DYNAMIC PARTIAL AWARENESS

JOSEPH Y. HALPERN
Cornell University,
Computer Science Department

EVAN PIERMONT
Royal Holloway, University of London,
Department of Economics

- Awareness is a pervasive form of ignorance:
- An agent is unaware of an object/event/statement/etc if it is *not on her radar screen*

Existing models have difficulties with:

- Awareness of Unawareness:

  "I know I am unaware of current research in immunological oncology."

- Partial Awareness:

  "I am aware of bitcoin but do not fully understand it."

- Growing Awareness:

  "I just discovered quantum computing, and it makes me re-evaluate cyber security."

Why do we care?

- More realistic/nuanced representation of ignorance.
- Agents who reason about thier own limitations, may take actions to temper exposure to unawareness
  - Motivate learning, preference for the well understood
- Has important implications in economic markets:
  - Incomplete contracting, over-confidence, context-dependent preferences, &c.

This talk:

1. Develop a formal model of partial awareness.
   - Modal logic extending Halpern Rego (2009, 2013).

2. Considers what happens when agents become more aware
   - How do the agent's beliefs change
   - Becoming aware is in-and-of-itself informative

The language has three building blocks:

1. objects
2. properties: unary predicates
3. concepts: Boolean combination of properties

In each state (possible world), each agent is aware of a subset of objects, properties, and concepts.

The logic also has two modalities:

- $A_i\varphi$: agent i is **aware** of the formula φ
- $K_i\varphi$: agent i **explicitly knows** formula φ.
    - Read: φ is true at all possible worlds and $A_i\varphi$

Suppose a quantum computer (Q) is defined as a computer (C) that possesses an additional "quantum property" QP.

$$Q := C \wedge QP$$

1. A *partially aware* agent is aware of Q but unaware of the specific Boolean combination of properties that characterizes it.
2. A *fully unaware* agent is unaware of even the concept of a quantum computer
3. A *fully aware* agent is unaware is aware of both the concept of a quantum computer and also what it means to be one, i.e., the properties C and QP.

We capture introspection via quantification (over both objects and properties).

- $K_i\Big(\exists P \forall x\big(Q(x) \Leftrightarrow C(x) \wedge P(x) \wedge \neg A_i P(x)\big)\Big)$
    - Although the agent is unaware of QP, she knows something about Q (quantum computers): she knows that they're computers that satisfy some extra property P.
- $K_i\big(\forall P \forall x((Q(x) \Leftrightarrow C(x) \wedge P(x)) \Rightarrow (P(x) \Rightarrow R(x))\big)$
    - She can even reason about this property.

So what is the limitation imposed by (partial) unawareness?

- “Shadow predicates” cannot be expressed.
- Limits information exchange.

Syntax

The syntax of our logic has the following building blocks:

- A countable set $\mathscr{O}$ of constant symbols, representing objects. Following Levesque (1990), $\mathscr{O}$ consists of *standard names* $d_1, d_2, \ldots$
- A countably infinite set $\mathcal{V}^{\mathscr{O}}$ of object variables, which range over objects.
- A countable set $\mathscr{P}$ of unary predicate symbols.
- A countably infinite set $\mathcal{V}^{\mathscr{P}}$ of predicate variables.
- A countable set $\mathscr{C}$ of concept symbols.

If $d \in \mathscr{O}$, $x \in \mathcal{V}^{\mathscr{O}}$, $P \in \mathscr{P}$, $Y \in \mathcal{V}^{\mathscr{P}}$, and $C \in \mathscr{C}$, then $P(d)$, $P(x)$, $Y(d)$, $Y(x)$, $C(d)$, and $C(x)$ are *atomic formulas*.

Our language $\mathcal{L}(\mathscr{O}, \mathscr{P}, \mathscr{C})$ is the closure under

- conjunction and negation
- quantification over objects and over unary predicates,
  - if φ is a formula, $x \in \mathcal{V}^{\mathscr{O}}$, and $Y \in \mathcal{V}^{\mathscr{P}}$, then so are $\forall x\varphi$ and $\forall Y\varphi$.
- the modal operators $A_1, \ldots, A_n$ and $K_1, \ldots, K_n$,
  - representing awareness and (explicit) knowledge.
  - if φ is a formula, then so are $A_i\varphi$ and $K_i\varphi$.

Semantics

- Ω of possible *states*.
- each state ω is associated with a language. Formally, there is a function Φ on states such that $\Phi(\omega) = (\mathscr{O}_\omega, \mathscr{P}_\omega, \mathscr{C}_\omega)$
- we take the domain D to be the standard names in $\mathscr{O}$.
- a binary relation $\mathcal{K}_i$ on states; states i considers possible.
- an awareness set $\mathcal{A}_i(\omega) \subseteq \Phi(\omega)$, the symbols i is aware of.

An interpretation I: for each state ω, we have a function I_ω:

- taking $\mathscr{O}$ to elements of the domain D, standard names are mapped to themselves, so that $I_w(d_i) = d_i$.
- $\mathscr{P}$ to subsets of D,
- $\mathscr{C}$ to $\mathcal{L}^{bc}(\Phi(\omega))$, Boolean combinations of properties (i.e., predicates).

The truth of a sentences at a state ω in

$$M = (\Omega, D, \Phi, \mathcal{A}_1 \ldots, \mathcal{A}_n, \mathcal{K}_1, \ldots, \mathcal{K}_n, I)$$

are defined as usual .

Of interest:

- $(M, \omega) \models C(d)$ iff $(M, \omega) \models C^I_\omega(d)$,
- $(M, \omega) \models \forall Y \varphi$ iff $(M, \omega) \models \varphi[Y/\psi]$, where $\psi \in \mathcal{L}^{bc}$
- $(M, \omega) \models A_i \varphi$ iff $\varphi \in \mathcal{L}(\mathcal{A}_i(\omega))$,
- $(M, \omega) \models K_i \varphi$ iff $(M, \omega) \models A_i \varphi$ and $(M, \omega') \models \varphi$ for all $\omega' \in \mathcal{K}_i(\omega)$.

We assume agents know what they are aware of:

- if $(\omega, \omega') \in \mathcal{K}_i$, then $\mathcal{A}_i(\omega) = \mathcal{A}_i(\omega')$

and $\mathcal{K}_i$ is an equivalence relation, and thus partitions the states in Ω.

Why different languages at different states? Otherwise:

$$\big(\neg K_i \neg \forall P (A_i P(d))\big) \Rightarrow K_i \forall P (A_i P(d))\big)$$

Because different languages at different states:

- φ is true at ω only if it is expressible at ω.
- Non-standard notion of validity:
  - φ is *valid* in M if $(M, \omega) \models \varphi$ for all $\omega \in \Omega$ such that $\varphi \in \mathcal{L}(\Phi(\omega))$.

How do beliefs change as awareness changes?

- Becoming aware of φ is informative
- Ex. playing a game with an irrational opponent
  - Become aware of a new rule (without knowing what the rule specifies)
  - More likely the opponent is rational

The state space Ω is God's state-space. The 'objective' view of the world.

But the agent's view is limited by her own language.

An event $E \subseteq \Omega$ is ω-**conceivable** of there is a sentence φ such that

1. The agent is aware of φ at ω.
2. φ is true exactly on E.

Under basic regularity conditions

$$\Sigma_\omega = \{E \subseteq \Omega \mid \text{ is } \omega\text{-conceivable}\}$$

is a σ-algebra on Ω.

A modeler can obtain a subjective probability $\pi \in \Delta(\Omega, \Sigma_\omega)$

- Use bets on the truth of φ (usual decision theory).
- Only consider sentences the agent is already aware of.
- The support of π is $\mathcal{K}(\omega)$.

Now what happens when the agent becomes more aware, e.g., of a new property P.

- Her language gets richer
  - Implies: She might be able to differentiate new states
- She learns that she used to be unaware of whatever she discovered
  - Implies: she might condition her beliefs

Formally: becoming aware of φ changes the model from M to $M^{[\varphi]}$:

- The ‘physical’ properties of the states do not change: $\Omega = \Omega^{[\varphi]}$, and $P(d)$ is true at ω if and only if it is true at $\omega^{[\varphi]}$.
- The agent becomes aware of all the symbols in φ

$$\mathcal{A}^{[\varphi]}(\omega) = \mathcal{A}(\omega) \cup \mathsf{SYM}(\varphi)$$

- The agent learns she was unaware of φ:

$$\mathcal{K}^{[\varphi]}(\omega^{[\varphi]}) = \mathcal{K}(\omega) \cap \{\omega \mid \text{the agent could have been unaware of } \varphi\}$$

Say π_0 is ex-ante probability and π_1 is ex-post (the agent becomes aware of φ):

- $\Sigma_\omega \subseteq \Sigma_\omega^{[\varphi]}$ (can differentiate new states)
- $\text{supp}(\pi_1) \subseteq \text{supp}(\pi_0)$ ('condition' her beliefs).

Bayes' rule states $\pi_1(E) = \frac{\pi_0(E \cap \text{supp}(\pi_1))}{\pi_0(\text{supp}(\pi_1))}$, but what if $\text{supp}(\pi_1) \notin \Sigma_\omega$?

Theorem.

If for all $E, E' \in \Sigma_\omega$ with $E, E' \subseteq \text{supp}(\pi_1)$, we have

$$\frac{\pi_0(E)}{\pi_0(E')} = \frac{\pi_1(E)}{\pi_1(E')}$$

and $\pi_0(E) \leq \pi_1(E)$ then it is *as if* π_1 is a conditional distribution of π_0.

- There exists a $\pi^* \in \Delta(\Omega, \Sigma_\omega^{[\varphi]})$ such that π^* is an extension of π_0 and

$$\pi_1(E) = \frac{\pi^*(E \cap \text{supp}(\pi_1))}{\pi^*(\text{supp}(\pi_1))}$$

Hence the model:

- Allows for growing awareness to change beliefs about previously describable events
  - Has testable predictions
- This is in juxtaposition to Karni and Viero (2015) and Dominiak and Tserenjigmid (2019)