

## Relating Belief Change and Reasoning about Action

Maurice Pagnucco

morri@ics.mq.edu.au

Department of Computing Science  
Division of Information and Communication Sciences  
Macquarie University  
NSW 2109, AUSTRALIA

## Relating Belief Change and Reasoning about Action



## Motivation

Belief revision and belief update are important types of belief change

They are similar in some respects however they are considered to be fundamentally different processes

They are motivated by different styles of belief change.

## Aims

Investigate difference between revision and update more closely

Can one be reduced to the other?

## Overview

Introduction (belief change)

Revision vs. update

AGM Belief Revision

KM Belief Update

Reconstructing Revision Functions from Similarity Structures

Fixed vs. Variable Similarity

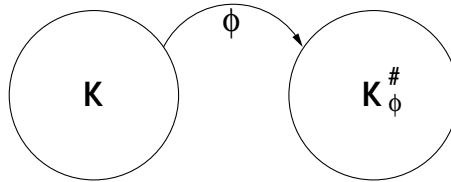
Discussion

Conclusions

## Belief Change

Consider a theory (belief state/knowledge base/. . .)  $K$  in some logic (we'll use propositional logic)

Upon obtaining new information  $\phi$ , how do we modify  $K$  to incorporate  $\phi$ ?



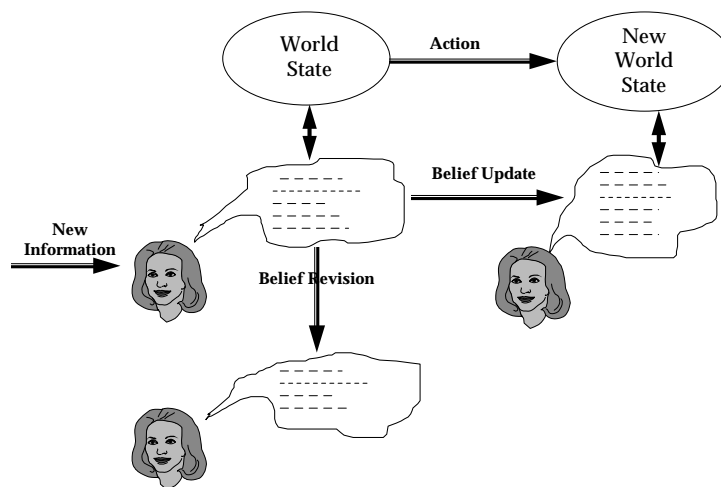
A number of proposals have been suggested in the literature

We shall distinguish two: *belief revision* and *belief update*

**Note:** there are many ways to effect revision and update

Belief change is related to nonmonotonic logics, conditionals, . . .

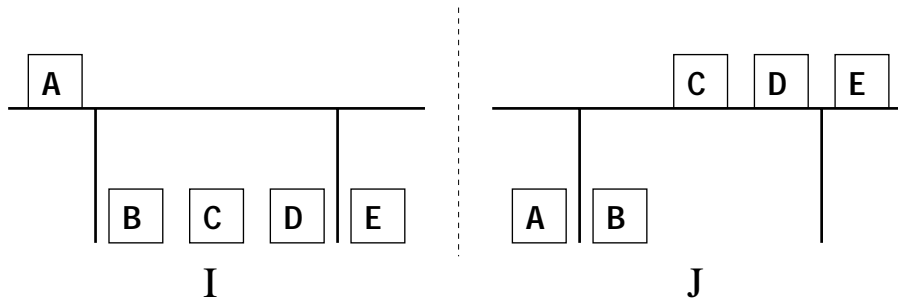
LSS 99, Friday 17 December, 1999 Relating Belief Change and Reasoning about Action5-1



## Revision versus Update

Consider the following example due to Katsuno & Mendelzon (1990)

$$(a \wedge \neg b \wedge \neg c \wedge \neg d \neg e) \vee (\neg a \wedge \neg b \wedge c \wedge d \wedge e)$$



## Measuring Similarity (or “Closeness”)

Let’s adopt a very simple measure of similarity suggested by Dalal (1988):

Count number of propositional letters on which two models differ

$$dist(I, L) = 4$$

$$dist(J, L) = 2$$

$$dist(I, K) = 1$$

$$dist(J, K) = 3$$

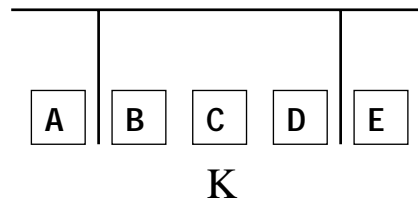
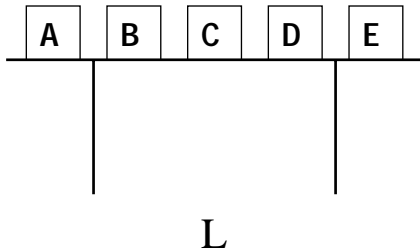
Revising  $[I \vee J]$  by  $[L \vee K]$  gives  $K$

Updating  $[I \vee J]$  by  $[L \vee K]$  gives  $L \vee K$

## Revision versus Update

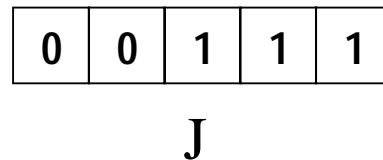
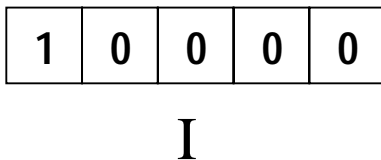
Instruct robot to enter room and “clean it up” (i.e., ensure all objects are on table or on floor)

$$(a \wedge b \wedge c \wedge d \wedge e) \vee (\neg a \wedge \neg b \wedge \neg c \wedge \neg d \wedge \neg e)$$

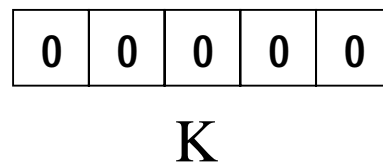
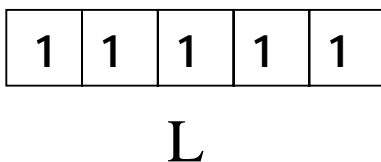


## Revision versus Update

Five-bit register



Learn that circuits have the same value



## Preliminaries

Language  $\mathcal{L}$  closed under all boolean connectives

- (i)  $\vdash \varphi$  for all truth functional tautologies  $\varphi$
- (ii) If  $\vdash (\phi \rightarrow \psi)$  and  $\vdash \phi$ , then  $\vdash \psi$
- (iii)  $\vdash$  consistent
- (iv)  $\vdash$  Satisfies deduction theorem
- (v)  $\vdash$  Compact

Belief Sets (theories) — closed under  $Cn$ .

$\mathcal{K}_{\mathcal{L}}$  — set of all belief sets.

$\mathcal{M}_{\mathcal{L}}$  — set of all consistent, complete theories (possible worlds).

$K + \phi = Cn(K \cup \{\phi\})$

## AGM Belief Revision

Rational agent changes beliefs about static world in light of new information

Revision function  $*$  :  $\mathcal{K}_{\mathcal{L}} \times \mathcal{L} \rightarrow \mathcal{K}_{\mathcal{L}}$ .

- (K\*1)  $K * \varphi$  is a theory of  $\mathcal{L}$ .
- (K\*2)  $\varphi \in K * \varphi$ .
- (K\*3)  $K * \varphi \subseteq K + \varphi$ .
- (K\*4) If  $\neg\varphi \notin K$  then  $K + \varphi \subseteq K * \varphi$ .
- (K\*5)  $K * \varphi = \mathcal{L}$  iff  $\vdash \neg\varphi$ .
- (K\*6) If  $\vdash \varphi \leftrightarrow \psi$  then  $K * \varphi = K * \psi$ .
- (K\*7)  $K * (\varphi \wedge \psi) \subseteq (K * \varphi) + \psi$ .
- (K\*8) If  $\neg\psi \notin K * \varphi$  then  $(K * \varphi) + \psi \subseteq K * (\varphi \wedge \psi)$ .

## Possible Worlds Construction

Proposed by Grove (1988) who adapted Lewis' (1973) system of spheres modelling for counterfactual conditionals.

Possible worlds — consistent, complete theories.

$[K]$  — worlds consistent with agent's beliefs.

$[\varphi]$  — worlds consistent with  $\varphi$ .

System of spheres centred on  $X \subseteq \mathcal{M}_{\mathcal{L}}$

(S1)  $\mathcal{S}$  is totally ordered with respect to set inclusion.

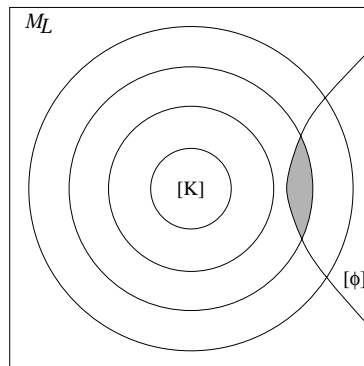
(S2)  $X$  is smallest sphere in  $\mathcal{S}$ .

(S3)  $\mathcal{M}_{\mathcal{L}} \in \mathcal{S}$ .

(S4) For every consistent  $\varphi \in \mathcal{L}$ , there is a smallest sphere in  $\mathcal{S}$  intersecting  $[\varphi]$ .

$C_{\mathcal{S}}(\varphi)$  — innermost (smallest) sphere intersecting  $[\varphi]$

$f_{\mathcal{S}}(\varphi) = C_{\mathcal{S}}(\varphi) \cap [\varphi]$ .



$$(\text{Def } \mathcal{S} \text{ to } *) \quad K * \varphi = \begin{cases} th(f_{\mathcal{S}}(\varphi)) & \text{if } [\varphi] \neq \emptyset \\ \mathcal{L} & \text{otherwise} \end{cases}$$

System of spheres centred on  $[K]$  is an ordering of relative similarity over possible worlds. Closer world is to centre, more “similar” it is to worlds in  $[K]$ .

(Def  $\mathcal{S}$  to  $*$ ) Revision given by  $\varphi$ -worlds most similar to  $[K]$ .

### **Belief Update (Katsuno & Mendelzon 1992)**

Rational agent changes belief about evolving world in light of new information.  $\diamond : \mathcal{K}_{\mathcal{L}} \times \mathcal{L} \rightarrow \mathcal{K}_{\mathcal{L}}$ .

- (K $\diamond$ 1)  $K \diamond \varphi$  is a theory.
- (K $\diamond$ 2)  $\varphi \in K \diamond \varphi$ .
- (K $\diamond$ 3) If  $\varphi \in K$  then  $K \diamond \varphi = K$ .
- (K $\diamond$ 4)  $K \diamond \varphi = \mathcal{L}$  iff  $K$  or  $\varphi$  is inconsistent.
- (K $\diamond$ 5) If  $\vdash \varphi \leftrightarrow \psi$  then  $K \diamond \varphi = K \diamond \psi$ .
- (K $\diamond$ 6)  $K \diamond (\varphi \wedge \psi) \subseteq (K \diamond \varphi) + \psi$ .
- (K $\diamond$ 7) If  $K$  complete,  $\neg \psi \notin K \diamond \varphi$  then  $(K \diamond \varphi) + \psi \subseteq K \diamond (\varphi \wedge \psi)$ .
- (K $\diamond$ 8) If  $[K] \neq \emptyset$  then  $K \diamond \varphi = \bigcap_{w \in [K]} w \diamond \varphi$ .

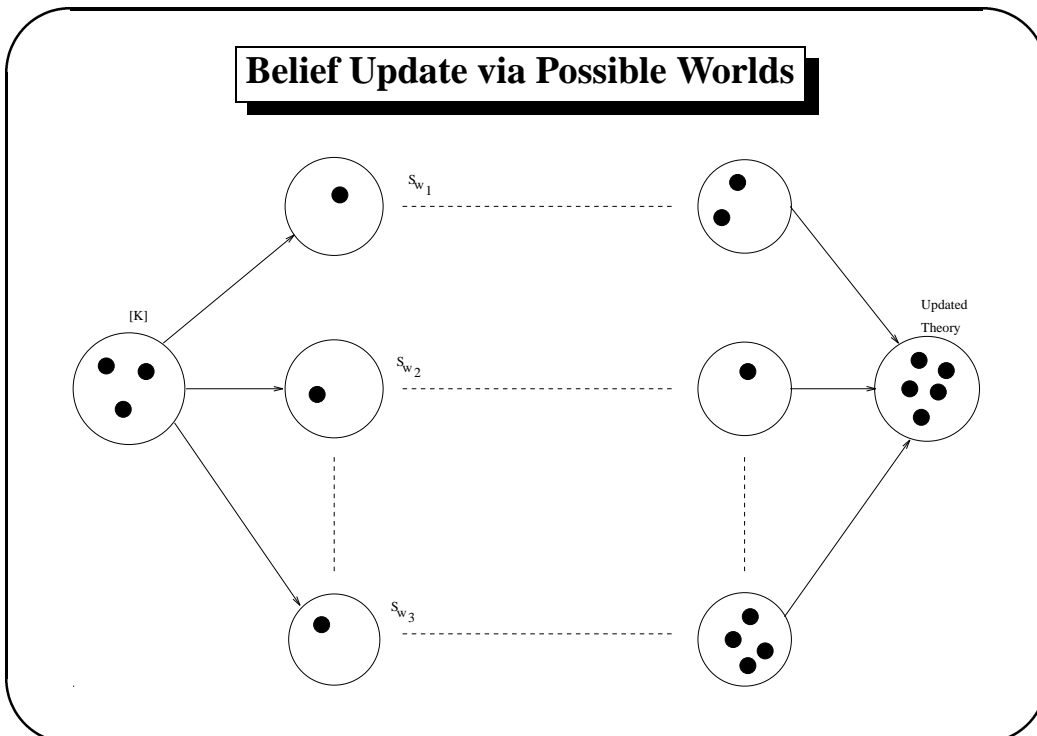
Equivalent to KM postulates (U1) — (U9).

Similarity structure  $\mathbf{S} : \mathcal{M}_{\mathcal{L}} \rightarrow \mathcal{S}$ .

To every possible world  $w$  assigns a system of spheres  $\mathcal{S}_w$  centred on  $w$ .

$$(\text{Def } \mathbf{S} \text{ to } \diamond) \quad K \diamond \varphi = \begin{cases} th \left( \bigcup_{w \in [K]} f_{\mathbf{S}_w}(\varphi) \right), & \text{if } [\varphi], [K] \neq \emptyset \\ \mathcal{L} & \text{otherwise} \end{cases}$$

Corresponds to KM *faithful assignment*.



## Reconstructing Revision Functions from Similarity Structures

$K_{\perp} * \varphi$  consistent whenever  $\varphi$  consistent

$$K_{\perp} \diamond \varphi = K_{\perp}.$$

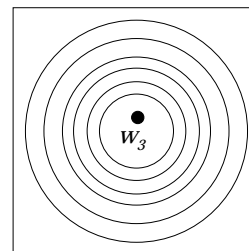
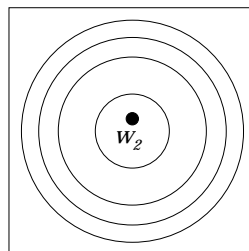
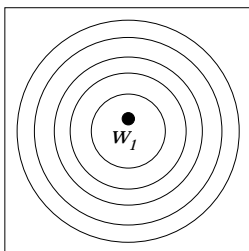
Shall disregard this limiting case.

Behaviour of  $*$  at  $K$  determined by single system of spheres, that of  $\diamond$  determined by a family  $\{\mathbf{S}_w\}_{w \in [K]}$  of systems of spheres (i.e., one for each  $w \in [K]$ ).

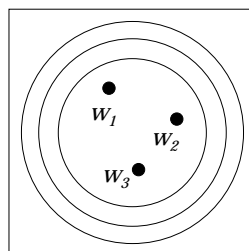
Main differences between possible worlds constructions wrt  $K$ .

<i>Revision</i>	<i>Update</i>
One SOS	Family of SOSs
Centred on $[K]$	Each centred on singleton $w \in [K]$

Collapse to same process when  $K$  is complete.



**Update**



**Revision**

## Revision and Update are not Fundamentally Different

At least when  $K$  is held fixed.

Revision function  $*$  can be constructed in exactly the same way as  $\diamond$ .

Similarity structure  $\mathbf{S}$  is *loyal* to a theory  $K$  of  $\mathcal{L}$  iff:

(SL1) For all  $w \in [K]$ ,  $[K] \in \mathbf{S}_w$ .

(SL2) For any two  $w, w' \in [K]$ ,  $\mathbf{S}_w - 2[K] = \mathbf{S}_{w'} - 2[K]$ .

$K * \varphi$  can be constructed from similarity structures loyal to  $K$  precisely like update for any  $\varphi \in \mathcal{L}$  and  $K \in \mathcal{M}_{\mathcal{L}}$ .

$$(SL^*) \quad K * \varphi = \begin{cases} th \left( \bigcup_{w \in [K]} f\mathbf{S}_w(\varphi) \right), & \text{if } [\varphi], [K] \neq \emptyset \\ \mathcal{L} & \text{otherwise} \end{cases}$$

### **Theorem 1:**

Let  $K$  be a consistent theory of  $\mathcal{L}$ . For every revision function  $*$  there exists a similarity structure  $\mathbf{S}$  that is loyal to  $K$ , such that (SL\*) is satisfied for all  $\varphi \in \mathcal{L}$ . Conversely, for any similarity structure  $\mathbf{S}$  that is loyal to  $K$ , there exists a revision function  $*$  that satisfies (SL\*) for all  $\varphi \in \mathcal{L}$ .

We now have an alternative constructive model for revision functions based on a family of systems of spheres centred on singletons rather than a single system of spheres centred on a family of worlds.

### **Theorem 2:**

Let  $K$  be a consistent theory of  $\mathcal{L}$ . For every revision function  $*$  there exists an update operator  $\diamond$  such that  $K * \varphi = K \diamond \varphi$ , for all  $\varphi \in \mathcal{L}$ .

## Are Revision Functions Special Kinds of Update?

Given a fixed consistent belief set  $K$  and revision function  $*$  Theorem 2 tells us that we can find an update operator  $\diamond$  behaving exactly like  $*$  (i.e.,  $K * \varphi = K \diamond \varphi$  for all  $\varphi \in \mathcal{L}$ ).

*However*, if we allow  $K$  to vary, will the *same* update operator  $\diamond$  keep on “simulating”  $*$ ?

Let  $H$  be a consistent theory different from  $K$ . Theorem 2 says that there exists a  $\diamond'$  such that  $H * \varphi = H \diamond' \varphi$  for all  $\varphi \in \mathcal{L}$ . But does  $\diamond = \diamond'$  hold?

Theorem 2 tells us nothing about this situation.

Is there a  $*$  for which there exists a single  $\diamond$  such that, for every consistent  $K$  and  $\varphi \in \mathcal{L}$ ,  $K * \varphi = K \diamond \varphi$ ?

Such a revision function is called *universal*.

Universal revision function is equivalent to existence of similarity structure  $\mathbf{S}$  loyal to every consistent theory  $K$  of  $\mathcal{L}$ .

Call this similarity structure  $\mathbf{S}$  *universally loyal*.

**Lemma 5.1:**

Let  $\mathbf{S}$  be a similarity structure defined over a language  $\mathcal{L}$ . If  $\mathbf{S}$  is universally loyal then  $\mathcal{L}$  is trivial.

**Theorem 3:**

Every universal revision function is defined over a trivial language.

### Fixed versus Variable Similarity

Fixed consistent theory  $K$ , revision operator  $*$  and update operator  $\diamond$  can be constructed from similarity structure in exactly the same way.

This ends when  $K$  allowed to vary.

$*$  requires  $\mathbf{S}$  associated with it to be loyal to current  $K$ . However,  $\mathbf{S}$  cannot be loyal to every  $K$  simultaneously. Therefore  $\mathbf{S}$  changes.

Similarity structure associated with  $\diamond$  fixed for all  $K$ .

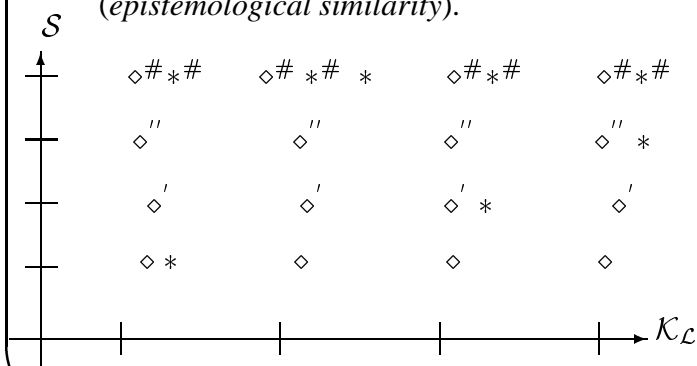
Semantic basis which differentiates between  $*$  and  $\diamond$ .

**Update:**  $\mathbf{S}$  remains fixed for all belief states.

Notion of similarity relates to intrinsic, external property of universe (*ontological similarity*).

**Revision:**  $\mathbf{S}$  changes with current belief state.

Similarity interpreted as agent's perception of (ontological) similarity (*epistemological similarity*).



### Discussion — Gärdenfors Triviality Theorem

Extend language by Ramsey test conditionals  $\varphi > \psi$

$$\varphi > \psi \text{ iff } \psi \in K * \varphi$$

Then there do not exist any  $*$ 's over a non-trivial language.

Culprits: *Preservation* ( $K^*4$ ) and *Monotonicity*:

(K\*M) If  $K \subseteq H$ , then  $K^*_\varphi \subseteq H^*_\varphi$ .

Loyal similarity structure implies Preservation.

Similarity structure of fixed (i.e., universal) similarity implies Monotonicity.

Any attempt to combine these will lead to triviality by Lemma 5.1.

Semantic reason for triviality?

Conjecture: Attempt to conflate loyal similarity structures with those of fixed similarity structures.

Need to show Preservation implies loyalty and Monotonicity implies universality.

Back to Lewis' original motivation for system of spheres modelling.

## Conclusions

Examined differences between two fundamental types of belief change — revision and update — from technical viewpoint.

For fixed consistent  $K$ , revising  $K$  same as updating  $K$ .

Difference between revision and update is not so much *how* to change a fixed  $K$  but rather way changes at different theories relate to one another.

This is due to a more fundamental difference in terms of ontological versus epistemological similarity adopted.

Semantic basis for Gärdenfors triviality result being pursued by considering basic rationality postulates common to both  $*$  and  $\diamond$  and examining semantics after addition of extra postulates.

## Areas of study in belief change

Belief bases and computational approaches

Coherence (how do beliefs ‘cohere’)

Contraction proposals (Lecture 3)

Iterated revision

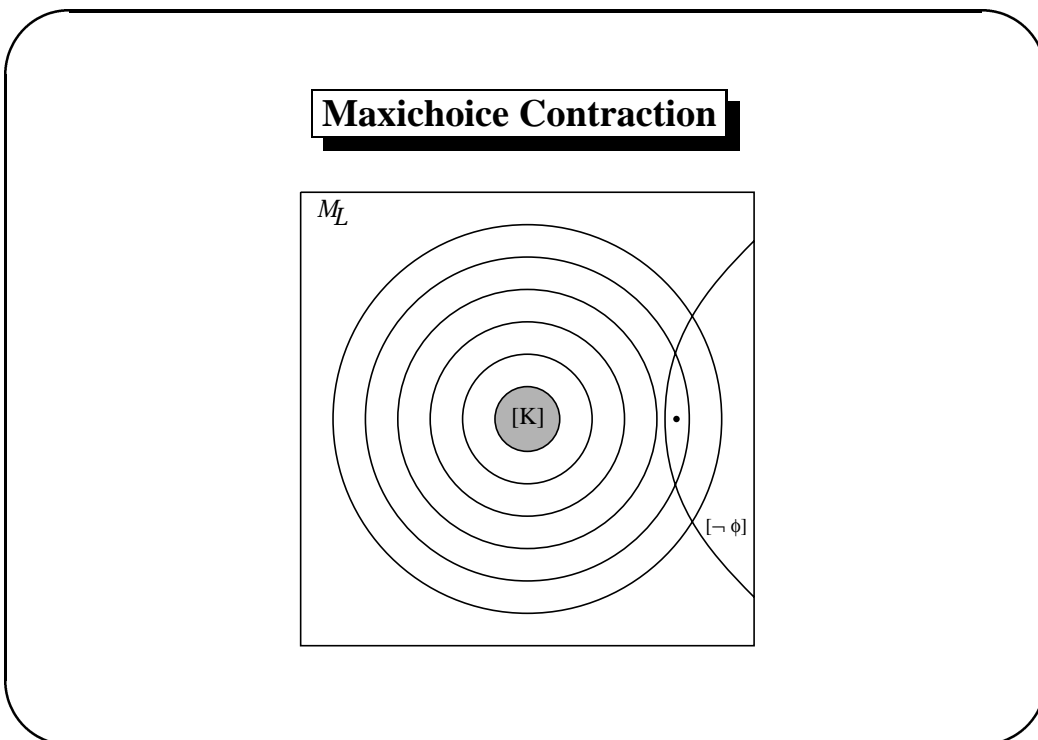
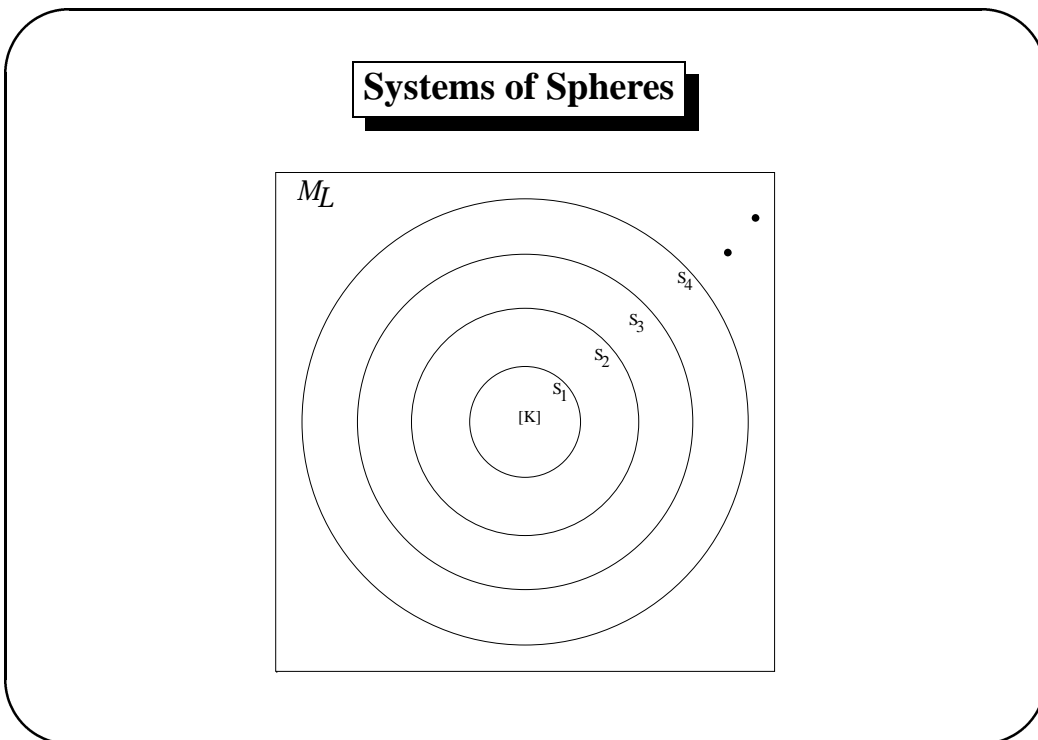
Relationships between belief change, nonmonotonic reasoning, conditionals, rational choice, ...

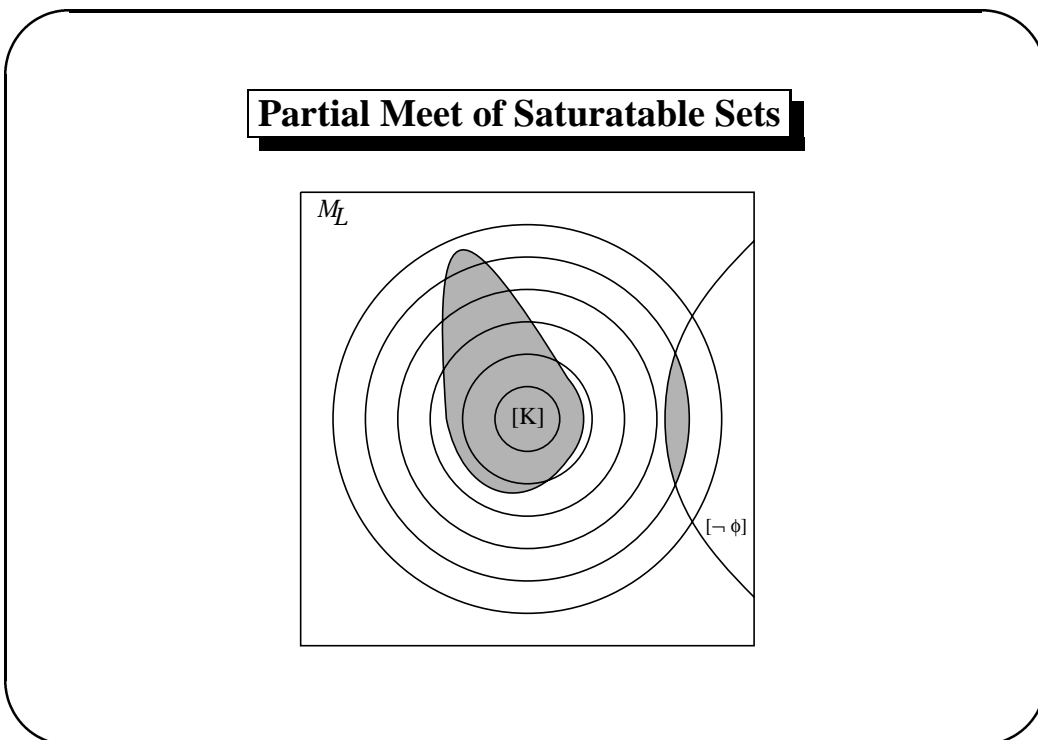
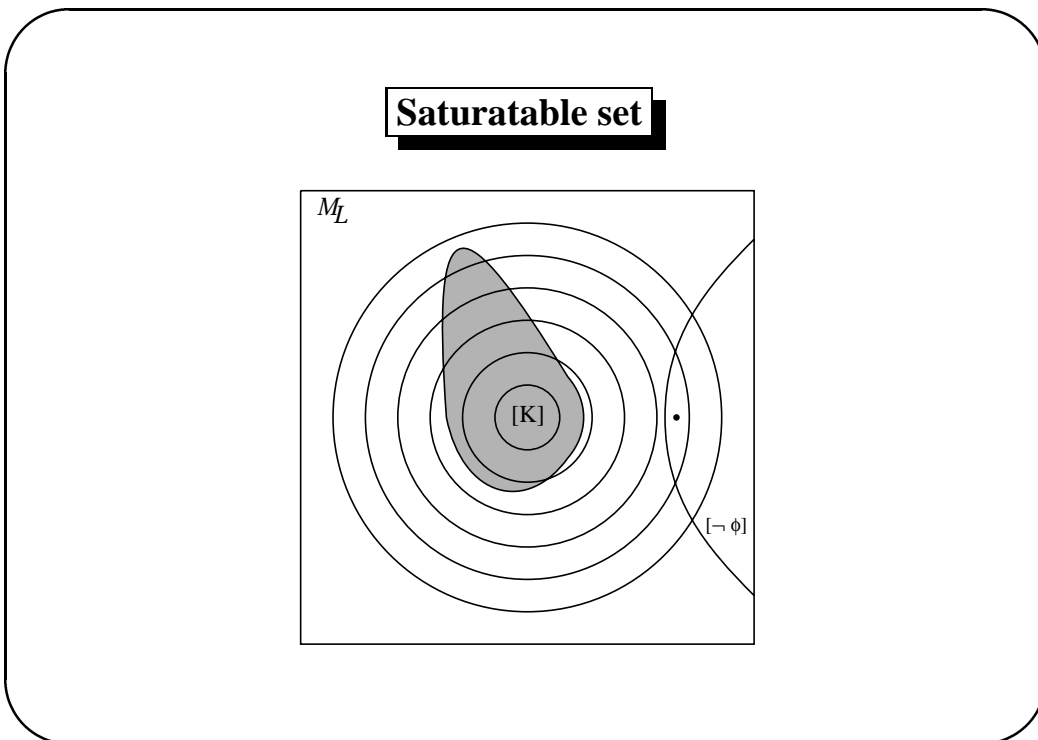
Non-prioritised belief change

Abductive belief change

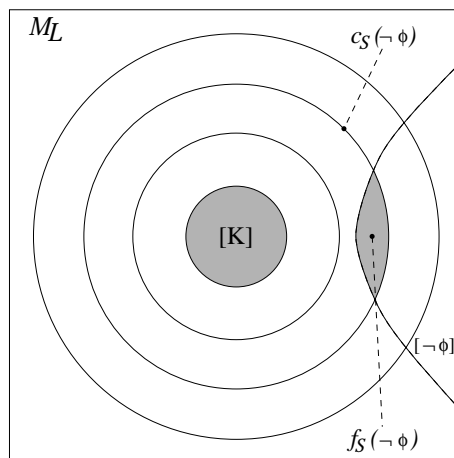
Reasoning about action and belief change

Dealing with uncertainty (Spohn, Bayesian networks, ...)



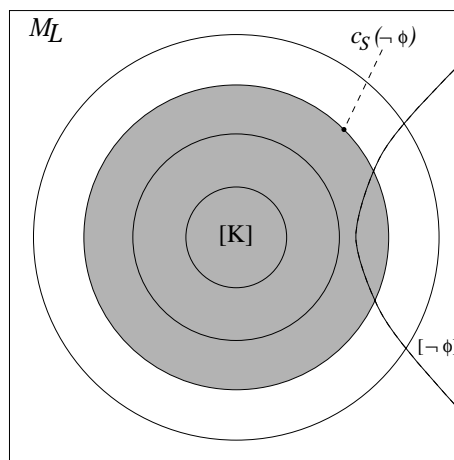


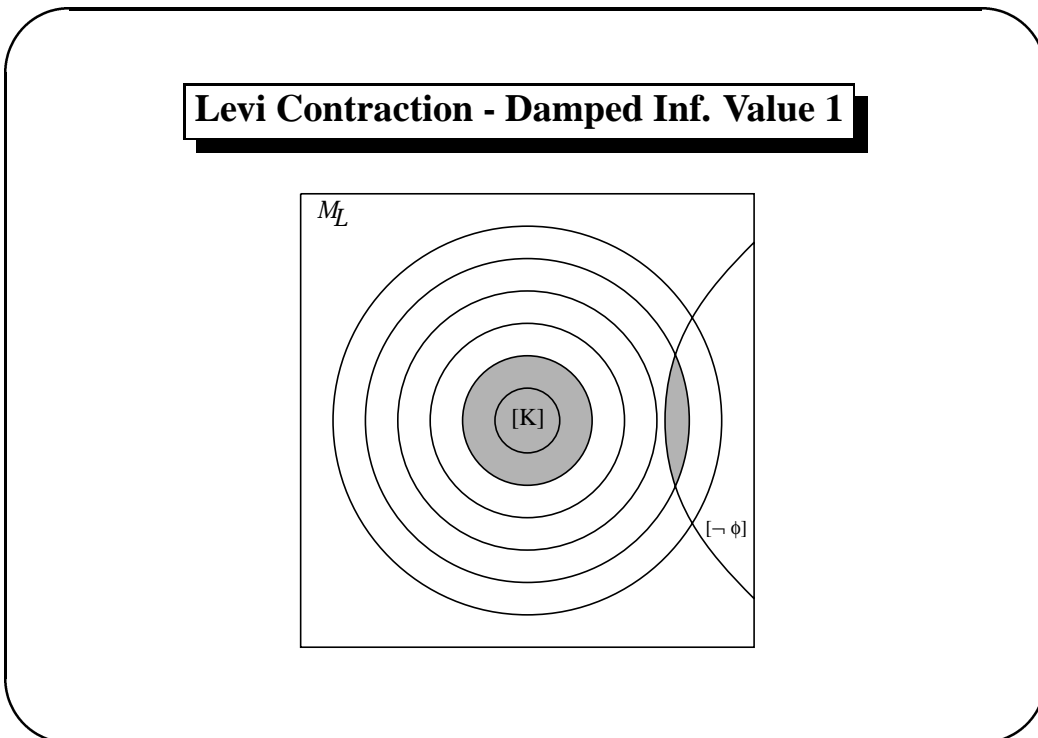
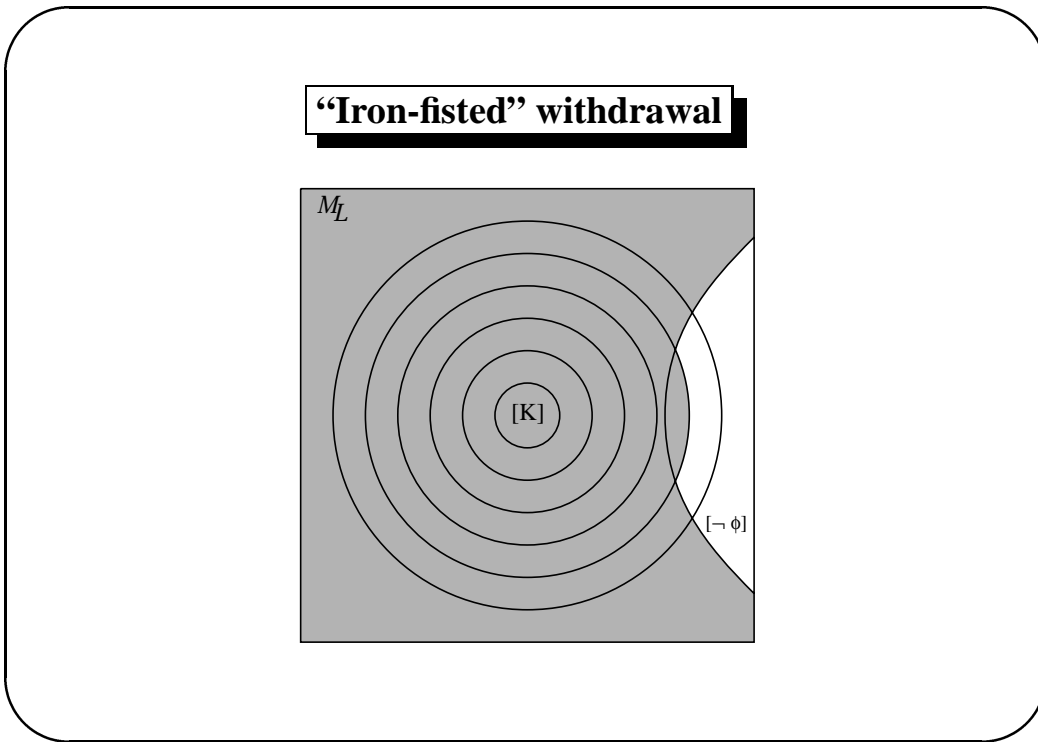
## AGM Partial Meet Contraction



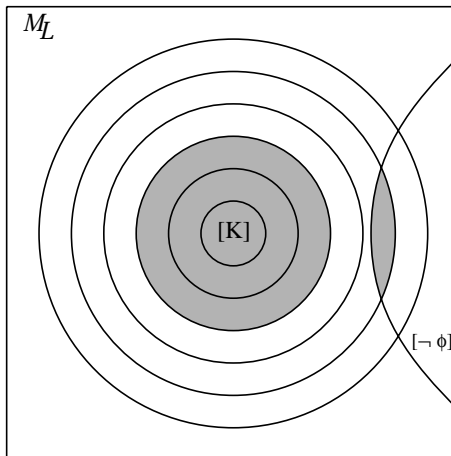
## Severe Withdrawal (Rott & Pagnucco 1996)

Also Levi 1996

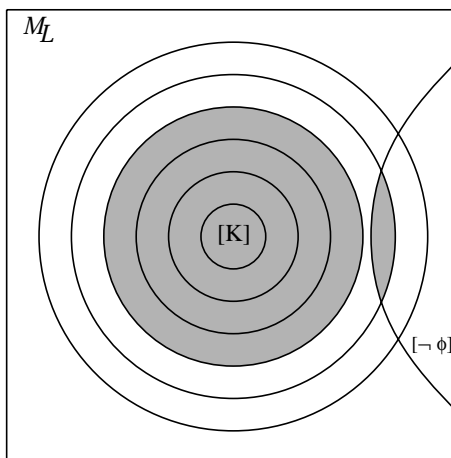




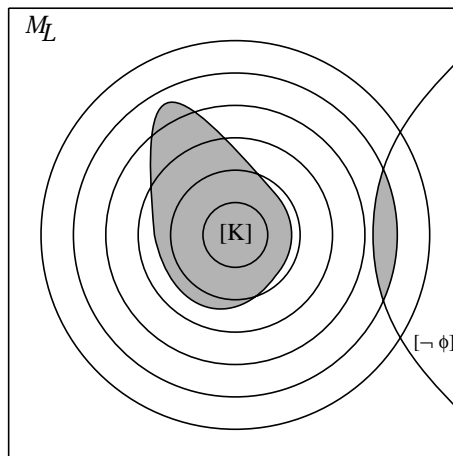
**Cantwell “fallback-based”**



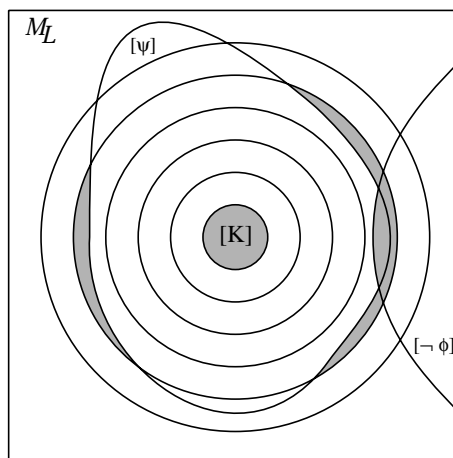
**Meyer *et al.* systematic withdrawal 1997**



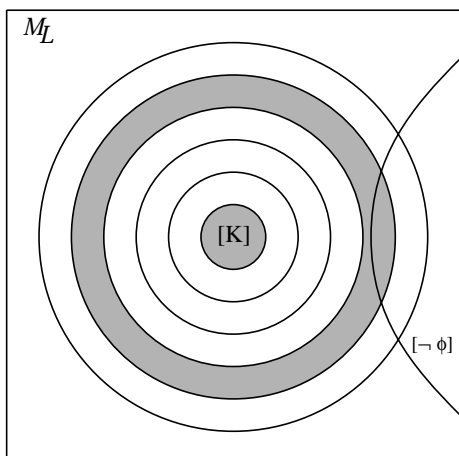
### Lindström and Rabinowicz (Interpolation) 1991



### Fermé and Rodriguez semi-contraction 1997



## Nayak (Personal Communication)



	$[K \dot{-} \phi]$	... within $[\neg \phi]$	... within $[\phi]$
1.	AGM (trans. rel.) partial meet	$[\neg \phi] \cap S_\phi$	$[K]$
2.	Severe withdrawal	$[\neg \phi] \cap S_\phi$	$[\phi] \cap S_\phi$
3.	AGM maxichoice	single $\neg \phi$ -world	$[K]$
4.	Saturatable set)	single $\neg \phi$ -world	some $X$ s.th. $[K] \subseteq X \subseteq [\phi]$
5.	Partial meet of saturatable sets	$[\neg \phi] \cap S_\phi$	some $X$ s.th. $[K] \subseteq X \subseteq [\phi]$
6.	Iron-fisted withdrawal	$[\neg \phi] \cap S_\phi$	$[\phi]$
7.	Levi - damped type 1	$[\neg \phi] \cap S_\phi$	$[\phi] \cap S_2$
8.	Cantwell fallback-based	$[\neg \phi] \cap S_\phi$	$[\phi] \cap S_i$ for some $i \in \{1, \dots, n\}$
9.	Systematic withdrawal	$[\neg \phi] \cap S_\phi$	$[\phi] \cap S_{\phi-1}$
10.	Lindström and Rabinowicz	$[\neg \phi] \cap S_\phi$	some $X$ s.th. $[K] \subseteq X \subseteq [\phi] \cap S_\phi$
11.	Semi-contraction	$[\neg \phi] \cap S_\phi$	some $X$ s.th. $[K] \subseteq X \subseteq [\phi] \cap S_\phi$
12.	Nayak [p.c.]	$[\neg \phi] \cap S_\phi$	$[\phi] \cap (S_\phi - S_{\phi-1})$