COSC462 Lecture 7: Belief change

Willem Labuschagne
University of Otago

Abstract

We briefly return to ideas from Lecture 1 in order to examine the belief architecture of an agent. We distinguish between definite and defeasible beliefs. Defeasible beliefs may need revision in the light of new evidence, even though the system may still be in the same state. We describe the AGM approach to belief revision, and give an indication of research directions.

1 Introduction

We know that defeasible reasoning is a process that takes an agent from an hypothesis to plausible conclusions, and preferential semantics is one of the ways in which to give a semantic justification for such defeasible leaps. Preferential semantics tells us which defeasible beliefs would be justified by this or that definite information. Can preferential semantics also tell us how the agent ought to change her beliefs when new information reveals that they are wrong? At roughly the same time that preferential semantics was being developed, but quite independently, three logicians called Alchourrón, Gärdenfors, and Makinson were suggesting general principles that ought to govern the process of belief change. It subsequently transpired that AGM belief change fits together extremely well with preferential semantics, and that preferential semantics provides all the ways of changing belief that conform to the AGM principles. This is one of the reasons we regard preferential semantics as the principal approach to nonmonotonic logic.

2 An architecture of belief

Suppose an agent is observing a finite system and using a knowledge representation language L_A with a finite set A of atoms. Let the states

of the system be represented by a finite set S, and assume there is some labelling function $V : S \longrightarrow W_A$ connecting the system and language. Recall that an agent's basic question is 'What is the actual state of the system?' and that the agent tries to narrow down the possibilities by applying fixed information, evidence, and default (indefinite, heuristic) information in turn.

The agent's definite information (composed of fixed information plus state-dependent evidence) is reflected by a division of the set S of states into two subsets: a set C_{fe} of candidates for being the actual state and the complementary set of excluded states. Thus we may represent an agent's definite information semantically by means of a finite ranked interpretation $\mathcal{I}_{fe} = (S, \preccurlyeq, V)$ where $\preccurlyeq$ is the total preorder on S which places all the states in C_{fe} into the bottom level and all the excluded states into the level above. (Candidates go to the bottom, because candidates are 'normal' or 'likely to be the case'. Excluded states go to the top, because excluded states are 'abnormal' or 'unlikely to be the case'.)

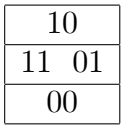
Example 1 *Let $A = \{p, q\}$, as is appropriate for an agent observing the Light-Fan System. The set of states is $S = \{11, 10, 01, 00\}$, in other words $S = W_A$ and V is just the identity function. Now suppose the agent can see that the light is on but cannot tell whether the fan is on. The agent would view 11 and 10 as candidates (i.e. $C_{fe} = \{11, 10\}$ in this case) and thus $\preccurlyeq$ would be the total preorder such that $11 \preccurlyeq 01$, $11 \preccurlyeq 00$, $10 \preccurlyeq 01$, $10 \preccurlyeq 00$ (the candidates are below the excluded states), $11 \preccurlyeq 10$, $10 \preccurlyeq 11$ (the candidates are all on the same level, which must therefore be the bottom level), $01 \preccurlyeq 00$, $00 \preccurlyeq 01$ (the excluded states are all on the same level, which must therefore be the only other level), and finally, for the sake of ensuring reflexivity and transitivity $11 \preccurlyeq 11$, $10 \preccurlyeq 10$, $01 \preccurlyeq 01$, $00 \preccurlyeq 00$. A diagram is easier to comprehend:*

01	00
11	10

Now suppose that, in addition to definite information, an agent has a default rule which ranks some states as being more normal than others. Assume that the default rule is represented by a total preorder $\sqsubseteq$ on S. This preorder may involve more than two levels, although it needn't. Here is an example.

Example 2 *Suppose the agent observing the Light-Fan System has a default rule according to which the Light-Fan System is most likely to be in state 00, a bit less likely to be in 01, just as likely to be in 11 as in 01, and hardly ever happens to be in state 10. (Perhaps the Light-Fan*

system is a nuclear powerplant which spends most of the time switched off so that leaks can be repaired, is equally likely to be testing the cooling subsystem as it is to be in full operation, and is never intentionally in heat-producing mode when the cooling system is off.) Then this default rule can be represented by the total preorder given by $00 \sqsubseteq 01$, $00 \sqsubseteq 11$, $01 \sqsubseteq 10$, $11 \sqsubseteq 10$ *(to establish the three levels with* 00 *in the bottom level),* $01 \sqsubseteq 11$, $11 \sqsubseteq 01$ *(so* 11 *and* 01 *are on the same level, which by the preceding must be the middle level), the pairs needed for reflexivity:* $00 \sqsubseteq 00$, $01 \sqsubseteq 01$, $11 \sqsubseteq 11$, $10 \sqsubseteq 10$, *and lastly those for transitivity:* $00 \sqsubseteq 10$.

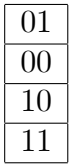


The agent may use its default rule to refine its definite information in the following way.

Definition 3 ***(Lexicographic refinement)*** *Given two total preorders* $\preccurlyeq$ *and* $\sqsubseteq$ *on the same set* S*, the lexicographic refinement of* $\preccurlyeq$ *by* $\sqsubseteq$ *is the total preorder* $\leq$ *given by* $s \leq t$ *iff either* $s \prec t$ *or else* $s \sqsubseteq t$.

In other words, if the original ordering $\preccurlyeq$ distinguishes between s and t by putting them on different levels, so does the refined ordering $\leq$. On the other hand, if s and t are placed on the same level by $\preccurlyeq$ but $\sqsubseteq$ distinguishes between them by (say) placing s on a level below t, then the lexicographic refinement $\leq$ distinguishes them by putting s below t.

Example 4 *For the agent observing the Light-Fan System, the lexicographic refinement of* $\preccurlyeq$ *by* $\sqsubseteq$ *is the ordering* $\leq$ *given by* $11 \leq 10$, $10 \leq 00$, $00 \leq 01$ *plus the pairs needed for reflexivity and transitivity. The set* $C_{fe} = \{11, 10\}$ *of candidates has now been reduced to* $C_{fed} = \{11\}$*, because the bottom level contains only* 11.



We use a ranked interpretation to represent what's in an agent's mind. So we need not restrict ourselves to the agent's definite information but may include the effect of a default rule.

Definition 5 *(**Semantic representation of information**) Given a finite ranked interpretation $\mathcal{I}_{fe} = (S, \preccurlyeq, V)$ as a semantic representation of the agent's definite information, and an ordering $\sqsubseteq$ on S representing a default rule, the result of refining the agent's definite information by the default rule is the interpretation $\mathcal{I}_{fed} = (S, \leq, V)$ where $\leq$ is the lexicographic refinement of $\preccurlyeq$ by $\sqsubseteq$.*

Now that we know how to semantically represent the agent's information, what about a symbolic representation by means of sentences of the knowledge representation language L_A?

Definition 6 *(**Definite beliefs**) If $\mathcal{I}_{fe} = (S, \preccurlyeq, V)$ is a semantic representation of the agent's definite information, the agent's set of definite beliefs is the set K_{fe} of all sentences satisfied by all the states in the bottom level of $\preccurlyeq$ (recall that these are the states in C_{fe}).*

It is not hard to see that the set of definite beliefs is closed under classical entailment, by which we mean that if α belongs to K_{fe} and $\alpha \vDash \beta$, then β also belongs to the belief set. (Why?) Another way to put it is that if $K_{fe} \vDash \beta$, then $\beta \in K_{fe}$.

Definition 7 *(**Defeasible beliefs**) Suppose that $\sqsubseteq$ is a default rule and let $\mathcal{I}_{fed} = (S, \leq, V)$ represent the result of refining the agent's definite information by the default rule, so that $\leq$ is the lexicographic refinement of $\preccurlyeq$ by $\sqsubseteq$. Then the set of states at the bottom level (according to the refined ordering $\leq$) is C_{fed}, and we may take the agent's set of defeasible beliefs to be the set K_{fed} of all sentences satisfied by all the states in C_{fed}.*

Note that the set of defeasible beliefs is also closed under classical entailment (*i.e.* if α is in K_{fed} and $\alpha \vDash \beta$ then β is also a defeasible belief). Another way to put it is that if $K_{fed} \vDash \beta$ then $\beta \in K_{fed}$.

What is the relationship between the set K_{fe} of definite beliefs and the set K_{fed} of defeasible beliefs?

Theorem 8 *Let $\alpha \in K_{fe}$ and let $\beta \in K_{fed}$. Then $\alpha \mathrel{|\!\sim} \beta$, where $\mathrel{|\!\sim}$ is the defeasible entailment relation induced by $\mathcal{I}_{fed}$.*

Proof. *Suppose $\alpha \in K_{fe}$. Then $C_{fe} \subseteq \mathcal{M}(\alpha)$ and $C_{fed} \subseteq C_{fe}$, so that $C_{fed} \subseteq \mathcal{M}(\alpha)$. Now in the ordering $\leq$ which represents the refinement of definite information $\preccurlyeq$ by a default rule $\sqsubseteq$, C_{fed} is the collection of states on the bottom level and since each of these is a model of α, these must be the minimal models of α. That is, in $\mathcal{I}_{fed}$, $\mathcal{M}in(\alpha) = C_{fed}$.*

Now $\beta \in K_{fed}$, so $C_{fed} \subseteq \mathcal{M}(\beta)$. Thus $\mathcal{M}in(\alpha) \subseteq \mathcal{M}(\beta)$. It follows that if $\alpha \in K_{fe}$ and $\beta \in K_{fed}$ then $\alpha \mathrel{|\!\sim} \beta$. The converse also follows, namely that if $\alpha \in K_{fe}$ and $\alpha \mathrel{|\!\sim} \beta$, then $\beta \in K_{fed}$. ■

The theorem tells us the defeasible beliefs of the agent are the sentences defeasibly entailed by K_{fe}, while the agent's definite beliefs are the sentences classically entailed by K_{fe}. This gives us another way to think of the difference between the entailment relations $\models$ and $\mid\!\sim$.

Exercise 9 *1. Consider the Light-Fan System, with atoms p and q expressing that the light and fan are on respectively, and think of it as a metaphor for a helicopter. The rotor (fan) provides lift and is augmented by a jet (since the jet uses heat to produce expansion of air, we model it by the light).*

- *Suppose the agent observing the system notices that the rotor is spinning but cannot tell, because of the distance, whether the jet is being used. Draw a diagram depicting the ordering on states that represents the agent's definite information.*
- *Suppose the agent has learnt that it is very unusual for the jet to be on while the rotor is off, it is normally the case that the helicopter is on the ground with everything switched off, and it is less normal but not very unusual for the helicopter to be flying, in which case it is equally likely to have just the rotor on as it is to have both rotor and jet on. Draw a diagram depicting the ordering on states that represents the agent's default rule.*
- *Now draw the lexicographic refinement of the first ordering by the second.*
- *Give an example of a sentence α that is in the agent's set of definite beliefs and a sentence β that is in the agent's set of defeasible beliefs but not in its set of definite beliefs.*

2. Consider the 3-Card System with $S = \{rgb, rbg, grb, gbr, brg, bgr\}$.

- *Suppose player 1 looks at his card and sees it is red. Draw a diagram depicting the ordering on states that represents player 1's definite information.*
- *Suppose player 1 suspects player 2 of being a bit of a cheat and rigging the deal whenever possible to get the green card (because the green card is worth more in the game that they are playing). Draw a diagram depicting player 1's default rule that player 2 more often than not manages to get the green card.*
- *Now draw the lexicographic refinement of the first ordering by the second.*

- *Give a sentence α that is in player 1's set of definite beliefs and a sentence β that is in player 1's set of defeasible beliefs but not in player 1's set of definite beliefs. (You may take the atoms to be of the form c_i where c is one of r, g, b and i is one of $1, 2, 3$.)*

3 AGM postulates

The AGM approach to belief change is very general. It can be formulated for any propositional language L_A, including those with infinite A. We shall simplify matters slightly by insisting that the ontology S be finite, because we want to use total preorders and minimal models, and don't want to grapple with the complications of avoiding infinite descending chains of models. To describe the AGM approach it is convenient to speak of beliefs in general without specifically distinguishing between definite and defeasible beliefs. In the previous section we noted that both K_{fe} and K_{fed} had the property of being closed under classical entailment, and so the following definition should seem natural.

Definition 10 *(**Belief sets**) A set K of sentences is a belief set iff K is **closed** under classical entailment in the sense that whenever β is a sentence such that $K \models \beta$, then β must be a member of K.*

We may also speak of K being closed under *consequence* — the (classical) consequences of K are the sentences (classically) entailed by K.

If we happen to be working with a language having a finite set A of atoms, then a belief set K is always of the form $Cn(\alpha)$, in other words we can always find a single sentence which belongs to K and whose consequences are precisely the consequences of K. After all, to build α, all we need do is to start with the models of K and to build a sentence in SDNF having precisely those models. This gives us α, and we know α was in K because α is entailed by K and K is closed under entailment.

Example 11 *Suppose we have the language L_A with $A = \{p, q\}$ and ontology $S = \{11, 10, 01, 00\}$. Take $K = \{p, p \vee q, p \vee \neg q, p \vee \neg p\}$. (We leave out all sentences equivalent to these, although strictly speaking such equivalent sentences are also in K. Just imagine them to be invisibly present in K.)*

Is it clear that K is a belief set? $\mathcal{M}(K) = \{11, 10\}$, and all sentences satisfied by these states are included in K or are equivalent to something in K.

$K = Cn(\{p\})$, since $\mathcal{M}(p) = \mathcal{M}(K)$. Of course, we could easily have constructed a sentence in SDNF that fulfils the same role, namely

$(p \wedge q) \vee (p \wedge \neg q)$. *This sentence is equivalent to* p *and so it makes no difference whether we write* $Cn(\{p\})$ *or write* $Cn(\{(p \wedge q) \vee (p \wedge \neg q)\})$.

Suppose K is a belief set and the sentence φ represents new information. The obvious way to incorporate the new belief is to keep all the old beliefs in K, add to them the new belief φ, and then add in all the new consequences that now follow.

Definition 12 *(**Expansion**)* $K + \varphi = Cn(K \cup \{\varphi\})$.

Example 13 *Suppose we again have the language* L_A *with* $A = \{p, q\}$, *ontology* $S = \{11, 10, 01, 00\}$, *and belief set* $K = Cn(\{p\})$.

Take $\varphi = q$. *The expansion is* $K + \varphi = Cn(K \cup \{\varphi\}) = Cn(\{p,\ p \vee q,\ p \vee \neg q,\ p \vee \neg p,\ q\})$ *and an example of one of the new consequences is the sentence* $p \wedge q$, *which is not in* $K \cup \{\varphi\}$ *but must be added because* $p \wedge q$ *is satisfied by every model of* $K \cup \{\varphi\}$. *Careful examination reveals that* $K + \varphi = \{p \wedge q,\ p,\ q,\ p \leftrightarrow q,\ p \vee q,\ p \vee \neg q,\ \neg p \vee q,\ p \vee \neg p\}$, *where we leave out equivalent sentences.*

How do we know that this is what $K \cup \{\varphi\}$ *looks like? Well, there are 4 states in* S, *and so there are 16 different sets of states, from the smallest,* $\varnothing$, *to the largest,* S *itself. For each set* X *of states, pick one sentence* α *such that* $\mathcal{M}(\alpha) = X$. *I have chosen the sentences* $p \wedge \neg p$, $p \wedge q$, $p \wedge \neg q$, $\neg p \wedge q$, $\neg p \wedge \neg q$, p, q, $p \leftrightarrow q$, $\neg(p \leftrightarrow q)$, $\neg p$, $\neg q$, $p \vee q$, $p \vee \neg q$, $\neg p \vee q$, $\neg p \vee \neg q$, *and* $p \vee \neg p$ *respectively. Next, find the set of models of* $K \cup \{\varphi\}$. *Clearly this is* $\{11\}$ *as every other state makes either* p *or* q *or both false. Now it is a simple matter to decide which sentences are satisfied by all models of* $K \cup \{\varphi\}$. *We simply look for the sentences whose sets of models include* 11.

Equally clearly, we may write $K + \varphi = Cn(\{p \wedge q\})$.

Although expansion is a simple way to incorporate new beliefs, there is a potential problem. It is possible that φ is *inconsistent* with K, in the sense that the information in φ contradicts some of the information in K. We don't want to prevent an agent learning something new, but equally we don't want the agent confused by contradictory information.

Why might φ be inconsistent with K? This can happen for two different reasons:

- If K is the agent's set of definite beliefs, and the system changes state, then a subsequent observation may reveal information which, when expressed declaratively, results in a sentence φ inconsistent with K. For example, the agent sees that the light is on, believes p, the system changes state so that the light is off, and subsequent observation causes the agent to form the belief $\neg p$.

- On the other hand, if K is the agent's set of defeasible beliefs, then even without the system changing its state the agent may learn something that shows some of its defeasible beliefs to have been in error. For example the agent observes that the light is on, and by virtue of a default rule concludes (defeasibly) that the light and fan are both on, only to learn subsequently that the state is actually that in which the light is on but the fan is off.

In either case, the set K must be changed, but in different ways. Let us focus on the second situation, calling this type of change *belief revision*. (The situation in which the system changes state will be omitted from this brief introduction, but for convenience we shall call the type of change involved *update*.) Before going on to explore belief revision, we must get a good grasp of what the word 'inconsistent' means in the present context.

Definition 14 ***(Inconsistency)*** *A sentence φ is inconsistent with a set of sentences Γ iff the set $\Gamma \cup \{\varphi\}$ is unsatisfiable, by which we mean that there is no $v \in W_A$ such that $v \in \mathcal{M}(\varphi)$ and $v \in \mathcal{M}(\Gamma)$. If $\Gamma = \{\gamma\}$ then φ is inconsistent with γ iff $\mathcal{M}(\varphi \wedge \gamma) = \emptyset$.*

For example, take $\Gamma = \{p, p \vee q, p \vee \neg q, p \vee \neg p\}$ and let $\varphi = \neg p$. Is it clear that φ is inconsistent with Γ?

Theorem 15 *φ is inconsistent with Γ iff $\neg\varphi \in Cn(\Gamma)$.*

Proof. *If φ is inconsistent with Γ, then $\mathcal{M}(\varphi) \cap \mathcal{M}(\Gamma) = \emptyset$, and so $\mathcal{M}(\Gamma) \subseteq \mathcal{M}(\neg\varphi)$. (Recall that $\mathcal{M}(\neg\varphi)$ is the complement of $\mathcal{M}(\varphi)$ in W_A.) Thus $\neg\varphi \in Cn(\Gamma)$.*

Conversely, if $\neg\varphi \in Cn(\Gamma)$ then $\Gamma \models \neg\varphi$ and thus all the models of Γ satisfy $\neg\varphi$. Hence there is no model of $\Gamma \cup \{\varphi\}$. ■

What is the cash value of this theorem? Well, suppose Γ is a belief set, say $\Gamma = \{p, p \vee q, p \vee \neg q, p \vee \neg p\}$. Then all the consequences of Γ are mentioned in the set (except that it is convenient to regard equivalent sentences as being invisibly present). If we want to know whether a new sentence φ is inconsistent with Γ, we can look through Γ and check whether Γ contains the sentence $\neg\varphi$ (or some sentence equivalent to $\neg\varphi$).

Theorem 16 *If a set of sentences Γ is unsatisfiable, then $Cn(\Gamma) = L_A$.*

Proof. *Suppose Γ is unsatisfiable. Then $\mathcal{M}(\Gamma) = \emptyset$. Let $\alpha \in L_A$. Clearly $\mathcal{M}(\Gamma) \subseteq \mathcal{M}(\alpha)$, so $\alpha \in Cn(\Gamma)$.* ■

This theorem tells us why we don't want inconsistencies to infect the beliefs of agents — an agent that builds an unsatisfiable belief set ends up believing everything! This is no better than believing nothing.

If new information φ is inconsistent with a belief set K, then in order to incorporate φ some of the information in K must be given up. Specifically, information contradicting φ must be sacrificed. By a contradiction we understand a sentence of the form $\alpha \wedge \neg\alpha$, so the idea is that if φ is inconsistent with K then the information expressed by $\neg\varphi$ must be lurking somewhere in K, either in explicit form as the sentence $\neg\varphi$ or expressed in disguised form by other sentences. Giving up the information in $\neg\varphi$ may occasionally mean simply taking the sentence $\neg\varphi$ out of K, but more usually it means we have to remove from K a whole lot of sentences that don't look like $\neg\varphi$. For example, suppose $K = \{p, p \vee q, p \vee \neg q, p \vee \neg p\}$ and we want to revise K so as to incorporate the new information $\varphi = \neg(p \vee \neg q)$. Clearly K does contain a sentence equivalent to $\neg\varphi$, namely the sentence $p \vee \neg q$. But it is not enough simply to take out the old $p \vee \neg q$ and put in the new $\neg(p \vee \neg q)$, because the result still contains p, and $p \vee \neg q$ is a consequence of p, so since belief sets are closed under consequence, the dratted $p \vee \neg q$ automatically slips back in, giving an unsatisfiable set $\{p, p \vee q, p \vee \neg q, \neg(p \vee \neg q), p \vee \neg p\}$. To really give up the information $p \vee \neg q$, we would need to remove p from K as well.

The example shows that removing information from K is more complicated than simply deleting a sentence from a list. It is even conceivable that there may be more than one way in which to remove from K the information that contradicts φ, and so the logicians Alchourrón, Gärdenfors, and Makinson gave some very general constraints on how the removal should be done. Specifically, the AGM postulates for *contraction* give eight conditions that should be fulfilled by the set $K - \neg\varphi$ that results from the removal of the information contradicting φ. (Note that no specific operation for accomplishing the removal is given at this stage.)

Similarly, the AGM postulates for *revision* describe eight conditions that should be fulfilled by the set $K * \varphi$ that results from the incorporation of the new information φ after removal of contradictory information. Recalling that $K + \varphi = Cn(K \cup \{\varphi\})$, we may think of revision as being accomplished as follows:

- (The Levi identity) $K * \varphi = (K - \neg\varphi) + \varphi$.

Thus we might arrive at the revised belief set $K * \varphi$ by first removing from K any information contradicting φ and then adding φ to the result (and adding in all the consequences). According to this approach, if we

know how to contract a belief set we will automatically be able to revise a belief set. Of course, this makes revision a clumsy two-step process, and we may be interested in more direct ways to do it. We shall describe a very simple and direct way to do revision later in this section.

Definition 17 *(**AGM postulates for contraction**) Given a belief set K, a contraction operation $-$ is an operation such that, for every sentence α, the set $K - \alpha$ (the result of contracting K with α) satisfies the following:*

1. $K - \alpha = Cn(K - \alpha)$
2. $K - \alpha \subseteq K$
3. $K - \alpha = K$ *if* $\alpha \notin K$
4. *If* α *is not a tautology,* $\alpha \notin K - \alpha$
5. *If* $\alpha \equiv \beta$*, then* $K - \alpha = K - \beta$
6. *If* $\alpha \in K$*, then* $(K - \alpha) + \alpha = K$ *(the Recovery Postulate)*
7. $(K - \alpha) \cap (K - \beta) \subseteq K - (\alpha \wedge \beta)$
8. *If* $\beta \notin K - (\alpha \wedge \beta)$ *then* $K - (\alpha \wedge \beta) \subseteq K - \beta$.

The first postulate says that $K - \alpha$ must again be a belief set.

The second ensures that contraction reduces the belief set and does not mysteriously expand it in some way.

The third ensures that information is not taken away unnecessarily.

The fourth ensures that contraction with any sentence that actually contains semantic information will take that information away.

The fifth ensures that it is semantic information that is playing a role and that the contraction operation is not sensitive to the syntactic form of α.

The sixth, known as the Recovery Postulate, requires the contraction of K with α to retain so much of the information of K that it is possible to recover the original K by putting α back (and adding in the consequences).

The last two postulates are included for technical reasons and are less intuitive, but may be thought of as saying that the information removed when contracting with $\alpha \wedge \beta$ must not be more than would be removed by contracting with α and with β separately.

Definition 18 *(**AGM postulates for revision**) Given a belief set K, a revision operation $*$ is an operation such that, for every sentence α, the set $K * \alpha$ (the result of revising K by α) satisfies the following:*

1. $K * \alpha = Cn(K * \alpha)$
2. $K * \alpha \subseteq K + \alpha$
3. $K * \alpha = K + \alpha$ *if* $\neg\alpha \notin K$
4. $\alpha \in K * \alpha$
5. *If* $\alpha \equiv \beta$, *then* $K * \alpha = K * \beta$
6. $K * \alpha = L_A$ *iff* α *is a contradiction*
7. $K * (\alpha \wedge \beta) \subseteq (K * \alpha) + \beta$
8. $(K * \alpha) + \beta \subseteq K * (\alpha \wedge \beta)$ *if* $\neg\beta \notin K * \alpha$.

The first revision postulate says that $K * \alpha$ must again be a belief set.

The second ensures that revision by α introduces no more information than is contained in α.

The third ensures that if K contains no information contradicting α then the revised belief set keeps all the information in K and adds to it all the information of α.

The fourth postulate gives precedence to the information in α by ensuring that if any information has to be sacrificed to avoid inconsistency, it will be information in K and not α.

The fifth postulate makes the point that it is semantic information that is at issue, not syntactic form.

The sixth postulate expresses the idea that any information removed from K during revision should be such as will render the resulting belief set satisfiable, but recognises that because of the fourth postulate, there is one case in which this is not possible.

The last two postulates may be thought of as saying that under certain conditions, the result of successive revisions by α and then β is the same as the result of simultaneous revision by these sentences in the form of $\alpha \wedge \beta$.

We are now at the point where we have to face the million dollar question: How, exactly, should contraction and revision operations be defined? From our perspective, these operations should modify agents' defeasible beliefs and therefore be determined by the orderings of states that represent the agents' default rules. This is by no means the way in

which Alchourrón, Gärdenfors and Makinson originally thought of the matter. But the desired connection follows from subsequent work by Grove, Katsuno & Mendelzon, and Boutilier and is summarised in the important theorem below.

Definition 19 *(**Faithfulness**) Suppose that K is a belief set and that $\leq$ is a total preorder on W_A. Then $\leq$ is faithful (with regard to K) iff $s < t$ for every $s \in \mathcal{M}(K)$ and $t \notin \mathcal{M}(K)$, and $s \not< t$ for all $s, t \in \mathcal{M}(K)$.*

Recall that if $\leq$ is a preorder then we may write $s < t$ to mean that $s \leq t$ but not $t \leq s$. So a total preorder $\leq$ is faithful relative to belief set K if it places the models of K at the bottom level and all nonmodels of K above them. Examples of faithful total preorders are

- if $\mathcal{I}_{fe} = (S, \preccurlyeq, V)$ represents the agent's definite information and we take $K = K_{fe}$, then $\preccurlyeq$ is faithful with regard to K
- if $\mathcal{I}_{fed} = (S, \leq, V)$ represents the refinement of the agent's definite information by its default rule $\sqsubseteq$ and we take $K = K_{fed}$ then $\leq$ is faithful with regard to K.

Theorem 20 *Let the belief set K be fixed.*

A contraction operation satisfies the 8 contraction postulates iff there is some total preorder $\leq$ which is faithful relative to K and such that, for all φ, $K - \varphi$ is the set of all sentences β such that $(\mathcal{M}(K) \cup \mathcal{M}in(\neg\varphi)) \subseteq \mathcal{M}(\beta)$.

A revision operation satisfies the 8 revision postulates iff there is some total preorder $\leq$ which is faithful with regard to K and such that, for all φ, $K * \varphi$ is the set of all sentences β such that $\mathcal{M}in(\varphi) \subseteq \mathcal{M}(\beta)$.

Proof. *Too hard to include, but the references at the end of the lecture give everything you might need.* ■

The theorem tells us how to work out $K - \varphi$ and $K * \varphi$.

If we have a finite ranked interpretation with ordering $\leq$ and K is the set of all sentences satisfied by the states in the bottom level of $\leq$, then $K * \varphi$ is the set of all β such that $\varphi \mathrel{|\!\sim} \beta$. Moreover, $K - \varphi$ consists of sentences true not only in the models of K but also in the most normal models of $\neg\varphi$. (Here is a potential for confusion. When we speak of $K - \varphi$, then φ stands not for new information to be added but for the old information we want to get rid of. The information carried by φ consists of its nonmodels, i.e. models of $\neg\varphi$. Which is why we now include (some of the) models of $\neg\varphi$ in working out which beliefs to keep in $K - \varphi$.)

Exercise 21 *1. Recall exercise 1 of the previous section. The Light-Fan System was taken to be a metaphor for a helicopter with a rotor (the fan) and a jet (the light). The agent observed the rotor to be spinning but could not tell, because of the distance, whether the jet was being used. You drew a diagram depicting the ordering on states that represents the agent's definite information, and therefore in effect built a ranked interpretation to semantically represent the agent's definite beliefs.*

Next it was assumed that the agent learnt it to be very unusual for the jet to be on while the rotor is off, very usual for the helicopter is on the ground with everything switched off, and less usual but not very unusual for the helicopter to be flying, in which case it was equally likely to have just the rotor on or to have both rotor and jet on. You drew a diagram depicting the ordering on states that represents the agent's default rule.

Finally you drew the lexicographic refinement of the first ordering by the second, thus in effect building the ranked interpretation which represents the result of refining the agent's definite beliefs by her default rule.

Taking for the belief set K the sentences satisfied by the states in the bottom level of the refined ordering, work out

- $K - p$
- $K - \neg p$
- $K * p$
- $K * \neg p$
- $K - q$
- $K - \neg q$
- $K * q$
- $K * \neg q$
- $K - (p \wedge q)$
- $K - \neg(p \wedge q)$
- $K * (p \wedge q)$
- $K * \neg(p \wedge q)$
- $K - (p \wedge \neg q)$
- $K * (p \wedge \neg q)$
- $K - (p \vee q)$

- $K * (p \vee q)$
- $K - (p \rightarrow q)$
- $K * (p \rightarrow q)$

Hint: To describe the new belief set, you could of course list all its sentences, but this can be tedious. Find the set of models and use this to write the contracted or revised belief set in the form $Cn(\{\alpha\})$.

2. *Return to exercise 2 of the previous section in which we represented player 1's defeasible beliefs in the 3-Card System. Let* K *be player 1's set of defeasible beliefs. Give, for each of the following, its set of models and then express the (contracted or revised) belief set in the form* $Cn(\{\alpha\})$*:*

   - $K - g_2$
   - $K * b_2$
   - $K - b_3$
   - $K * g_3$
   - $K * r_3$
   - $K - (g_2 \rightarrow b_3)$
   - $K * (g_2 \rightarrow b_1)$

3. *Recall the Light-Fan-Heater System. Take* $A = \{p, q, r\}$*, where* p *says that the light is on,* q *that the fan is on, and* r *that the heater is on. We take the ontology to be* (S, V) *where* $S = W_A = \{111, 110, 101, 011, 100, 010, 001, 000\}$ *and* $V(s) = s$ *for all* $s \in S$*.*

   Assume that the agent can see the heater is on, and has a default rule saying that normally two of the three components are off.

   Give the orderings representing the agent's definite beliefs, default rule, and the lexicographic refinement of the former by the latter.

   - *What is the agent's set of defeasible beliefs,* K*?*
   - *What is* $K - \neg p$*?*
   - *What is* $K * p$*?*
   - *What is* $K - \neg(p \vee q)$*?*
   - *What is* $K * (p \vee q)$*?*
   - *What is* $K - (\neg p \vee \neg q)$*?*

- *What is* $K * (p \wedge q)$*?*
- *What is* $K - (p \leftrightarrow q)$*?*
- *What is* $K * \neg(p \leftrightarrow q)$*?*

4. *Use the revision postulates to prove that*

   - $K * \alpha = K * \beta$ *iff* $\beta \in K * \alpha$ *and* $\alpha \in K * \beta$
   - $(K * \alpha) \cap (K * \beta) \subseteq K * (\alpha \vee \beta)$

5. *Use the contraction postulates to prove that*

   - $K - (\alpha \wedge \beta) \subseteq K - \alpha$ *or* $K - (\alpha \wedge \beta) \subseteq K - \beta$
   - $K - (\alpha \wedge \beta) =$ *either* $K - \alpha$ *or* $K - \beta$ *or* $(K - \alpha) \cap (K - \beta)$

4 Review

The basic paper on belief change was:

- Alchourrón, Gärdenfors and Makinson: On the logic of theory change: Partial meet functions for contraction and revision. *Journal of Symbolic Logic* **50**:510-530 1985.

The papers describing how to connect AGM belief change with preferential semantics are:

- Grove: Two modellings for theory change. *Journal of Philosophical Logic* **17**:157-170 1988.
- Katsuno and Mendelzon: Knowledge base revision and minimal change. *Artificial Intelligence* **52**:263-294 1991.
- Boutilier: Unifying default reasoning and belief revision in a modal framework. *Artificial Intelligence* **68**:33-85 1994.

If you read further in the literature on belief change, keep in mind that (as in the case of nonmonotonic logic) the object languages are usually not restricted to having a finite set A of atoms or a finite set S of states, and so the orderings on S have to satisfy additional properties such as 'smoothness'. (The paper by Kraus, Lehmann and Magidor listed in the references of Lecture 5&6 contains a discussion of smoothness.)

So what should you read if you find the topic of belief change interesting?

4.1 Current research

There are a number of research directions being actively pursued. I will give a highly prejudiced sampling.

Iterated belief revision:

Although AGM belief revision is a step toward completing the picture of how agents ought to change their beliefs in the light of new information, it is still incomplete. The approach describes how to produce, from a belief set K and a faithful ordering $\leq$, a new belief set $K * \varphi$ for any sentence φ, but no clue is given to suggest what the new order relation should be. It cannot be expected that the new ordering should be the same as $\leq$, because $\leq$ need not be faithful with regard to $K * \varphi$. Currently one of the most active areas of research in the field is *iterated belief revision*, which seeks to describe not only how the new belief set is to be produced but also how the new ordering (the *epistemic state*) is to be produced. For an introduction to the latest research on iterated belief change, see

- Darwiche A & Pearl J: On the logic of iterated belief revision. *Artificial Intelligence* 89:1-29 1997.
- Meyer T: Merging Epistemic States. In Mizoguchi and Slaney (eds): *PRICAI 2000: Topics in Artificial Intelligence*, volume 1886 of *Lecture Notes in Artificial Intelligence*, pp 286-296, Springer 2000.

Other kinds of contraction:

Not everyone is convinced that the Recovery Postulate (the 6th contraction postulate) is correct. The Recovery Postulate says that $(K - \varphi) + \varphi = K$ as long as we start with a sentence $\varphi \in K$. Hansson gives an example to show why one might not always want this postulate to hold. Suppose I read a book about Cleopatra, in which the claim is made that she had a son and a daughter. Suppose I then discover that the book is a work of fiction. It would be quite reasonable for me to remove my belief that Cleopatra had any children. In other words, if we use the language with $A = \{p, q\}$ and let p stand for 'Cleopatra had a son' and q for 'Cleopatra had a daughter', then initially $K = Cn\{p \wedge q\}$ $= \{p \wedge q,\ p,\ q,\ p \leftrightarrow q,\ p \vee q,\ p \vee \neg q,\ \neg p \vee q,\ p \vee \neg p\}$. Upon learning that the book was fictional, I change my belief set to $K - (p \vee q)$. For the moment, it is not clear exactly what the new belief set contains (no ordering on states has been given that I can use to construct the new belief set), but at the very least I know that the new belief set cannot contain $p \wedge q$, because if it did then it would also contain $p \vee q$, which has

been successfully removed. Now imagine that I go on to consult a history book and discover that Cleopatra did indeed have at least one child, although no details are given of whether there was more than one nor what the sex(es) may have been. This new information is thus no more than $p \vee q$. I now expand my belief set with the new information $p \vee q$, and by the Recovery Postulate the result should be K. In other words, having learned that Cleopatra had at least one child I should suddenly believe that she had both a son and a daughter (because $p \wedge q \in K$). This would strike many people as unreasonable. Accordingly, research has been conducted on alternative forms of 'contraction' that do not have to satisfy the Recovery Postulate. The paper below describes a family of such approaches, known as 'withdrawal' operations:

- Meyer T, Heidema J, Labuschagne W & Leenen L: Systematic withdrawal. *Journal of Philosophical Logic* **31**:415-443 2002.

Entrenchment:

When an agent is forced to give up some of the information in its belief set, it is not equally ready to give up all beliefs — some beliefs are more entrenched than others. A total preorder on the set of states induces an entrenchment ordering on the sentences of the language. For different approaches to entrenchment, see:

- Gärdenfors P & Makinson D: Revisions of knowledge systems using epistemic entrenchment. In Vardi MY (ed): *Proceedings of the Second Conference on Theoretical Aspects of Reasoning about Knowledge* pp83-95. Morgan Kaufmann 1988.

- Meyer T, Labuschagne W & Heidema J: Refined epistemic entrenchment. *Journal of Logic, Language, and Information* **9**:237-259 2000.

Infobases:

Where does the ordering on the set of states come from? One way to get such an ordering is to use the notion of an infobase. An infobase is a list of sentences representing information derived from independent sources (e.g. sensors or other agents). Every state can now be given a weight by checking to see how many sentences in the infobase it satisfies. Thus if several sources have provided the information p, say, then p appears several times in the infobase and the models of p get a heavy weight. Now an ordering on models may be constructed by using the weights to make the states sink down (heavy states sink down further). For more about infobases, see:

- Meyer T, Labuschagne W & Heidema J: Infobase change: A first approximation. *Journal of Logic, Language, and Information* **9**:353-377 2000.

Update:

It should not be forgotten that there is a distinction between belief *revision* and belief *update*. Intuitively, revision is what is supposed to occur when the agent gets new information but the system has persisted in the same state. Thus the new information is not inconsistent with the agent's definite beliefs, although it may be inconsistent with the agent's defeasible beliefs. Update is what is supposed to occur when the agent gets new information after the system has changed state. Some proposals have been made, for example by Katsuno and Mendelzon. However, the proposals so far published lack a clear criterion which the agent might use to choose between revision and update, and have other weaknesses which suggest that this is a profitable area for research. From the perspective outlined previously, in which we distinguish between definite beliefs and defeasible beliefs, there is a way for the agent to become aware that update is called for. If the new information φ is obtained in some reliable fashion such as making an observation, and φ is inconsistent with the agent's definite beliefs, then the system must have changed state and so update is needed rather than revision. The difference between update and revision is of course that revision should retain all the agent's state-dependent evidence, whereas update should replace the agent's state-dependent evidence with the semantic information in the observation φ. Two difficulties inherent in this suggestion are that the system may be able to change state without the observation φ revealing anything inconsistent with the out-of-date state-dependent evidence of the agent, and that there is heavy reliance on the assumption that agents do not make errors in acquiring their evidence.

For one (rather unsatisfactory) proposal on how to do updates, see:

- Katsuno and Mendelzon: On the difference between updating a knowledge base and revising it. In Gärdenfors (ed): *Belief Revision*, pp 183-203, Cambridge University Press 1992.

Glossary

- **contraction** — changing a belief set K into a belief set $K - \varphi$ by removing the information in φ (which is harder than it seems at first glance).

- **entrenchment** — not all beliefs are held with the same fervour, and two papers are cited in the Review which explain how the ordering on states induces an ordering on sentences that reflects how reluctant an agent is to give up each sentence.

- **expansion** — naively adding a belief φ to a belief set K by forming the new set $K + \varphi = Cn(K \cup \{\varphi\})$.

- **inconsistent** — a sentence φ is inconsistent with a set of sentences Γ if together $\Gamma \cup \{\varphi\}$ is unsatisfiable.

- **lexicographic refinement** — a way of putting together two order relations to get a combined order.

- **revision** — incorporating new information φ into a belief set K by giving up the information in K that contradicts φ; the idea is that the system hasn't changed state but the agent has learned something new which shows some of her defeasible beliefs to have been mistaken, and so now these mistaken beliefs must be removed and the new information added.

- **update** — the way in which a belief set should be changed if the old beliefs become wrong because the system has changed state.