Imaging and *Sleeping Beauty*

Mikaël Cozic

DEC (Ecole Normale Supérieure Ulm)

TARK 2007 - 25/06/2007

1. Halfers & Thirders

SB's scenario

- on sunday evening (t_0), SB is put to sleep. A fair coin is tossed, SB doesn't know the outcome of the toss.
- on monday morning (t_1), SB is awaken; she is not told which day it is.
- some minutes later (t_2), SB is told that it is monday
- what follows depends on the result of the toss :
  (i) if the coin lands heads (*HEADS*), SB is put to sleep until the end of the week.
  (ii) if the coin lands tails (*TAILS*), SB is awaken on tuesday morning but before a drug is given to her s.t. her tuesday's and monday's awakenings are not distinguishable

2 questions

- Focus : SB's degree of belief that *HEADS*
- 2 questions

  Q1 what should be SB's degree of belief that *HEADS* à t_1 ?

  Q2 what should be SB's degree of belief that *HEADS* à t_2 ?
- Notation:
  - P_0 = SB's credence at t_0 (sunday evening)
  - P_1 = SB's credence at t_1 (monday morning at her awakening)
  - P_2 = SB's credence at t_2 (monday morning after having learned that it is monday)

Halfers and Thirders

- Thirders' Claim (Elga, 2000): $P_1(HEADS) = 1/3$
- Halfers' Claim (Lewis, 2001): $P_1(HEADS) = 1/2$
- But answers to $Q1$ are connected to answers to $Q2$:

	A. Elga	D. Lewis
Q1	1/3	1/2
Q2	1/2	2/3

common ground

- state space ("centered worlds") $W = \{HM, TM, TT\}$ where
  - in *HM* the coin lands heads and it's monday
  - in *TM* the coin lands tails and it's monday
  - in *TT* the coin lands tails and it's tuesday
- *common ground*:
  - $P_1(TM) = P_1(TT)$
    (Indifference or Laplacean Principle)
  - $P_2(HEADS) = P_1(HEADS|MONDAY) = P_1(HEADS|\{HM, TM\})$
    (belief change by conditionalization)
  - $P_0(HEADS) = P_0(TAILS) = 1/2$
    ($\approx$ Principal Principle)

Elga's argument

- basic idea: the coin could be tossed on monday night. Hence, by the Principal Principle,
  (E) $P_2(HEADS) = P_0(HEADS) = 1/2$
- From (E) and the common ground, it follows that $P_1(HEADS) = 1/3$ by "backtracking" conditionalization since
  $P_2(HEADS) = P_1(HEADS|MONDAY) = 1/2$.
- "Bottom-Up" argument which answers to $Q1$ by answering antecendently to $Q2$

Lewis's argument

- basic idea: when SB is awakened on tuesday morning (t_1), she acquires no relevant evidence w.r.t. *HEADS* vs. *TAILS*. Hence her credence in *HEADS* should be unchanged: (L) $P_1(HEADS) = 1/2 = P_1(TAILS)$
- From (L) and the common ground, it follows that $P_2(HEADS) = 2/3$
- "Top-Down" argument which answers to $Q2$ by answering antecendently to $Q1$

2. Conditionalizing vs. Imaging

starting point

- starting point:
  1) both Elga's and Lewis's intuitions are appealing. If one would put them together, one would obtain a *double halfer* position according to which
  $P_1(HEADS) = P_2(HEADS) = 1/2$
  2) given the common ground, these intuitions are not compatible
  Why ? Since credence is changed by conditionalization, necessarily, $P_1(HEADS) \neq P_2(HEADS)$

conditionalization (1)

- the situation could be different with another rule of belief change. But is there any reason to question conditionalization ?

conditionalization (1)

- the situation could be different with another rule of belief change. But is there any reason to question conditionalization ?
- the proposition that SB learns at t_2 bears on her *temporal location* and is *context(time)-sensitive*
- context-sensitive propositions are *in general* problematic for conditionalization.
- two properties of conditionalization are problematic:

(i) concentration

(ii) partiality

conditionalization (2)

(i) concentration: the beliefs of a conditionalizer become more and more concentrated when she learns more and more information.
If information I is compatible with initial beliefs P ($I \cap Supp(P) \neq \emptyset$), then $Supp(P(.|I)) \subseteq Supp(P)$

- Particular cases:
  - if a proposition A is certain and compatible with the information, it will remain certain (*preservation*, Gardenförs (1988))
  - if a proposition has null probability, its probability will never be positive

conditionalization (3)

- SB: (a) the probability of *HM* necessarily increases when SB learns that it's monday (b) if at t_0 SB believes that it's sunday, she cannot at t_1 believe that it's monday or tuesday

(ii) partiality :conditionalization is undefined when the information is incompatible with initial beliefs ($I \cap Supp(P) = \emptyset$)

SB: conditionalization doesn't say how to go from P_0 to P_1

conditionalization and SB

- these properties suggest that with context-sensitive propositions, conditionalization may not be a reliable guide
- maybe the discomfort with both Halfers and Thirders could come from a mistaken use of conditionalization...

$\Rightarrow$ is there another probabilistic change rule available ?

imaging

- Lewis (1976) introduces the *imaging* rule. Let $A \subseteq W$ be a proposition. w_A is the closest world to w where A is true (cf. Stalnaker's semantics for conditionals)
- Suppose that the agent learns that A ; the imaging rule says that the weight of world w is entirely allocated to world w_A.
  If P is the initial distribution, then the posterior probability is defined as follows:

  $$P^{Im(A)}(w) = \sum_{\{w' \in W : w = w'_A\}} P(w')$$

- Lewis: "no gratuitous movement of probability from worlds to dissimilar worlds"

example: Apple & Banana

- the basket of fruits state space

AB	$A\neg B$
$\neg AB$	$\neg A\neg B$

- initial probability P:

1/3	1/3
1/3	0

- change of P by imaging on $I = \{A\neg B, \neg A\neg B\}$ with $AB_I = A\neg B$ and $\neg AB_I = \neg A\neg B$:

0	2/3
0	1/3

is imaging serious ?

- in general, *imaging* is not considered as a serious rule of credence change.
- Lewis (1976) introduces imaging because it is the rule that matches Stalnaker's conditional
- Gardenförs (1988) rejects imaging because it violates *preservation*.
- but a (cognitive) justification of imaging has been recently proposed by Walliser & Zwirn (2002).
- basic idea : conditionalization is appropriate in some kind of contexts (revising), imaging in other kinds of contexts (updating)

3. Revising vs. Updating

contexts of belief change

- 2 contexts of belief change:

  1) *revising* contexts: the agent learns an information about an environment that is supposed to be stable.

  2) *updating context*: the agent learns an information about a (potential) change in the environment

contexts of belief change

- 2 contexts of belief change:

  1) *revising* contexts: the agent learns an information about an environment that is supposed to be stable.

  2) *updating context*: the agent learns an information about a (potential) change in the environment

- Katzuno & Mendelzon (1992) argue that principles of belief change should be different in revising contexts and updating contexts

example

- initial belief set $K = \{AB, A\neg B, \neg AB\}$:

AB	$A\neg B$
$\neg AB$	

- information $I = \{A\neg B, \neg A\neg B\}$

revising
"there is no banana"

$A\neg B$

updating
"there is no more banana (if there was any)"

$A\neg B$
$\neg A\neg B$

rationality postulates

- **revising** (AGM postulates):

(A5) If $(K^r A) \cap B \neq \emptyset$, then $K^r(A \cap B) \subseteq (K^r A) \cap B$ (Super-expansion)

- **updating** (KM postulates):

(A6) If $\exists w_0 \in W$ t.q. $K = \{w_0\}$ and $(K^m A) \cap B \neq \emptyset$, then $K^m(A \cap B) \subseteq (K^m A) \cap B$ (Pointwise Super-expansion)

(A7) $(K \cup K')^m A = (K^m A) \cup (K'^m A)$ (Left Distributivity)
(A7) "gives each of the possible worlds equal consideration" (KM 1992)

similarity relations

(i) **revising :** the primitive is a family $\leq_K$ of similarity relations on worlds. $K^r A$ are the closer worlds of K where A is the case.
$K^r A = \{w' \in A \text{ s.t. } \forall w'' \in A, w' \leq_K w''\}$

$\Rightarrow$ *global* minimal change of belief set

(ii) **updating:** the primitive is a family $\leq_w$ of similarity relations on worlds. $K^u A$ is the union of the closer worlds of w where A is the case, for each $w \in K$:
$K^u A = \{w' \in A \text{ s.t. } \exists w \in K \text{ and } \forall w'' \in A : w' \leq_w w''\}$

$\Rightarrow$ *local* minimal change of belief set

4. Updating, Imaging and SB

justification of imaging

- Katzuno & Mendelzon (1992) suggest that conditionalization corresponds to belief revision whereas imaging corresponds to belief updating.
- Walliser & Zwirn (2002) have proven the following results:

(i) conditionalization-like change rules may be derived from probabilistic transcription of AGM postulates for belief revision

(ii) imaging-like change rules may be derived from probabilistic transcription of KM postulates for belief updating

imaging and SB

- basic idea: when SB is told that it is monday (t_2), she has an information about a feature of her situation that has changed since her initial credence.
- hence, the imaging rule seems to be more appropriate to model SB's belief when she learns that it is monday
- question: which similarity relation ?
- *assumption: TM* is the closest *MONDAY*-world to *TT*

double-halfer's case

- Lewis's starting point:
  $P_1(HEADS) = P_1(HM) = 1/2$ and
  $P_1(TM) = P_1(TT) = 1/4$
  By *imaging*,
  $P_2(TM) = P_1(TM) + P_1(TT) = 1/2$
  $P_2(HEADS) = P_2(HM) = P_1(HM) = 1/2$
- Elga's starting point:
  $P_2(HEADS) = P_2(HM) = 1/2 = P_2(TAILS) = P_2(TT)$
  By backtracking l'*imaging*
  $P_1(HEADS) = P_2(HEADS) = 1/2$

5. Discussion

the argument

- summary of the argument:

(P1) in a (probabilistic) updating context, one should rely on imaging and not on conditionalization

(P2) when SB learns that it is monday, it is an updating context

(C) SB should rely on imaging when she learns that it is monday

main objection

- when SB is aware at t_1 that she is on monday or tuesday, this is a true updating context since the day it is is different from t_0.
  But when SB learns that it is monday at t_2, this is not about a change that took place between t_1 and t_2
- therefore, one could be tempted to think that even if the imaging rule is appropriate in updating contexts, it is not appropriate at t_2 in SB scenario since it is a revising context
- general problem: when two successive pieces of information I_1, I_2 at $t_1 < t_2$ bear on a change that took place between t_0 and t_1, should the second information be processed by conditionalization or imaging?

example: Apple, Banana & Coconut

- state space

ABC	$AB\neg C$	$A\neg BC$	$A\neg B\neg C$
$\neg ABC$	$\neg AB\neg C$	$\neg A\neg BC$	$\neg A\neg B\neg C$

- initial probability at t_0:

0	1/4	1/4	1/4
1/4	0	0	0

- information $I_1 = \{A\neg BC, A\neg B\neg C, \neg A\neg BC, \neg A\neg B\neg C\}$ ("there is no more banana, if there was any") at t_1
- *imaging* on I_1 : $P_1 = P_0^{Im(I_1)}$

0	0	1/4	1/2
0	0	1/4	0

example, cont.

- information $I_2 = \{A\neg B\neg C, \neg A\neg B\neg C\}$ (“there is no more banana, if there was any, and there is no more coconut, if there was any”) at t_2
- *imaging* on I_2 :

0	0	0	3/4
0	0	0	1/4

Rem: $P_0^{Im(I_2)} = P_1^{Im(I_2)}$

- conditionalization on I_2:

0	0	0	1
0	0	0	0

Rem: $P_0^{Cond(I_2)} = P_1^{Cond(I_2)}$

analysis

- in both cases, the change takes place only between t_0 and t_1
- in both cases, the second piece of information I_2 bears on a change that took place between t_0 and t_1 and *refines* I_1 (i.e. $I_2 \subset I_1$)
- *claim:* if one is convinced of the distinction between revising and updating by examples like Apple & Banana, one should prefer $P_1^{Im(I_2)}$ to $P_1^{Cond(I_2)}$ in Apple, Banana & Coconut

conclusion

- the imaging rule opens the way to a true reconciliation between Halfers and Thirders basic intuitions
- the distinction between revising and updating is not clear-cut enough for the argument to be definitive in a scenario like SB
- an unified theoretical framework that makes explicit both types of contexts and timing of information is needed

Dutch Book dynamique

- P_0 distribution équiprobable sur $Supp(P_0) = \{AB, \neg AB, A\neg B\}$. $I = \{\neg A\neg B, A\neg B\}$.
- $P^{Im(I)}(AB) = P^{Im(I)}(\neg AB) = 0$ $P^{Im(I)}(A\neg B) = 2/3$. On suppose que $P_1 = P^{Im(I)}$.
- notons que $P^{Im(I)}(A) = 2/3$ alors que $P(A|I) = 1$
- Trois paris :
  (a) [1 si $A \wedge \neg B$, 0 sinon]
  (b) [$x = P_0(A|I) = 1$ si B, 0 sinon]
  (c) [$y = P_0(A|I) - P_1(A) = 1/3$ si $\neg B$, 0 sinon]
  L'agent les achète à $1/3 + 2/3 + 1/9 = 11/9$.

Dutch Book dynamique

- Cas 1 : il est faux que $I = \neg B$, alors l'agent a une perte sèche de $y.P_0(I) = 1/3.1/3 = 1/9$
- Cas 2 : on informe l'agent que $I = \neg B$, alors dans le *Dutch Book* diachronique, le *bookie* lui rachète le pari (d) [1 si A, 0 sinon] pour $P_1(A)$. L'agent essuie aussi une perte sèche de $y.P_0(I) = 1/3.1/3 = 1/9$.
- Problème : le rachat du pari.
  Pari (a) porte sur : il est vrai à t_0 que A et que $\neg B$
  Pari (d) porte sur : il est vrai à t_1 que A