Who’s the Expert? On Multi-source Belief Change

Joe Singleton and Richard Booth

singletonj1@cardiff.ac.uk

Overview

- **Problem:** suppose we receive conflicting reports from multiple non-expert information sources
  - What should we believe?
  - Who should we trust?
- We develop a logical framework to reason about expertise of multiple sources
- A belief change problem is expressed in this framework
  - Extends AGM revision
  - Allows us to explore how belief and trust interact
- We put forward several postulates and families of change operators

Motivating example

Motivating example

- Patients Alice and Bob are checked for condition p
- Blood test confirms Alice does have p (📋, Alice, ✅)
    - Assuming the test is reliable, we can revise beliefs with **AGM revision**

Motivating example

- Patients Alice and Bob are checked for condition p
- Blood test confirms Alice does have p (📋, Alice, ✅)
  - Assuming the test is reliable, we can revise beliefs with **AGM revision**
- Dr. X: "Bob has p" (🧑‍⚕️, Bob, ✅)
- Dr. Y: "Bob does not have p" (🧑‍⚕️, Bob, ❌)
  - Need to use **non-prioritised** revision: drop the *Success* postulate

Motivating example

- Patients Alice and Bob are checked for condition p
- Blood test confirms Alice does have p (📋, Alice, ✅)
  - Assuming the test is reliable, we can revise beliefs with AGM revision
- Dr. X: “Bob has p” (👨‍⚕️, Bob, ✅)
- Dr. Y: “Bob does not have p” (👨🏽‍⚕️, Bob, ❌)
  - Need to use non-prioritised revision: drop the *Success* postulate
- It emerges that earlier, X said Alice did *not* have p (👨‍⚕️, Alice, ❌)
  - There is now reason to distrust X on diagnosing p
  - We need to revise beliefs about X’s expertise *and* Bob’s condition

Motivating example

- Patients Alice and Bob are checked for condition p
- Blood test confirms Alice does have p (📋, Alice, ✅)
  - Assuming the test is reliable, we can revise beliefs with AGM revision
- Dr. X: "Bob has p" (🧑‍⚕️, Bob, ✅)
- Dr. Y: "Bob does not have p" (🧑🏽‍⚕️, Bob, ❌)
  - Need to use non-prioritised revision: drop the *Success* postulate
- It emerges that earlier, X said Alice did *not* have p (🧑‍⚕️, Alice, ❌)
  - There is now reason to distrust X on diagnosing p
  - We need to revise beliefs about X's expertise *and* Bob's condition
- **Our questions:**
  - How should the revision be performed?
  - Can trust and belief aspects be unified?

Logical framework for expertise

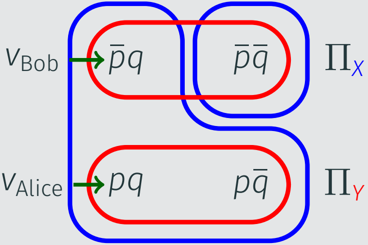
Syntax

- $\mathcal{S}$: finite set of information sources (e.g. test results, X, Y, ...)
- A distinguished source $* \in \mathcal{S}$ is completely reliable (e.g. test results)
- $\mathcal{P}$: finite set of propositional variables (e.g. p, ...)
- $\mathcal{L}$: extension of the propositional language $\mathcal{L}_0$:
  - $E_i(\varphi)$: source i **has expertise** on φ
    - In any situation, i is able to determine the correct value of φ
  - $S_i(\varphi)$: the formula φ is **sound** for i to report
    - φ is true *up to the expertise of i*
    - i does not *know* φ is false

Semantics

- $\mathcal{C}$: finite set of cases (e.g. Alice, Bob)
- $\mathcal{V}$: propositional valuations over $\mathcal{P}$
- We model expertise of sources via partitions
    - States in the same cell are indistinguishable
- A world is a pair $W = \langle \{v_c\}_{c \in \mathcal{C}}, \{\Pi_i\}_{i \in \mathcal{S}} \rangle$
    - Each v_c is a valuation
    - Each Π_i is a partition of $\mathcal{V}$, s.t. $\Pi_* = \{\{v\} \mid v \in \mathcal{V}\}$

Example



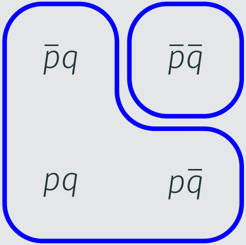
Semantics (cont'd)

- We evaluate formulas with respect to a world W and a case c
- i **has expertise** on φ if i can always determine the correct value of φ:

$$W, c \models E_i(\varphi) \iff (u \in \mathrm{mod}_0\, \varphi \implies \Pi_i[u] \subseteq \mathrm{mod}_0\, \varphi)$$

  - φ states always distinguishable from $\neg\varphi$ states

Example



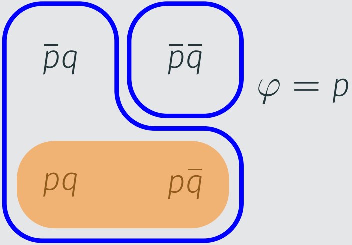
Semantics (cont'd)

- We evaluate formulas with respect to a world W and a case c
- i **has expertise** on φ if i can always determine the correct value of φ:

$$W, c \models E_i(\varphi) \iff (u \in \mathrm{mod}_0\, \varphi \implies \Pi_i[u] \subseteq \mathrm{mod}_0\, \varphi)$$

  - φ states always distinguishable from $\neg\varphi$ states

Example



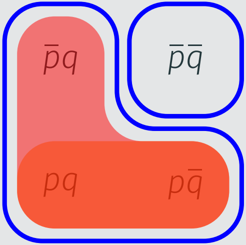
Semantics (cont'd)

- We evaluate formulas with respect to a world W and a case c
- i **has expertise** on φ if i can always determine the correct value of φ:

$$W, c \models E_i(\varphi) \iff (u \in \mathrm{mod}_0\, \varphi \implies \Pi_i[u] \subseteq \mathrm{mod}_0\, \varphi)$$

  - φ states always distinguishable from $\neg\varphi$ states

Example



$\varphi = p$

$W, c \not\models E_X(p)$ ✗

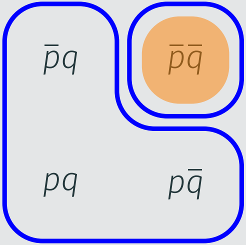
Semantics (cont'd)

- We evaluate formulas with respect to a world W and a case c
- i **has expertise** on φ if i can always determine the correct value of φ:

$$W, c \models E_i(\varphi) \iff (u \in \mathrm{mod}_0\, \varphi \implies \Pi_i[u] \subseteq \mathrm{mod}_0\, \varphi)$$

  - φ states always distinguishable from $\neg\varphi$ states

Example



$\psi = \neg p \wedge \neg q$

$W, c \models E_X(\psi)$ ✅

- We evaluate formulas with respect to a world W and a case c
- i **has expertise** on φ if i can always determine the correct value of φ:

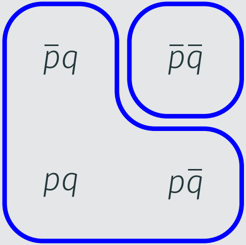
$$W, c \models E_i(\varphi) \iff (u \in \mathrm{mod}_0\,\varphi \implies \Pi_i[u] \subseteq \mathrm{mod}_0\,\varphi)$$

  - φ states always distinguishable from $\neg\varphi$ states
- φ is **sound** for i if φ is true *up to lack of expertise* of i

$$W, c \models S_i(\varphi) \iff \Pi_i[v_c] \cap \mathrm{mod}_0\,\varphi \neq \emptyset$$

  - true state indistinguishable from some φ state

Example



Semantics (cont'd)

- We evaluate formulas with respect to a world W and a case c
- i **has expertise** on φ if i can always determine the correct value of φ:

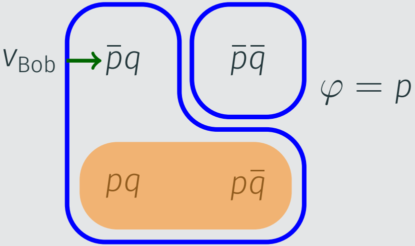
$$W, c \models E_i(\varphi) \iff (u \in \mathrm{mod}_0\, \varphi \implies \Pi_i[u] \subseteq \mathrm{mod}_0\, \varphi)$$

  - φ states always distinguishable from $\neg\varphi$ states
- φ is **sound** for i if φ is true *up to lack of expertise* of i

$$W, c \models S_i(\varphi) \iff \Pi_i[v_c] \cap \mathrm{mod}_0\, \varphi \neq \emptyset$$

  - true state indistinguishable from some φ state

Example



Semantics (cont'd)

- We evaluate formulas with respect to a world W and a case c
- i **has expertise** on φ if i can always determine the correct value of φ:

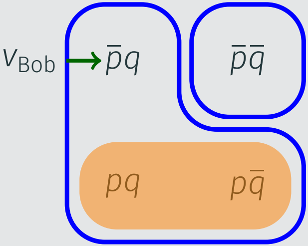
$$W, c \models E_i(\varphi) \iff (u \in \mathrm{mod}_0\, \varphi \implies \Pi_i[u] \subseteq \mathrm{mod}_0\, \varphi)$$

  - φ states always distinguishable from $\neg\varphi$ states
- φ is **sound** for i if φ is true *up to lack of expertise* of i

$$W, c \models S_i(\varphi) \iff \Pi_i[v_c] \cap \mathrm{mod}_0\, \varphi \neq \emptyset$$

  - true state indistinguishable from some φ state

Example



$\varphi = p$

$W, \text{Bob} \models \neg p \wedge S_X(p)$ ✓

Some validities

- Symmetric expertise:

$$E_i(\varphi) \leftrightarrow E_i(\neg\varphi)$$

- Closure under conjunctions:

$$E_i(\varphi) \wedge E_i(\psi) \rightarrow E_i(\varphi \wedge \psi)$$

- Expertise and soundness interaction:

$$E_i(\varphi) \wedge S_i(\varphi) \rightarrow \varphi$$

- Reliable source properties:

$$E_*(\varphi)$$

$$S_*(\varphi) \leftrightarrow \varphi$$

The belief change problem

The belief change problem

- **Input:** a finite sequence of reports σ, where each report is a triple $\langle i, c, \varphi \rangle$, with $\varphi \not\equiv \bot$
- **Output:** a pair $\langle B^\sigma, K^\sigma \rangle$, where
  - $B^\sigma = \{B_c^\sigma\}_{c \in \mathcal{C}}$ is a collection of belief sets $B_c^\sigma \subseteq \mathcal{L}$
  - $K^\sigma = \{K_c^\sigma\}_{c \in \mathcal{C}}$ is a collection of knowledge sets $K_c^\sigma \subseteq \mathcal{L}$
- Output is expressed in the extended language: describes both beliefs about the state of affairs in each case, *and* the expertise of the sources

The belief change problem

- **Input:** a finite sequence of reports σ, where each report is a triple $\langle i, c, \varphi \rangle$, with $\varphi \not\equiv \bot$
- **Output:** a pair $\langle B^\sigma, K^\sigma \rangle$, where
  - $B^\sigma = \{B_c^\sigma\}_{c \in \mathcal{C}}$ is a collection of **belief sets** $B_c^\sigma \subseteq \mathcal{L}$
  - $K^\sigma = \{K_c^\sigma\}_{c \in \mathcal{C}}$ is a collection of **knowledge sets** $K_c^\sigma \subseteq \mathcal{L}$
- Output is expressed in the extended language: describes both beliefs about the state of affairs in each case, *and* the expertise of the sources

- Belief sets represent the **defeasible** part of the output: beliefs can be retracted; e.g.
  - $E_X(p) \in B_c^{\langle *, Alice, p \rangle \cdot \langle X, Bob, p \rangle}$
  - $E_X(p) \notin B_c^{\langle *, Alice, p \rangle \cdot \langle X, Bob, p \rangle \cdot \langle Y, Bob, \neg p \rangle}$
  - $\neg E_X(p) \in B_c^{\langle *, Alice, p \rangle \cdot \langle X, Bob, p \rangle \cdot \langle Y, Bob, \neg p \rangle \cdot \langle X, Alice, \neg p \rangle}$

Basic postulates

- We introduce some basic postulates, including:
  - *Closure*: $B^\sigma = \mathrm{Cn}(B^\sigma)$ and $K^\sigma = \mathrm{Cn}(K^\sigma)$
  - *Equivalence*: If $\varphi \equiv \psi$ then $B^{\sigma \cdot \langle i,c,\varphi \rangle} = B^{\sigma \cdot \langle i,c,\psi \rangle}$ and $K^{\sigma \cdot \langle i,c,\varphi \rangle} = K^{\sigma \cdot \langle i,c,\psi \rangle}$
  - *Containment*: $K_c^\sigma \subseteq B_c^\sigma$
  - *K-conjunction*: $K^{\sigma \cdot \rho} = \mathrm{Cn}(K^\sigma \sqcup K^\rho)$
    - Knowledge grows monotonically
  - *Soundness*: If $\langle i, c, \varphi \rangle \in \sigma$, then $S_i(\varphi) \in K_c^\sigma$
    - All reports are sound. Sources are honest

Basic postulates

- We introduce some basic postulates, including:
  - *Closure*: $B^\sigma = \mathrm{Cn}(B^\sigma)$ and $K^\sigma = \mathrm{Cn}(K^\sigma)$
  - *Equivalence*: If $\varphi \equiv \psi$ then $B^{\sigma \cdot \langle i,c,\varphi \rangle} = B^{\sigma \cdot \langle i,c,\psi \rangle}$ and $K^{\sigma \cdot \langle i,c,\varphi \rangle} = K^{\sigma \cdot \langle i,c,\psi \rangle}$
  - *Containment*: $K^\sigma_c \subseteq B^\sigma_c$
  - *K-conjunction*: $K^{\sigma \cdot \rho} = \mathrm{Cn}(K^\sigma \sqcup K^\rho)$
    - Knowledge grows monotonically
  - *Soundness*: If $\langle i, c, \varphi \rangle \in \sigma$, then $S_i(\varphi) \in K^\sigma_c$
    - All reports are sound. Sources are **honest**
- Immediate consequences:
  - *Success* holds for the reliable source $*$:
    - If $\langle *, c, \varphi \rangle \in \sigma$, then $\varphi \in K^\sigma_c$

Basic postulates

- We introduce some basic postulates, including:
  - *Closure*: $B^{\sigma} = \mathrm{Cn}(B^{\sigma})$ and $K^{\sigma} = \mathrm{Cn}(K^{\sigma})$
  - *Equivalence*: If $\varphi \equiv \psi$ then $B^{\sigma \cdot \langle i,c,\varphi \rangle} = B^{\sigma \cdot \langle i,c,\psi \rangle}$ and $K^{\sigma \cdot \langle i,c,\varphi \rangle} = K^{\sigma \cdot \langle i,c,\psi \rangle}$
  - *Containment*: $K_c^{\sigma} \subseteq B_c^{\sigma}$
  - *K-conjunction*: $K^{\sigma \cdot \rho} = \mathrm{Cn}(K^{\sigma} \sqcup K^{\rho})$
    - Knowledge grows monotonically
  - *Soundness*: If $\langle i, c, \varphi \rangle \in \sigma$, then $S_i(\varphi) \in K_c^{\sigma}$
    - All reports are sound. Sources are **honest**

- Immediate consequences:
  - *Success* holds for the reliable source $*$:
    - If $\langle *, c, \varphi \rangle \in \sigma$, then $\varphi \in K_c^{\sigma}$
  - We distrust sources making reports known to be false:
    - If $\varphi \in K_c^{\sigma}$, then $\neg E_i(\varphi) \in K_c^{\sigma \cdot \langle i,c,\neg\varphi \rangle}$

Basic postulates

- We introduce some basic postulates, including:
  - *Closure*: $B^\sigma = \mathrm{Cn}(B^\sigma)$ and $K^\sigma = \mathrm{Cn}(K^\sigma)$
  - *Equivalence*: If $\varphi \equiv \psi$ then $B^{\sigma \cdot \langle i,c,\varphi \rangle} = B^{\sigma \cdot \langle i,c,\psi \rangle}$ and $K^{\sigma \cdot \langle i,c,\varphi \rangle} = K^{\sigma \cdot \langle i,c,\psi \rangle}$
  - *Containment*: $K_c^\sigma \subseteq B_c^\sigma$
  - *K-conjunction*: $K^{\sigma \cdot \rho} = \mathrm{Cn}(K^\sigma \sqcup K^\rho)$
    - Knowledge grows monotonically
  - *Soundness*: If $\langle i, c, \varphi \rangle \in \sigma$, then $S_i(\varphi) \in K_c^\sigma$
    - All reports are sound. Sources are **honest**

- Immediate consequences:
  - *Success* holds for the reliable source $*$:
    - If $\langle *, c, \varphi \rangle \in \sigma$, then $\varphi \in K_c^\sigma$
  - We distrust sources making reports known to be false:
    - If $\varphi \in K_c^\sigma$, then $\neg E_i(\varphi) \in K_c^{\sigma \cdot \langle i,c,\neg\varphi \rangle}$
  - We believe reports from trusted sources:
    - If $\langle i, c, \varphi \rangle \in \sigma$ and $E_i(\varphi) \in B_c^\sigma$ then $\varphi \in B_c^\sigma$

Belief change operators

Conditioning operators

- **Idea:** select a set of possible worlds $\mathcal{X}_\sigma$, and use a plausibility ordering to choose the most plausible worlds
- These sets induce knowledge and beliefs, respectively
- Given a mapping $\sigma \mapsto \mathcal{X}_\sigma$ and total preorder $\leq$ on worlds, the conditioning operator is defined by

$$K^\sigma = \mathrm{Th}(\mathcal{X}_\sigma)$$
$$B^\sigma = \mathrm{Th}(\min_\leq \mathcal{X}_\sigma)$$

Example

Set $\mathcal{X}_\sigma = \{W \mid \forall \langle i, c, \varphi \rangle \in \sigma, W, c \models S_i(\varphi)\}$, and $W \leq W'$ iff $r(W) \leq r(W')$, where

$$r(W) = -\sum_{i \in \mathcal{S}} |\{p \in \mathcal{P} \mid W, c_0 \models E_i(p)\}|$$

That is: we only know soundness statements, and aim to trust sources on as many propositional variables as possible.

Conditioning operators: example

- $\sigma_1 = \langle *, Alice, p \rangle \cdot \langle X, Bob, p \rangle \cdot \langle Y, Bob, \neg p \rangle$
- We have:

$$B^{\sigma_1}_{Alice} \cap \mathcal{L}_0 = \mathrm{Cn}_0(p)$$
$$B^{\sigma_1}_{Bob} \cap \mathcal{L}_0 = \mathrm{Cn}_0(\emptyset)$$
$$E_i(p), \neg E_i(p) \notin B^{\sigma_1}_c, \text{ for both } i \in \{X, Y\}$$
$$(p \to E_X(p)) \wedge (\neg p \to E_Y(p)) \in B^{\sigma_1}_{Bob}$$

Conditioning operators: example

- $\sigma_1 = \langle *, Alice, p\rangle \cdot \langle X, Bob, p\rangle \cdot \langle Y, Bob, \neg p\rangle$
- We have:

$$B^{\sigma_1}_{Alice} \cap \mathcal{L}_0 = \mathrm{Cn}_0(p)$$
$$B^{\sigma_1}_{Bob} \cap \mathcal{L}_0 = \mathrm{Cn}_0(\emptyset)$$
$$E_i(p), \neg E_i(p) \notin B^{\sigma_1}_c, \text{ for both } i \in \{X, Y\}$$
$$(p \rightarrow E_X(p)) \wedge (\neg p \rightarrow E_Y(p)) \in B^{\sigma_1}_{Bob}$$

- $\sigma_2 = \sigma_1 \cdot \langle X, Alice, \neg p\rangle$

$$B^{\sigma_2}_{Bob} \cap \mathcal{L}_0 = \mathrm{Cn}_0(\neg p)$$
$$\neg E_X(p) \wedge E_Y(p) \in B^{\sigma_2}_c$$

- **Note:** a new report from Dr. X on Alice affects beliefs about Bob and the expertise of Drs. X and Y!

A characterisation of conditioning operators

- Conditioning is characterised by the following postulates:
  - *Duplicate-removal*: If $\rho_1 = \sigma \cdot \langle i, c, \varphi \rangle$ and $\rho_2 = \rho_1 \cdot \langle i, c, \varphi \rangle$ then $B^{\rho_1} = B^{\rho_2}$ and $K^{\rho_1} = K^{\rho_2}$
  - *Conditional-consistency*: If K^σ is consistent then so is B^σ
  - *Inclusion-vacuity*: $B^{\sigma \cdot \rho} \sqsubseteq \mathrm{Cn}(B^\sigma \sqcup K^\rho)$, with equality if $B^\sigma \sqcup K^\rho$ is consistent
  - *Acyc*: If $\sigma_0, \ldots, \sigma_n$ are such that $K^{\sigma_j} \sqcup B^{\sigma_{j+1}}$ is consistent for all $0 \leq j < n$ and $K^{\sigma_n} \sqcup B^{\sigma_0}$ is consistent, then $K^{\sigma_0} \sqcup B^{\sigma_n}$ is consistent

Theorem

Suppose an operator satisfies the basic postulates. Then it is an elementary conditioning operator if and only if it satisfies Duplicate-removal, Conditional-consistency, Inclusion-vacuity and Acyc.

(elementary $\iff \mathcal{X}_\sigma = \mathrm{mod}(G)$ for some G, and similarly for $\min_{\leq} \mathcal{X}_\sigma$)

One-step revision

Extending AGM revision

- The reliable source $*$ allows us to extend AGM revision
- **Notation:** $[B_c^\sigma] = B_c^\sigma \cap \mathcal{L}_0$
  - *AGM-$*$*: For any σ and $c \in \mathcal{C}$ there is an AGM operator $\star$ for $[B_c^\sigma]$ such that $\left[B_c^{\sigma \cdot \langle *, c, \varphi \rangle}\right] = [B_c^\sigma] \star \varphi$ whenever $\neg\varphi \notin K_c^\sigma$
- That is, a new report from $*$ for case c is just AGM revision on (propositional part of) the c-th belief set
- Satisfied by our example operator ✅

Weakenings of Success

- We have *Success* (and other AGM postulates) for the reliable source $*$
- What about unreliable sources? We offer two weaker formulations:
  - *Cond-success*: If $E_i(\varphi) \in B_c^\sigma$ and $\neg\varphi \notin B_c^\sigma$, then $\varphi \in B_c^{\sigma \cdot \langle i,c,\varphi \rangle}$
  - *Strong-cond-success*: If $\neg(\varphi \wedge E_i(\varphi)) \notin B_c^\sigma$, then $\varphi \in B_c^{\sigma \cdot \langle i,c,\varphi \rangle}$

Weakenings of Success

- We have *Success* (and other AGM postulates) for the reliable source $*$
- What about unreliable sources? We offer two weaker formulations:
  - *Cond-success*: If $E_i(\varphi) \in B_c^{\sigma}$ and $\neg\varphi \notin B_c^{\sigma}$, then $\varphi \in B_c^{\sigma \cdot \langle i,c,\varphi \rangle}$
  - *Strong-cond-success*: If $\neg(\varphi \wedge E_i(\varphi)) \notin B_c^{\sigma}$, then $\varphi \in B_c^{\sigma \cdot \langle i,c,\varphi \rangle}$
- Our example satisfies *Cond-success* ✅
- ...but **not** *Strong-cond-success* ❌
- We have an impossibility result: basic conditioning operators with some additional reasonable properties *cannot* satisfy *Strong-cond-success*

In the paper...

- We overcome this impossibility result by introducing the new class of score-based operators:
- **Idea:** assign a plausibility score to each pair $(W, \langle i, c, \varphi \rangle)$
- We give an example score-based operator satisfying all the postulates

Conclusion

Summary and future work

- Summary:
  - Introduced a belief change problem to reason about reports from non-expert information sources
  - Explored the connections between trust and belief
  - Representation result for the class of conditioning operators
- Future work:
  - Computational issues
  - Graded or probabilistic expertise