Unawareness and Impossible States

Mikaël Cozic

DEC (Ecole Normale Supérieure, Paris)

Workshop Decisions, Games and Logic - LSE
20/07/2007

cognitive idealizations

- two main families of doxastic models:
  (i) **epistemic logic** for *full beliefs*
  (ii) **probability** for *partial beliefs*
- *both* doxastic models suffer from two main cognitive idealizations:
  (i) **logical omniscience**: closure under logical consequence, substituability of logically equivalent formulas, etc.
  (ii) **full awareness**: full understanding of the state space
- to weaken these cognitive idealizations is intrinsically valuable *and* crucial for the development of bounded rationality in decision theory and game theory

a plea for impossible states

- lots of papers on each of the two cognitive idealizations:
  - Hintikka 1975, Fagin & Halpern 1988, Wansing 1991, Stalnaker 1991 & 1999, FHMV 1995 on logical omniscience
  - Fagin & Halpern 1988, Modica & Rustichini (MR) 1994, Dekel, Lipman & Rustichini 1998, Modica & Rustichini 1999, Halpern 2001, Heifetz, Meier & Schipper (HMS) 2006, HMS 2007a, HMS 2007b on unwareness
- **broad aim:** defend the impossible states (or worlds) approach as a unifying way to weaken cognitive idealizations = **a plea for impossible states**

impossible states and LO

(i) what is a “unifying” solution ?
(1) a solution to both logical omniscience and unawareness
(2) a solution for models of *full and partial* beliefs

impossible states and LO

(i) what is a "unifying" solution ?
(1) a solution to both logical omniscience and unawareness
(2) a solution for models of *full and partial* beliefs

(ii) what are impossible states? states of the state space where *a priori* "anything goes"

- it is well-known that impossible states allow one to introduce any form of logical ignorance in epistemic logic

impossible states and LO

(i) what is a "unifying" solution ?
(1) a solution to both logical omniscience and unawareness
(2) a solution for models of *full and partial* beliefs

(ii) what are impossible states? states of the state space where *a priori* "anything goes"

- it is well-known that impossible states allow one to introduce any form of logical ignorance in epistemic logic
- basic idea: suppose that Pierre believes that ϕ but not that ψ where ψ is a logical consequence of ϕ. Modeling: Pierre considers as epistemically possible a state s^* where ϕ is true but ψ is false. By assumption, s^* is an impossible state.
- Cozic (2007) shows how to extend the impossible state solution to probabilistic logic (logical version of *type*

what is (un)awareness ?

- *Modica & Rustichini 1999:*
  • "ignorance about the state space"
  • "some of the facts that determine which state of nature occurs are not present in the subject's mind"
  • "the agent does not know, does not know that she does not know, does not know that she does not know that she does not know, and so on..."
- *Heifetz, Meier & Schipper 2007b:*
  • "Unawareness refers to lack of conception rather than to lack of information."

example

- Pierre plans to rent a house for the holiday; three main factors from the modeler point of view:
- p: the house is no more than 1 km far from the sea
- q: the house is no more than 1 km far from a bar
- r: the house is no more than 1 km far from an airport

example

- Pierre plans to rent a house for the holiday; three main factors from the modeler point of view:
  - p: the house is no more than 1 km far from the sea
  - q: the house is no more than 1 km far from a bar
  - r: the house is no more than 1 km far from an airport
- "simple", factual, ignorance of r: Pierre doesn't know whether there is an airport no more than 1 km far from the house - there are both r-states and $\neg r$-states which are epistemically accessible

example

- Pierre plans to rent a house for the holiday; three main factors from the modeler point of view:
  - p: the house is no more than 1 km far from the sea
  - q: the house is no more than 1 km far from a bar
  - r: the house is no more than 1 km far from an airport
- "simple", factual, ignorance of r: Pierre doesn't know whether there is an airport no more than 1 km far from the house - there are both r-states and $\neg r$-states which are epistemically accessible
- unawareness: Pierre doesn't ask to himself: "is there an airport no more than 1 km far from the house?"
  [See states in *small worlds* in Savage (1954/72)]

example, *cont.*

- the possibility that r is not simply excluded: it is out of Pierre's state space
- modeler's point of view:

pqr	$p\neg qr$	$pq\neg r$	$p\neg q\neg r$
$\neg pqr$	$\neg p\neg qr$	$\neg pq\neg r$	$\neg p\neg q\neg r$

- Pierre's point of view:

pq	$p\neg q$
$\neg pq$	$\neg p\neg q$

[See Savage: "...a smaller world is derived from a larger by neglecting some distinctions between states"]

properties of (un)awareness

- some intuitive properties of (un)awareness:

$A\phi \leftrightarrow A\neg\phi$	(Symmetry)
$A(\phi \wedge \psi) \leftrightarrow A\phi \wedge A\psi$	
$A\phi \leftrightarrow AA\phi$	(Self-Reflection)
$U\phi \rightarrow UU\phi$	(U-introspection)
$U\phi \rightarrow \neg B\phi \wedge \neg B\neg B\phi$	(Plausibility)
$U\phi \rightarrow (\neg B)\phi^n \; \forall n \in N$	(Strong Plausibility)
$\neg BU\phi$	(BU-introspection)

the modeling of (un)awareness

- it is impossible to devise a non-trivial (un)awareness operator that satisfies most of the intuitively appealing properties above mentioned
- for instance, in *"Standard State-Space Models Preclude Unawareness"* (1998) , Dekel, Lipman & Rustichini show that it is impossible to have

  (i) a non-trivial awareness operator which satisfies Plausibility, U-introspection and BU-introspection
  (ii) a belief operator which satisfies either Necessitation or Monotonicity

main models of unawareness

- two main ways to circumvent the issue:

(i) endogenous characterization: awareness defined in terms of beliefs : Modica & Rustichini (1999), Heifetz, Meier & Schipper (2006), (2007a)

$$\mathcal{M}, s \vDash A\phi \Leftrightarrow \mathcal{M}, s \vDash B\phi \vee B\neg B\phi$$

main models of unawareness

- two main ways to circumvent the issue:

(i) endogenous characterization: awareness defined in terms of beliefs : Modica & Rustichini (1999), Heifetz, Meier & Schipper (2006), (2007a)

$$\mathcal{M}, s \vDash A\phi \Leftrightarrow \mathcal{M}, s \vDash B\phi \vee B\neg B\phi$$

(ii) exogenous characterization: Fagin & Halpern (1988), Halpern (2001) awareness© structures

$$\mathcal{M}, s \vDash A\phi \Leftrightarrow \phi \in \mathcal{A}(s)$$

where $\mathcal{A} : S \rightarrow \wp(\mathcal{L}(At))$

GSM structures, example

- state space S based on $At = \{p, q, r\}$:

pqr	$p\neg qr$	$pq\neg r$	$p\neg q\neg r$
$\neg pqr$	$\neg p\neg qr$	$\neg pq\neg r$	$\neg p\neg q\neg r$

- in actual state $s = pqr$, Pierre believes that p, does not know whether q and is unaware of r ; his non-standard state space $S_{p,q}$ and accessibility correspondance in s are

pq	$p\neg q$
$\neg pq$	$\neg p\neg q$

GSM structures, example

pqr	$p\neg qr$	$pq\neg r$	$p\neg q\neg r$
$\neg pqr$	$\neg p\neg qr$	$\neg pq\neg r$	$\neg p\neg q\neg r$

pq	$p\neg q$
$\neg pq$	$\neg p\neg q$

- conditions:
  - pqr in projected in pq: $\rho(pqr) = pq$
  - if pqr and $pq\neg r$ are projected in pq, pqr and $pq\neg r$ agree on p and q
  - if pqr and $pq\neg r$ are projected in pq, $R(pqr) = R(pq\neg r)$
  - $R(pqr) \subseteq S_{pq}$

GSM structures

A GSM structure is a t-uple $\mathcal{M} = (S, S', \pi, R, \rho)$

(i) S is a state space

(ii) $S' = \bigcup_{X \subseteq At} S'_X$ (where S'_X are disjoint) is a (non-standard) state space

(iii) $\pi : At \times S \to \{0, 1\}$ is a valuation for S

(iv) $R : S \to \wp(S')$ is an accessibility correspondence

(v) $\rho : S \to S'$ is a projection s.t. (1) if $\rho(s) = \rho(t) \in S'_X$, then (a) for each atomic formula $p \in X$, $\pi(s, p) = \pi(t, p)$ and (b) $R(s) = R(t)$ and (2) if $\rho(s) \in S'_X$, then $R(s) \subseteq S'_X$

- each state s is associated to a subjective state space S'_X
- one can extend R and π to the whole state space with π^*: if $s' \in S'_X$, then $\pi^*(s', p) = 1$ iff (a) $p \in X$ and (b) for all $s \in \rho^{-1}(s'), \pi(s, p) = 1$.

satisfaction relation

- one may then define as follows the satisfaction relation for each $s^* \in S^* = S \cup S'$ (Halpern's 2001 version):

(i) $\mathcal{M}, s^* \models p$ iff $\pi^*(s^*, p) = 1$

(ii) $\mathcal{M}, s^* \models \phi \wedge \psi$ iff $\mathcal{M}, s^* \models \phi$ and $\mathcal{M}, s^* \models \psi$

satisfaction relation

- one may then define as follows the satisfaction relation for each $s^* \in S^* = S \cup S'$ (Halpern's 2001 version):

(i) $\mathcal{M}, s^* \models p$ iff $\pi^*(s^*, p) = 1$

(ii) $\mathcal{M}, s^* \models \phi \wedge \psi$ iff $\mathcal{M}, s^* \models \phi$ and $\mathcal{M}, s^* \models \psi$

(iii) $\mathcal{M}, s^* \models \neg\phi$ iff $\mathcal{M}, s^* \not\models \phi$ and either $s^* \in S$, or $s^* \in S'_X$ and $\phi \in \mathcal{L}(X)$

satisfaction relation

- one may then define as follows the satisfaction relation for each $s^* \in S^* = S \cup S'$ (Halpern's 2001 version):

(i) $\mathcal{M}, s^* \vDash p$ iff $\pi^*(s^*, p) = 1$

(ii) $\mathcal{M}, s^* \vDash \phi \wedge \psi$ iff $\mathcal{M}, s^* \vDash \phi$ and $\mathcal{M}, s^* \vDash \psi$

(iii) $\mathcal{M}, s^* \vDash \neg\phi$ iff $\mathcal{M}, s^* \nvDash \phi$ and either $s^* \in S$, or $s^* \in S'_X$ and $\phi \in \mathcal{L}(X)$

(iv) $\mathcal{M}, s^* \vDash B\phi$ iff for each $t^* \in R^*(s^*)$, $\mathcal{M}, t^* \vDash \phi$

- crucial point: (iii) introduces partiality: if $p \notin X$ and $s^* \in S'_X$ then neither $\mathcal{M}, s^* \vDash p$ nor $\mathcal{M}, s^* \vDash \neg p$ (for short, $\mathcal{M}, s^* \Uparrow p$). More generally,

$$\mathcal{M}, s^* \Downarrow \phi \text{ for } s^* \in S'_X \text{ iff } \phi \in \mathcal{L}(X)$$

what (un)awareness is not

- MR 1999 and HMS 2006 define unawareness in terms of beliefs; this is OK *given their assumption that accessibility correspondences are partitional*

what (un)awareness is not

- MR 1999 and HMS 2006 define unawareness in terms of beliefs; this is OK *given their assumption that accessibility correspondences are partitional*
- **but** this endogenous characterization is not robust under change of the accessibility relation's properties. In the general case,
  - it is plausible that if Pierre is unaware of ϕ, he doesn't believe that ϕ nor believe that he doesn't believe that ϕ $(U\phi \rightarrow (\neg B\phi \wedge \neg B\neg B\phi))$
  - it is not plausible that if Pierre doesn't believe that ϕ nor believe that he doesn't believe that ϕ, he is necessarily unaware of ϕ $((\neg B\phi \wedge \neg B\neg B\phi) \rightarrow U\phi)$

example

- the actual state s is projected in $s_1 \in S'_{\{p,q\}}$
- $R(s_1) = \{s_2, s_3\}$ (hence $s_2, s_3 \in S'_{\{p,q\}}$ as well) ; $R(s_2) = \{s_2\}$; $R(s_3) = \{s_3\}$
- $\mathcal{M}, s_2 \vDash \neg p$, hence $\mathcal{M}, s_2 \vDash \neg Bp \wedge B\neg p$
- $\mathcal{M}, s_3 \vDash p$, hence $\mathcal{M}, s_3 \vDash Bp$
- $\mathcal{M}, s \vDash \neg Bp \wedge \neg B\neg Bp$ hence $\mathcal{M}, s \vDash U^{MR}p$
  But, for instance, $\mathcal{M}, s \vDash B(B\neg p \vee Bp)$

unawareness as partiality

- hence: keep the underlying GSM structure but change the definition of (un)awareness

unawareness as partiality

- hence: keep the underlying GSM structure but change the definition of (un)awareness
- the possible states that Pierre conceives do not “answer” to $?p$, $?q$ and $?r$: they answer only to $?p$ and $?q$
- unawareness may be seen as **partiality**: the possible states that Pierre conceives make true neither r nor $\neg r$

unawareness as partiality

- hence: keep the underlying GSM structure but change the definition of (un)awareness
- the possible states that Pierre conceives do not "answer" to $?p$, $?q$ and $?r$: they answer only to $?p$ and $?q$
- unawareness may be seen as **partiality**: the possible states that Pierre conceives make true neither r nor $\neg r$
- semantic characterization of unawareness in terms of **partiality**:

$$\mathcal{M}, s \vDash A\phi \text{ iff } \exists t \in R(s), \mathcal{M}, t \Downarrow \phi$$

- Let's call a P-GSM structure a GSM structure where the truth conditions of the unawareness operator are given in terms of partiality

partiality and awareness©

- Halpern 2001 relates GSM structures and awareness© structures; one obtains a still closer connection with P-GSM
- an awareness© structure $\mathcal{M} = (S, R, \mathcal{A}, \pi)$ is *propositionally determined* (pd) if (1) for each state s, $\mathcal{A}(s)$ is generated by some atomic formulas $X \subseteq At$ i.e. $\mathcal{A}(s) = \mathcal{L}(X)$ and (2) if $t \in R(s)$, then $\mathcal{A}(s) = \mathcal{A}(t)$

partiality and awareness©

- Halpern 2001 relates GSM structures and awareness© structures; one obtains a still closer connection with P-GSM
- an awareness© structure $\mathcal{M} = (S, R, \mathcal{A}, \pi)$ is *propositionally determined* (pd) if (1) for each state s, $\mathcal{A}(s)$ is generated by some atomic formulas $X \subseteq At$ i.e. $\mathcal{A}(s) = \mathcal{L}(X)$ and (2) if $t \in R(s)$, then $\mathcal{A}(s) = \mathcal{A}(t)$
- **Proposition** (see Halpern 2001 Thm 4.1)

1. For every pd awareness© structure $\mathcal{M}$ there exists a P-GSM structure $\mathcal{M}'$ based on the same state space S and the same valuation π s.t. for all formulas $\phi \in \mathcal{L}^{BA}(At)$ and each possible state s
$\mathcal{M}, s \vDash_{a©} \phi$ iff $\mathcal{M}', s \vDash_{P-GSM} \phi$

partiality and awareness©, *cont.*

2. For every P-GSM structure $\mathcal{M}$ there exists a awareness© structure $\mathcal{M}'$ based on the same state space S and the same valuation π s.t. for all formulas $\phi \in \mathcal{L}^{BA}(At)$ and each possible state s
$\mathcal{M}, s \vDash_{P-GSM} \phi$ iff $\mathcal{M}', s \vDash_{a©} \phi$

- **Corollary**: the axiom system K_X in Halpern 2001 is sound and complete for P-GSM structures.

axiom system K_X (Halpern 2001)

(PROP) Instances of propositional tautologies
(MP) From ϕ and $\phi \rightarrow \psi$ infer ψ
(A0) $B\phi \Rightarrow A\phi$
(K) $B\phi \wedge B(\phi \rightarrow \psi) \rightarrow B\psi$
(Gen) From ϕ infer $A\phi \rightarrow B\phi$
(A1) $A\phi \leftrightarrow A\neg\phi$
(A2) $A(\phi \wedge \psi) \leftrightarrow (A\phi \wedge A\psi)$
(A3) $A\phi \leftrightarrow AA\phi$
(A4) $AB\phi \leftrightarrow A\phi$
(A5) $A\phi \rightarrow BA\phi$
(Irr) If no atomic formulas in ϕ appear in ψ, from $U\phi \rightarrow \psi$ infer ψ

probabilistic unawareness

- as stressed by Cozic (2007) and Pucella & Halpern (2007), one of the main advantages of the impossible states framework is that it can be straightforwardly extended to the probabilistic case

probabilistic unawareness

- as stressed by Cozic (2007) and Pucella & Halpern (2007), one of the main advantages of the impossible states framework is that it can be straightforwardly extended to the probabilistic case
- actually, P-GSM structures can be given probabilistic analogues
- sketch of the probabilistic extension: explicit probabilistic structures (EPS) are logical versions of type spaces from GT (see Aumann 1999, Heifetz & Mongin 2001) and true probabilistic analogues of Kripke structures

explicit probabilistic structures (EPS)

- **Definition** : the set $\mathcal{L}^L(At)$ of formulas of an explicit probabilistic language based on a set At of propositional variables is defined by :

$$\phi ::= p | \neg\phi | \phi \vee \psi | L_a\phi$$

where $p \in At$ and $a \in [0, 1] \subseteq \mathbb{Q}$.

$L_a\phi$ means intuitively that the agent believes at least to degree a that ϕ

explicit probabilistic structures (EPS), cont.

- **Definition** : an **explicit probabilistic structure** for $\mathcal{L}^L(At)$ is a 3-tuple $\mathcal{M} = (S, \pi, P)$ where $P : S \rightarrow \Delta(S)$ assigns to every state a probability distribution on the state space.
- Satisfaction condition for L_a :
$\mathcal{M}, s \vDash L_a\phi \Leftrightarrow P(s)([[\phi]]) \geq a$

explicit probabilistic structures (EPS), cont.

- **Definition** : an **explicit probabilistic structure** for $\mathcal{L}^L(At)$ is a 3-tuple $\mathcal{M} = (S, \pi, P)$ where $P : S \rightarrow \Delta(S)$ assigns to every state a probability distribution on the state space.
- Satisfaction condition for L_a :
  $\mathcal{M}, s \vDash L_a\phi \Leftrightarrow P(s)([[\phi]]) \geq a$
- higher-order (partial) beliefs are induced in the same way that higher-order (full) beliefs are induced by the accessibility relation

probabilistic unawareness, first attempt

- **Definition** : an **P-GSM explicit probabilistic structure** for $\mathcal{L}^{LA}(At)$ is a t-tuple $\mathcal{M} = (S, S', \pi, P, \rho)$ where

(i) S is a state space

(ii) $S' = \bigcup_{\Phi \subseteq At} S'_{\Phi}$ (where S'_{Φ} are disjoint) is a state space

(iii) $\pi : At \times S \rightarrow \{0, 1\}$ is a valuation for S

(iv) $P : S \rightarrow \Delta(S')$

(v) $\rho : S \rightarrow S'$ is a projection s.t. (1) if $\rho(s) = \rho(t) \in S'_{\Phi}$, then (a) for each atomic formula $p \in \Phi$, $\pi(s, p) = \pi(t, p)$ and (b) $P(s) = P(t)$ and (2) if $\rho(s) \in S'_{\Phi}$, then $Supp(P(s)) \subseteq S'_{\Phi}$

probabilistic unawareness, first attempt

- some good news:
  for all GSM-EPS $\mathcal{M}$ and all standard state s, unawareness precludes positive probability:
  $\mathcal{M}, s \models U\phi \rightarrow \neg L_a \phi$ for $a > 0$
  $\mathcal{M}, s \models U\phi \rightarrow \neg L_a \neg \phi$ for $a > 0$
  $\mathcal{M}, s \models \neg L_a U\phi$ for $a > 0$

probabilistic unawareness, first attempt

- some good news:
  for all GSM-EPS $\mathcal{M}$ and all standard state s, unawareness precludes positive probability:
  $\mathcal{M}, s \vDash U\phi \rightarrow \neg L_a\phi$ for $a > 0$
  $\mathcal{M}, s \vDash U\phi \rightarrow \neg L_a\neg\phi$ for $a > 0$
  $\mathcal{M}, s \vDash \neg L_a U\phi$ for $a > 0$
- but some (very) bad news:
  for all GSM-EPS $\mathcal{M}$ and all standard state s,
  $\mathcal{M}, s \vDash U\phi \rightarrow L_0\phi$
  $\mathcal{M}, s \vDash U\phi \rightarrow L_0\neg\phi$
  $\mathcal{M}, s \vDash U\phi \rightarrow L_1L_0\phi$ (!!)

probabilistic unawareness, second attempt

- satisfaction condition for $L_a\phi$:
  $\mathcal{M}, s \vDash L_a\phi \Leftrightarrow P(s)([[\phi]]) \geq a$ **and** $\mathcal{M}, \rho(s) \Downarrow \phi$

probabilistic unawareness, second attempt

- satisfaction condition for $L_a\phi$:
  $\mathcal{M}, s \vDash L_a\phi \Leftrightarrow P(s)([[\phi]]) \geq a$ **and** $\mathcal{M}, \rho(s) \Downarrow \phi$
- in this case, the following holds:

$A\phi \leftrightarrow A\neg\phi$	(Symmetry)
$A\phi \leftrightarrow AA\phi$	(Self-Reflection)
$U\phi \rightarrow UU\phi$	(U-introspection)
$U\phi \rightarrow \neg L_a\phi \wedge \neg L_a\neg L_a\phi$	(Plausibility)
$U\phi \rightarrow (\neg L_a)^n\phi \ \forall n \in N$	(Strong Plausibility)
$\neg L_aU\phi$	(L_aU-introspection)
$L_0\phi \leftrightarrow A\phi$	

further issues

1. axiomatizing probabilistic unawareness
2. becoming aware
3. multi-agent unawareness
4. applications to decision theory and game theory

axiomatizing probabilistic unawareness

- Heifetz & Mongin 2001 have axiomatized explicit probabilistic structures i.e. probabilistic structures with full awareness

axiomatizing probabilistic unawareness

- Heifetz & Mongin 2001 have axiomatized explicit probabilistic structures i.e. probabilistic structures with full awareness
- Halpern 2001 has axiomatized pd awareness© structures hence P-GSM structures

axiomatizing probabilistic unawareness

- Heifetz & Mongin 2001 have axiomatized explicit probabilistic structures i.e. probabilistic structures with full awareness
- Halpern 2001 has axiomatized pd awareness$^{©}$ structures hence P-GSM structures
- the next step is to axiomatize P-GSM probabilistic structures

becoming aware

- Pierre may be initially unaware of ϕ and become aware of ϕ :

(1) when someone gives Pierre an information that involves ϕ

(2) when someone asks Pierre what he thinks about ϕ

becoming aware

- Pierre may be initially unaware of ϕ and become aware of ϕ :

(1) when someone gives Pierre an information that involves ϕ

(2) when someone asks Pierre what he thinks about ϕ

- when one thinks about (1) for full beliefs, things may look simple:

- initially, Pierre is only aware of p, neither q nor p: $R(s) \subseteq S_p$
- Pierre is informed that q
  (i) first, the structure is modified such that $R'(s) \subseteq S_{p,q}$
  (ii) then, the $\neg q$-states are eliminated

becoming aware, cont.

- but *even for scenario of type (1)*, the probabilistic case is much more tricky
- one could reason like this:
- initially, Pierre is only aware of p, neither q nor p: $Supp(P(s)) \subseteq S_p$
- Pierre is informed that q
  (i) first, the structure is modified: $Supp(P'(s)) \subseteq S_{p,q}$ and for each ϕ of $\mathcal{L}(\{p\})$, $P(s)([[\phi]] = P'(s)([[\phi]])$
  (ii) then, Pierre conditionalizes on the information that q
- but the new probability of p could be affected by the fact that the agent learns that q (intuitively, if p and q are not independent)