, willing to assign a confident "true" or "false" to every statement that can be made in the language. Needless to say, my experience of coming to believe that I had no milk did not involve an intermediate

[4] See Section 4.4 for a formal characterization of the preservation of success conditions under intersection of belief sets.

stage in which I vacillated between different forms of purported omniscience. A reconstruction of the process in such terms seems far-fetched.

Remainders do not have this property, but they have another problematic property:

Observation 2.2 ([3, p. 20]) *Let $p \in K$ and $X \in K \perp p$, and let q be any sentence. Then either $p \vee q \in X$ or $p \vee \neg q \in X$.*

As Alchourrón and Makinson noted, this is a "rather counterintuitive" property, in particular when q is (intuitively speaking) content-wise unconnected with both p and the rest of K [3]. To see that, we can again consider my belief change when I found no milk in the fridge. Let q denote that Socrates had the hiccups on his sixth birthday. According to the partial meet account of belief revision, my adoption of the belief $\neg p$ began with retraction of p from my original belief set. The retraction followed the select-and-intersect pattern. Therefore, in the initial selection phase I chose among a collection of belief sets, in each of which I believed in one of the two statements "either there is milk in the fridge or Socrates had the hiccups on his sixth birthday" ($p \vee q$) and "either there is milk in the fridge or Socrates did not have the hiccups on his sixth birthday" ($p \vee \neg q$). Both of these are strange beliefs for someone to hold who has no idea what happened to Socrates on his sixth birthday. I should be able to give up my belief that I have milk in the fridge without passing through an intermediate stage in which I vacillate between such outlandish belief states.[5]

Furthermore, a plausible belief state should be one that a human mind can harbour. Since we are finite beings, we cannot have belief states that require infinite representations. If a belief state can only be represented by infinite sets, then it is not a belief state that human beings can have or entertain having. (Nor is it representable in a computer.) We should therefore expect all belief sets that are considered in a belief change process to satisfy the following condition:

Definition 2.3 *A logically closed set X of sentences is* finite-based *if and only if there is some finite set X' such that $X = \mathrm{Cn}(X')$.*

A simple way to achieve this would be to use a logically finite language, i.e. a language that contains only a finite number of (pairwise) non-equivalent sentences. This would mean that the language has only a finite number of atoms.[6] However, such a language is bound to have gratuitous limits to its expressive power [108, 109]. Consider the following list of sentences:

---

[5] On the implausibility of maxichoice contraction of belief sets, see also [1], [99, pp. 76–77], and [109, p. 33]. Maxichoice contraction is less implausible for belief bases (that are not logically closed) than for belief sets, see [175] and [99, p. 77].

[6] A language is syntactically finite if it has only a finite number of non-identical sentences. All syntactically finite languages are logically finite, but the converse does not hold. For instance, a language that contains the atom a and the conjunction sign is syntactically infinite since it contains the infinite set of sentences $\{a, a\&a, a\&a\&a, \ldots\}$. Contrary to logical finiteness, syntactic finiteness is a property of the language itself (rather than a property of the logic).

v_{50} = Less than 50 paintings by Johannes Vermeer are extant.
v_{51} = Less than 51 paintings by Johannes Vermeer are extant.
v_{52} = Less than 52 paintings by Johannes Vermeer are extant.
...
$v_{1.000.000}$ = Less than 1.000.000 paintings by Johannes Vermeer are extant.
...

I believe in each of the sentences on this list, and therefore my set of beliefs contains infinitely many logically non-equivalent sentences. A logically finite language cannot treat all pairs of sentences on the list as non-equivalent. This is a serious restriction on its expressive power. However, my belief in all of these sentences can be represented by a finite-based belief set. The reason for this is that all the sentences on this infinite list follow logically from the first of them, viz. v_{50}. Therefore a belief set that contains v_{50} implies all of the others. This example shows that the requirement of finite-basedness allows for much more expressive power than that of a logically finite language.

Let us now apply the criterion of finite-basedness to the two types of intermediates used in the AGM approach, namely remainders and possible worlds. It is fairly easy to show that neither of them can be finite-based if the language is logically infinite. In addition we can show that both of them will come in infinite numbers:

Observation 2.4 ([109]) *Let the language $\mathcal{L}$ consist of infinitely many logically independent atoms and their truth-functional combinations. Let K be a belief set and let $p \in K \setminus \mathrm{Cn}(\varnothing)$. Then:*

(1) If W is a possible world (i.e. $W \in \mathcal{L} \perp \perp$), then W is not finite-based.

(2) There are infinitely many W such that $p \in W \in \mathcal{L} \perp \perp$.

(3) If $X \in K \perp p$, then X is not finite-based, and

(4) $K \perp p$ is infinite.

Hence, if the language is logically infinite, then all remainders and all possible worlds lack a finite representation. Furthermore, the remainders or possible worlds that we have to select among in a partial meet contraction or a sphere-based revision are always infinite in number.

Thus, even if both the original belief set (K) and the outcome of an operation ($K \div p$ or $K * p$) are finite-based, the transition from the former to the latter requires that we create an infinite set of irreducibly infinite entities, which are then eliminated (through intersection). In other words, the road from a finite-based belief set to another finite-based belief set takes a detour into Cantor's paradise. For those of us who are in favour of cognitive realism and linguistic representability, this is not a desirable deviation.

Someone might wish to argue that this excursion into infinity is useful and perhaps even necessary since we are trying to model the doxastic behaviour of rational agents rather than that of actual agents. Supposedly, results obtained for ideal rational agents with transfinite reasoning powers have normative force as ideals for actual agents. However, the best use of limited cognitive resources may require that one

follows principles and processes that would not be useful for logically omniscient beings. Therefore, normative guidance is best obtained from studies of another type of ideal agents, namely agents that have limited cognitive capacity of which they make rational use.