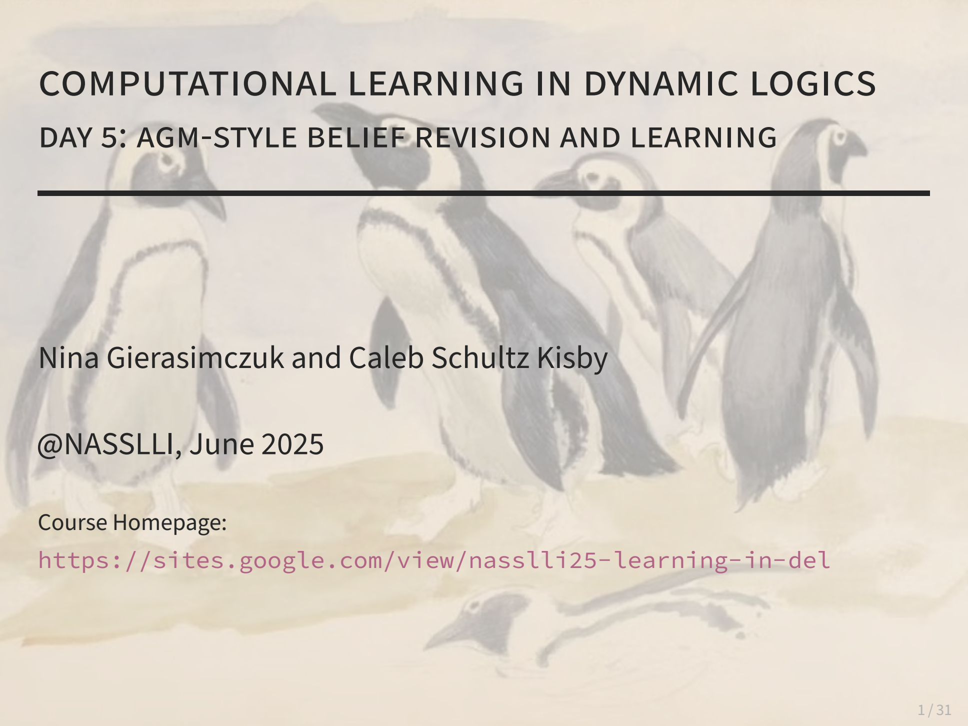


COMPUTATIONAL LEARNING IN DYNAMIC LOGICS

DAY 5: AGM-STYLE BELIEF REVISION AND LEARNING

Nina Gierasimczuk and Caleb Schultz Kisby

@NASSLLI, June 2025

Course Homepage:

https://sites.google.com/view/nasslli25-learning-in-del

PLAN FOR TODAY

1. Introduction to AGM-Style Belief Revision
2. Machine Learning and AGM Belief Revision

OUTLINE

1. Introduction to AGM-Style Belief Revision

2. Machine Learning and AGM Belief Revision

PLAN FOR TODAY

1. Introduction to AGM-Style Belief Revision

2. Machine Learning and AGM Belief Revision

THE PROBLEM OF BELIEF REVISION

> Belief revision is a topic of much interest in theoretical computer science and logic, and it forms a central problem in research into artificial intelligence. In simple terms: how do you update a database of knowledge in the light of new information? What if the new information is in conflict with something that was previously held to be true?
>
> Gärdenfors, Belief Revision

- CS: updating databases (Doyle 1979 and Fagin et al. 1983)
- Philosophy (epistemology):
  - scientific theory change and revisions of probability assignments;
  - belief change (Levi 1977, 1980, Harper 1977) and its rationality.

AGM BELIEF REVISION MODEL

- Names: Carlos **A**lchourrón, Peter **G**ärdenfors, and David **M**akinson.
- 1985 paper in the Journal of Symbolic Logic.
- Starting point of belief revision theory.

BELIEF REPRESENTATION IN AGM

We are talking about **beliefs** rather than **knowledge**.

Here, the difference is that beliefs are changeable and can be false.

BELIEF REPRESENTATION IN AGM

We are talking about **beliefs** rather than **knowledge**.

Here, the difference is that beliefs are changeable and can be false.

- belief := sentence

BELIEF REPRESENTATION IN AGM

We are talking about **beliefs** rather than **knowledge**.

Here, the difference is that beliefs are changeable and can be false.

- belief := sentence
- belief := sentence in some formal language

BELIEF REPRESENTATION IN AGM

We are talking about **beliefs** rather than **knowledge**.

Here, the difference is that beliefs are changeable and can be false.

- belief := sentence
- belief := sentence in some formal language
- beliefs of an agent := a set of such sentences

BELIEF REPRESENTATION IN AGM

We are talking about **beliefs** rather than **knowledge**.

Here, the difference is that beliefs are changeable and can be false.

- belief := sentence
- belief := sentence in some formal language
- beliefs of an agent := a set of such sentences

Language of Beliefs in AGM

Beliefs are expressed in propositional logic:

- propositions $p, q, r, \ldots$
- connectives: negation ($\neg$), conjunction ($\wedge$), disjunction ($\vee$), implication ($\rightarrow$), and biconditional ($\leftrightarrow$).

BELIEF SETS

Example (What are the consequences of my beliefs?)

BELIEF SETS

Example (What are the consequences of my beliefs?)

1. John is a bachelor. John is handsome.

BELIEF SETS

Example (What are the consequences of my beliefs?)

1. John is a bachelor. John is handsome. **John is a handsome bachelor.**

BELIEF SETS

Example (What are the consequences of my beliefs?)

1. John is a bachelor. John is handsome. **John is a handsome bachelor.**
2. If we charge high fees for university, only the rich enroll. We charge high fees for university.

BELIEF SETS

Example (What are the consequences of my beliefs?)

1. John is a bachelor. John is handsome. **John is a handsome bachelor.**
2. If we charge high fees for university, only the rich enroll. We charge high fees for university. **Only the rich enroll.**

BELIEF SETS

Example (What are the consequences of my beliefs?)

1. John is a bachelor. John is handsome. **John is a handsome bachelor.**
2. If we charge high fees for university, only the rich enroll. We charge high fees for university. **Only the rich enroll.**
3. The barber is male. The barber shaves only those men in town who do not shave themselves.

BELIEF SETS

Example (What are the consequences of my beliefs?)

1. John is a bachelor. John is handsome. **John is a handsome bachelor.**
2. If we charge high fees for university, only the rich enroll. We charge high fees for university. **Only the rich enroll.**
3. The barber is male. The barber shaves only those men in town who do not shave themselves. **The barber is female.**

BELIEF SETS

Example (What are the consequences of my beliefs?)

1. John is a bachelor. John is handsome. **John is a handsome bachelor.**
2. If we charge high fees for university, only the rich enroll. We charge high fees for university. **Only the rich enroll.**
3. The barber is male. The barber shaves only those men in town who do not shave themselves. **The barber is female.**

Belief set is a set of formulas that is **deductively closed**.

BELIEF SETS

Example (What are the consequences of my beliefs?)

1. John is a bachelor. John is handsome. **John is a handsome bachelor.**
2. If we charge high fees for university, only the rich enroll. We charge high fees for university. **Only the rich enroll.**
3. The barber is male. The barber shaves only those men in town who do not shave themselves. **The barber is female.**

Belief set is a set of formulas that is **deductively closed**.

Definition

For any set B of sentences, $Cn(B)$ is the set of **logical consequences** of B.

If φ can be derived from B by classical propositional logic, then $\varphi \in Cn(B)$.

THREE WAYS OF TAKING IN NEW INFORMATION

What can I do to my belief set?

1. **Revision**: $B * \varphi$; φ is added and other things are removed, so that the resulting new belief set B' is consistent.
2. **Contraction**: $B \div \varphi$; φ is removed from B giving a new belief set B'.
3. **Expansion**: $B + \varphi$; φ is added to B giving a new belief set B'.

AGM$^{\div}$ RATIONALITY POSTULATES OF CONTRACTION

1. **Closure**: $B \div \varphi = Cn(B \div \varphi)$ (the outcome is logically closed)
2. **Success**: If $\varphi \notin Cn(\varnothing)$, then $\varphi \notin Cn(B \div \varphi)$

   the outcome does not contain φ
3. **Inclusion**: $B \div \varphi \subseteq B$ (the outcome is a subset of the original set)
4. **Vacuity**: If $\varphi \notin Cn(B)$, then $B \div \varphi = B$

   if the incoming sentence is not in the original set then there is no effect
5. **Extensionality**: If $\varphi \leftrightarrow \psi \in Cn(\varnothing)$, then $B \div \varphi = B \div \psi$.

   the outcomes of contracting with equivalent sentences are the same
6. **Recovery**: $B \subseteq (B \div \varphi) + \varphi$.

   contraction leads to the loss of as few previous beliefs as possible
7. **Conjunctive inclusion**: If $\varphi \notin B \div (\varphi \wedge \psi)$, then $B \div (\varphi \wedge \psi) \subseteq B \div \varphi$.
8. **Conjunctive overlap**: $(B \div \varphi) \cap (B \div \psi) \subseteq B \div (\varphi \wedge \psi)$.

AGM* RATIONALITY POSTULATES OF REVISION

1. **Closure**: $B * \varphi = Cn(B * \varphi)$
2. **Success**: $\varphi \in B * \varphi$
3. **Inclusion**: $B * \varphi \subseteq B + \varphi$
4. **Vacuity**: If $\neg\varphi \notin B$, then $B * \varphi = B + \varphi$
5. **Consistency**: $B * \varphi$ is consistent if φ is consistent.
6. **Extensionality**: If $(\varphi \leftrightarrow \psi) \in Cn(\varnothing)$, then $B * \varphi = B * \psi$.
7. **Superexpansion**: $B * (\varphi \wedge \psi) \subseteq (B * \varphi) + \psi$
8. **Subexpansion**: If $\neg\psi \notin B * \varphi$, then $(B * \varphi) + \psi \subseteq B * (\varphi \wedge \psi)$.

HARPER IDENTITIY

One formal way to combine those two is to use:

Harper identity (HR)

$$B \div \varphi := (B * \neg\varphi) \cap K.$$

Given an AGM $*$ function, the $\div$ obtained by HR is an AGM-contraction.

REVISION AND CONTRACTION ON PLAUSIBILITY ORDERS

p, q	$p, \bar{q}$	$\bar{p}, q$	$\bar{p}, \bar{q}$
		z	
			w
	y		
x			

↓ more plausible

Plausibility order over valuations

REVISION AND CONTRACTION ON PLAUSIBILITY ORDERS

p, q	$p, \bar{q}$	$\bar{p}, q$	$\bar{p}, \bar{q}$
		z	
			w
	y		
x			

↓ more plausible

B is determined by the most plausible world(s)

REVISION AND CONTRACTION ON PLAUSIBILITY ORDERS

p, q	$p, \bar{q}$	$\bar{p}, q$	$\bar{p}, \bar{q}$
		z	
			w
	y		
x			

↓ more plausible

$B * \neg p$ is determined by min world(s) with $\neg p$

REVISION AND CONTRACTION ON PLAUSIBILITY ORDERS

p, q	$p, \bar{q}$	$\bar{p}, q$	$\bar{p}, \bar{q}$
		z	
			w
	y		
x			

↓ more plausible

$B \div p$ is the union of the previous two

FORMALLY

p, q	$p, \bar{q}$	$\bar{p}, q$	$\bar{p}, \bar{q}$
		z	
			w
	y		
x			

↓ more plausible

Definition

Let P be a set of propositions (e.g. above, $P = \{p, q\}$). A **plausibility order** is a total preorder $\leq$ over the possible truth assignments W on P. A total preorder on X is a binary relation that is:

- transitive: for all $x, y, z \in X$, if $x \leq y$ and $y \leq z$, then $x \leq z$;
- complete: for all $x, y \in X$, $x \leq y$ or $y \leq x$.

FORMALLY

p, q	$p, \bar{q}$	$\bar{p}, q$	$\bar{p}, \bar{q}$
		z	
			w
	y		
x			

↓ more plausible

Let B be a belief set, φ a formula, and let $[\varphi] := \{x \in W \mid \varphi \text{ is true in } x\}$.

- $\varphi \in B$ iff $min_{\leq}(W) \subseteq [\varphi]$;
- $\varphi \in B * \psi$ iff $min_{\leq}([\psi]) \subseteq [\varphi]$;
- $\varphi \in B \div \psi$ iff $min_{\leq}([\neg\psi]) \cup min_{\leq}(W) \subseteq [\varphi]$

DALAL'S REVISION: HAMMING DISTANCE IN ACTION

The **difference** between two $r, r' \in W$, $Diff(r, r')$ is the set of propositional variables that have different truth values in the two worlds.

DALAL'S REVISION: HAMMING DISTANCE IN ACTION

The **difference** between two $r, r' \in W$, $Diff(r, r')$ is the set of propositional variables that have different truth values in the two worlds.

Given a belief set K and and a world r, the **distance** between K and r, $D(K, r)$, is the cardinality-minimum difference between r and the K-worlds.

DALAL'S REVISION: HAMMING DISTANCE IN ACTION

The **difference** between two $r, r' \in W$, $Diff(r, r')$ is the set of propositional variables that have different truth values in the two worlds.

Given a belief set K and and a world r, the **distance** between K and r, $D(K, r)$, is the cardinality-minimum difference between r and the K-worlds.

Definition (Dalal's revision)

$[K *_D \varphi] = min([\varphi], \sqsubseteq_K)$, where $\sqsubseteq_K$ is a total preorder on K, s.t.:

$$r \sqsubseteq_K r' \text{ iff } D(K, r) \leq D(K, r').$$

DALAL'S REVISION: HAMMING DISTANCE IN ACTION

The **difference** between two $r, r' \in W$, $Diff(r, r')$ is the set of propositional variables that have different truth values in the two worlds.

Given a belief set K and and a world r, the **distance** between K and r, $D(K, r)$, is the cardinality-minimum difference between r and the K-worlds.

Definition (Dalal's revision)
$[K *_D \varphi] = min([\varphi], \sqsubseteq_K)$, where $\sqsubseteq_K$ is a total preorder on K, s.t.:

$$r \sqsubseteq_K r' \text{ iff } D(K, r) \leq D(K, r').$$

The **distance between the two belief sets** K_1 and K_2 is defined as:

$$Dist(K_1, K_2) = |([K_1] \setminus [K_2]) \cup ([K_2] \setminus [K1])|.$$

DALAL IS AGM

Dalal's revision satisfies the AGM postulates.

OUTLINE

1 Introduction to AGM-Style Belief Revision

2 Machine Learning and AGM Belief Revision

ARTIFICIAL NEURAL NETWORKS

An ANN is a directed acyclic graph $G = (V, E)$, with V neurons and E connections between neurons. (Here we only consider feed-forward nets.)

V comes as a set of distinct, ordered subsets (*layers*) $V_0, \ldots, V_L$, where:

- V_0 is the *input layer*: $X_1, \ldots, X_n$.
- V_L is the *output layer*: $y_1, \ldots, y_m$.
- $V_1, V_2, \ldots, V_{L-1}$ are the *hidden layers*.

A *layer* V_l, for $l = 0, \ldots, L$, is a set of neurons such that:

- For $l = 0$, V_0 receives the external inputs $X_1, \ldots, X_n$.
- For $l = L$, V_L produces outputs $y_1, \ldots, y_m$.
- For $l = 1, 2, \ldots, L-1$, V_l receives only from V_{l-1}, and sends only to V_{l+1}.

$E = \{(u, v) \mid u \in V_{l-1}, v \in V_l, \text{ for } l = 1, \ldots, L\}$.

COMPUTATION OF AN ACTIVATION

Each neuron $v \in V \setminus V_0$ computes a *weighted sum* of its inputs, adds a bias term, and then applies a non-linear *activation function* σ. Specifically, for a neuron $v \in V_l$ in layer l (where $l = 1, 2, \ldots, L$), the output z_v is given by

$$z_v = \sigma\left(\sum_{u \in V_{l-1}} w_{uv} \cdot x_u + b_v\right).$$

Here, w_{uv} is the weight of the edge from neuron u in layer V_{l-1} to neuron v in layer V_l, x_u is the output of neuron u, and b_v is the bias of neuron v.

COMPUTATION OF AN ACTIVATION

Each neuron $v \in V \setminus V_0$ computes a *weighted sum* of its inputs, adds a bias term, and then applies a non-linear *activation function* σ. Specifically, for a neuron $v \in V_l$ in layer l (where $l = 1, 2, \ldots, L$), the output z_v is given by

$$z_v = \sigma\left(\sum_{u \in V_{l-1}} w_{uv} \cdot x_u + b_v\right).$$

Here, w_{uv} is the weight of the edge from neuron u in layer V_{l-1} to neuron v in layer V_l, x_u is the output of neuron u, and b_v is the bias of neuron v.

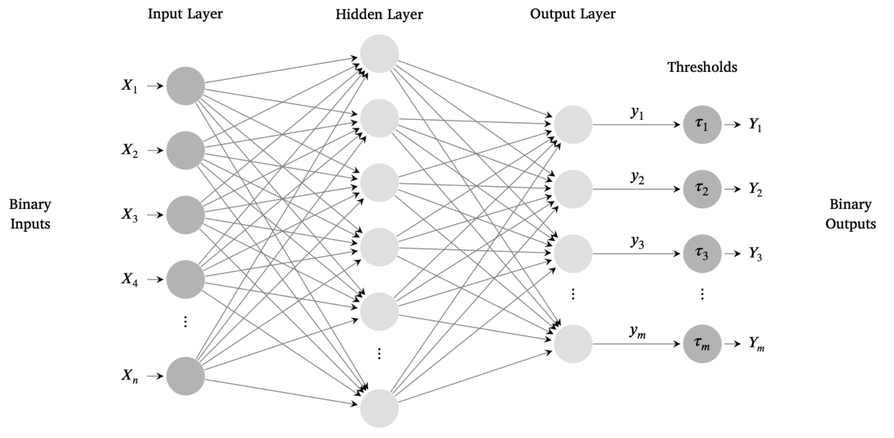
Activation functions: sigmoid, Rectified Linear Unit (ReLU), and softmax.

BINARY ANN

We only consider feed-forward ANNs, with inputs $X_1, \ldots, X_n \in \{0, 1\}$.
Given a threshold $\tau_i \in [0, 1]$, output y_i becomes a binary Y_i:

$$Y_i = \begin{cases} 1 & \text{if } y_i \geq \tau_i \\ 0 & \text{if } y_i < \tau_i \end{cases}$$

BINARY ANN



Input Layer
Hidden Layer
Output Layer
Thresholds
Binary Inputs
X_1
X_2
X_3
X_4
X_n
y_1
y_2
y_3
y_m
τ_1
τ_2
τ_3
τ_m
Y_1
Y_2
Y_3
Y_m
Binary Outputs

BINARY ANN AS A BELIEF SET

Binary ANN computes the boolean function:

$$Y = f(X_1, \ldots X_n)$$

which can be represented as a propositional formula ψ.

Then, for the belief set $K = Cn(\psi)$, we have: $[K] = [\psi]$.

Similarly, ANN with multiple outputs $Y_1, \ldots, Y_m$ can be represented as $S = \langle K_1, \ldots, K_m \rangle$ of belief sets (a belief state/epistemic space).

TRAINING AN ANN

Training an ANN involves iteratively tuning its parameters (i.e., the w_{uv}'s and b_v corresponding to every neuron) in order to minimize the disparity between the desired/actual outputs and the predictions of the network, thereby improving its **predictive accuracy**.

Forward propagation: input propagates according to the weights.

Backpropagation: computing the gradient of the **loss function** $\mathcal{L}$ (measure of prediction error) relative to each parameter (weight and bias).

ASSUMPTIONS ABOUT THE LOSS FUNCTION

- **Smoothness:** In each iteration the value of loss function decreases.
- **Monotonicity:** Loss function is monotonically related to the sum of absolute errors across all predictions (of all samples).

RESULTS

Consider a single-output binary ANN whose training process satisfies smoothness and monotonicity. Let Y be the output.

Theorem

Let K_n be the Boolean function of Y corresponding to the labels and $K_1, \ldots, K_n$ be belief sets, s.t. for any $i \in \{1, \ldots, n-1\}$, K_i and K_{i+1} are Boolean functions of Y just before and after the i-th update. Then for any $i, j \in \{1, \ldots, n\}$, such that $i < j$: $Dist(K_i, K_n) \geq Dist(K_j, K_n)$.

Theorem

Let K_1 and K_2 represent the boolean functions of Y before and after an arbitrary update of ANN's parameters. Then there are AGM-style $*$ and $\div$, and formulas φ, φ', s.t.

$$K_2 = (K_1 * \varphi_1) \div \varphi_2$$

WHAT HAVE WE LEARNED IN THIS COURSE?

We have introduced three paradigms of learning:

- **Model-Theoretic Learning:**

  Belief Revision Theory & Dynamic Epistemic Logic (DEL)

- **Function Learning:** Machine Learning, Neural Network Learning
- **Set Learning:** Computational Learning Theory, Learning in the Limit

WHAT HAVE WE LEARNED IN THIS COURSE?

We have introduced three paradigms of learning:

- **Model-Theoretic Learning:**

  Belief Revision Theory & Dynamic Epistemic Logic (DEL)

- **Function Learning:** Machine Learning, Neural Network Learning
- **Set Learning:** Computational Learning Theory, Learning in the Limit

And showed that these perspectives are compatible with each other!

WHAT HAVE WE LEARNED IN THIS COURSE?

We have introduced three paradigms of learning:

- **Model-Theoretic Learning:**

  Belief Revision Theory & Dynamic Epistemic Logic (DEL)

- **Function Learning:** Machine Learning, Neural Network Learning
- **Set Learning:** Computational Learning Theory, Learning in the Limit

And showed that these perspectives are compatible with each other!

- Belief revision can be modelled as DEL updates

WHAT HAVE WE LEARNED IN THIS COURSE?

We have introduced three paradigms of learning:

- **Model-Theoretic Learning:**

  Belief Revision Theory & Dynamic Epistemic Logic (DEL)
- **Function Learning:** Machine Learning, Neural Network Learning
- **Set Learning:** Computational Learning Theory, Learning in the Limit

And showed that these perspectives are compatible with each other!

- Belief revision can be modelled as DEL updates
- Neural network updates can be modelled as DEL updates (DEL can give us **complete characterizations** of learning)

WHAT HAVE WE LEARNED IN THIS COURSE?

We have introduced three paradigms of learning:

- **Model-Theoretic Learning:**

  Belief Revision Theory & Dynamic Epistemic Logic (DEL)
- **Function Learning:** Machine Learning, Neural Network Learning
- **Set Learning:** Computational Learning Theory, Learning in the Limit

And showed that these perspectives are compatible with each other!

- Belief revision can be modelled as DEL updates
- Neural network updates can be modelled as DEL updates (DEL can give us **complete characterizations** of learning)
- We can study belief revision operators' ability to learn in the limit

WHAT HAVE WE LEARNED IN THIS COURSE?

We have introduced three paradigms of learning:

- **Model-Theoretic Learning:**

  Belief Revision Theory & Dynamic Epistemic Logic (DEL)

- **Function Learning:** Machine Learning, Neural Network Learning
- **Set Learning:** Computational Learning Theory, Learning in the Limit

And showed that these perspectives are compatible with each other!

- Belief revision can be modelled as DEL updates
- Neural network updates can be modelled as DEL updates (DEL can give us **complete characterizations** of learning)
- We can study belief revision operators' ability to learn in the limit
- Backpropagation in a neural net is AGM-compatible

Thank you for attending our NASSLLI'25 class on Computational Learning in Dynamic Logics!

Please get in touch with us at

- Nina: nigi@dtu.dk
- Caleb: cckisby@gmail.com

If you have further questions, comments, or feedback. We are happy to help with the course exercises and chat about open problems in the area!