Théories formelles de la rationalité : logique & théorie de la décision

D. Bonnay / M. Cozic / H. Galinon

Cogmaster rentrée 2011

Le triptyque du jour

- *Logique*
  théorie des inférences déductives

Si j'accepte un certain nombre de propositions,
quelles autres propositions suis-je *par là même* commis à accepter ?

- *Théorie subjective des probabilités*
  théorie du raisonnement dans l'incertain

Si je crois certaines propositions à certains degrés,
quelles contraintes de cohérences pèsent sur ces croyances,
comment doivent-elles évoluer ?

- *Théorie de la décision*
  théorie du choix

Comment dois-je agir
étant donné ce que je crois et ce que je désire ?

Un tryptique cohérent

Ces trois théories ont en commun

- leur objet
  la rationalité
- leur statut
  elles sont (d'abord) *normatives*
- leur méthode
  elles sont *mathématisées* et *intégrées* entre elles

The penguin curse

La logique a d'abord affaire à un certain type d'activités mentales : *les inférences déductives*

The penguin curse

La logique a d'abord affaire à un certain type d'activités mentales : *les inférences déductives*

The penguin curse

La logique a d'abord affaire à un certain type d'activités mentales : *les inférences déductives*

Intuitivement, le raisonnement du pingouin n'est pas correct.

Intuitivement, le raisonnement du pingouin n'est pas correct.

... Pourquoi ?

Intuitivement, le raisonnement du pingouin n'est pas correct.

... Pourquoi ?

La logique développe des *langages formels*
dans lesquels peuvent être définis :

- la notion de raisonnement valide
  (comme une certaine relation entre prémisses et conclusion)
- des systèmes de preuve
  (qu'on peut utiliser pour montrer qu'une conclusion particulière suit bien de certaines prémisses)

La négation, ¬

Exemple : ‘Pierre *n’*est *pas* content.’

p : Pierre est content

p	$\neg p$
1	
0	

La négation, ¬

Exemple : ‘Pierre *n’*est *pas* content.’

p : Pierre est content

p	$\neg p$
1	0
0	

La négation, ¬

Exemple : ‘Pierre *n’*est *pas* content.’

p : Pierre est content

p	$\neg p$
1	0
0	1

La conjonction, $\wedge$

Exemple : 'Pierre est content *et* Marie est triste.'

p : Pierre est content
q : Marie est triste

p	q	$p \wedge q$
1	1	
1	0	
0	1	
0	0	

La conjonction, $\wedge$

Exemple : 'Pierre est content *et* Marie est triste.'

p : Pierre est content
q : Marie est triste

p	q	$p \wedge q$
1	1	1
1	0	
0	1	
0	0	

La conjonction, $\wedge$

Exemple : 'Pierre est content *et* Marie est triste.'

p : Pierre est content
q : Marie est triste

p	q	$p \wedge q$
1	1	1
1	0	0
0	1	
0	0	

La conjonction, $\wedge$

Exemple : 'Pierre est content *et* Marie est triste.'

p : Pierre est content
q : Marie est triste

p	q	$p \wedge q$
1	1	1
1	0	0
0	1	0
0	0	

La conjonction, $\wedge$

Exemple : 'Pierre est content *et* Marie est triste.'

p : Pierre est content
q : Marie est triste

p	q	$p \wedge q$
1	1	1
1	0	0
0	1	0
0	0	0

L'implication, $\rightarrow$

Exemple : '*Si* Pierre est content *alors* Marie est triste.'

p : Pierre est content
q : Marie est triste

p	q	$p \rightarrow q$
1	1	
1	0	
0	1	
0	0	

L'implication, $\rightarrow$

Exemple : '*Si* Pierre est content *alors* Marie est triste.'

p : Pierre est content
q : Marie est triste

p	q	$p \rightarrow q$
1	1	
1	0	0
0	1	
0	0	

L'implication, $\rightarrow$

Exemple : '*Si* Pierre est content *alors* Marie est triste.'

p : Pierre est content
q : Marie est triste

p	q	$p \rightarrow q$
1	1	1
1	0	0
0	1	
0	0	

L'implication, $\rightarrow$

Exemple : '*Si* Pierre est content *alors* Marie est triste.'

p : Pierre est content
q : Marie est triste

p	q	$p \rightarrow q$
1	1	1
1	0	0
0	1	1
0	0	1

The penguin strikes back

Analyzing the penguin

If John is lucky, Mary is angry
Mary is angry

John is lucky

Analyzing the penguin

$$\frac{\text{If John is lucky, Mary is angry} \quad \text{Mary is angry}}{\text{John is lucky}}$$

p : John is lucky
q : Mary is angry

Analyzing the penguin

$$\frac{\begin{array}{l}\text{If John is lucky, Mary is angry}\\ \text{Mary is angry}\end{array}}{\text{John is lucky}}$$

p : John is lucky
q : Mary is angry

$$\frac{\begin{array}{l}p \rightarrow q\\ q\end{array}}{p}$$

Analyzing the penguin

If John is lucky, Mary is angry
Mary is angry

John is lucky

p : John is lucky
q : Mary is angry

$$\frac{p \rightarrow q \quad q}{p}$$

p	q	$p \rightarrow q$
1	1	1
1	0	0
0	1	1
0	0	1


GLASBERGEN

Mondes possibles et vérité

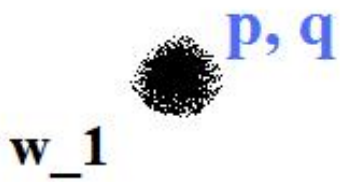
p	q	
1	1	w_1
1	0	w_2
0	1	w_3
0	0	w_4

Les lignes du tableau représentent des mondes possibles.

On vient de définir inductivement la notion de vérité dans un monde possible.

notation : $w \vDash \phi$ est mis pour ϕ est vrai en w

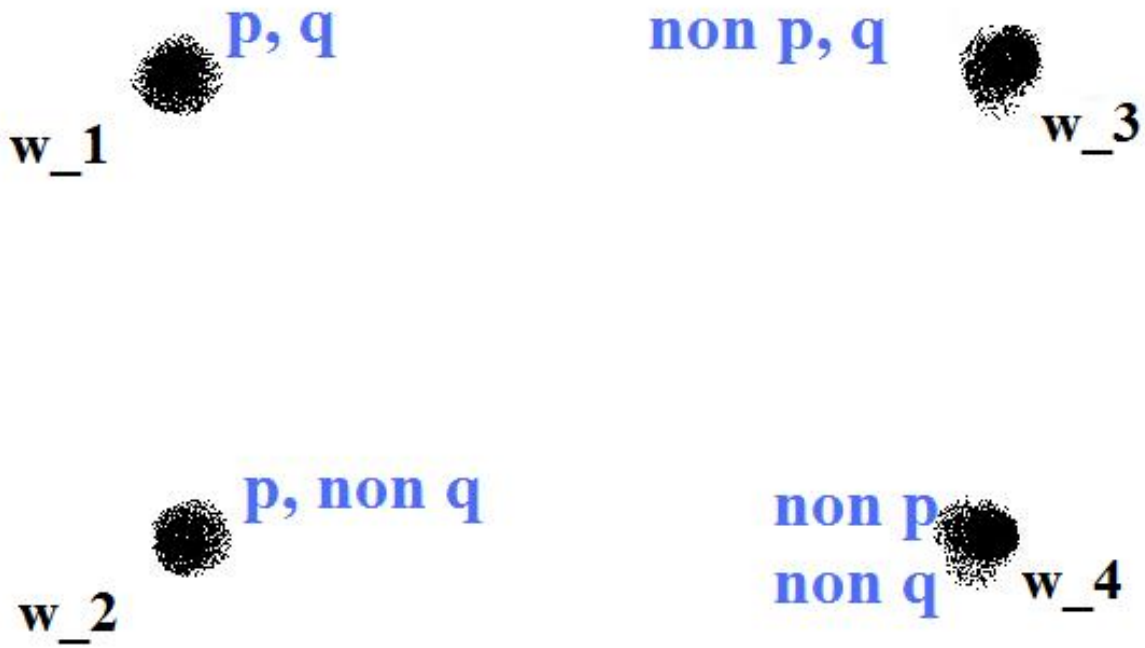
Mondes possibles et vérité



- $w_1 \models p \wedge q$
- $w_2 \not\models p \wedge q$
- $w_4 \models \neg p \wedge \neg q$

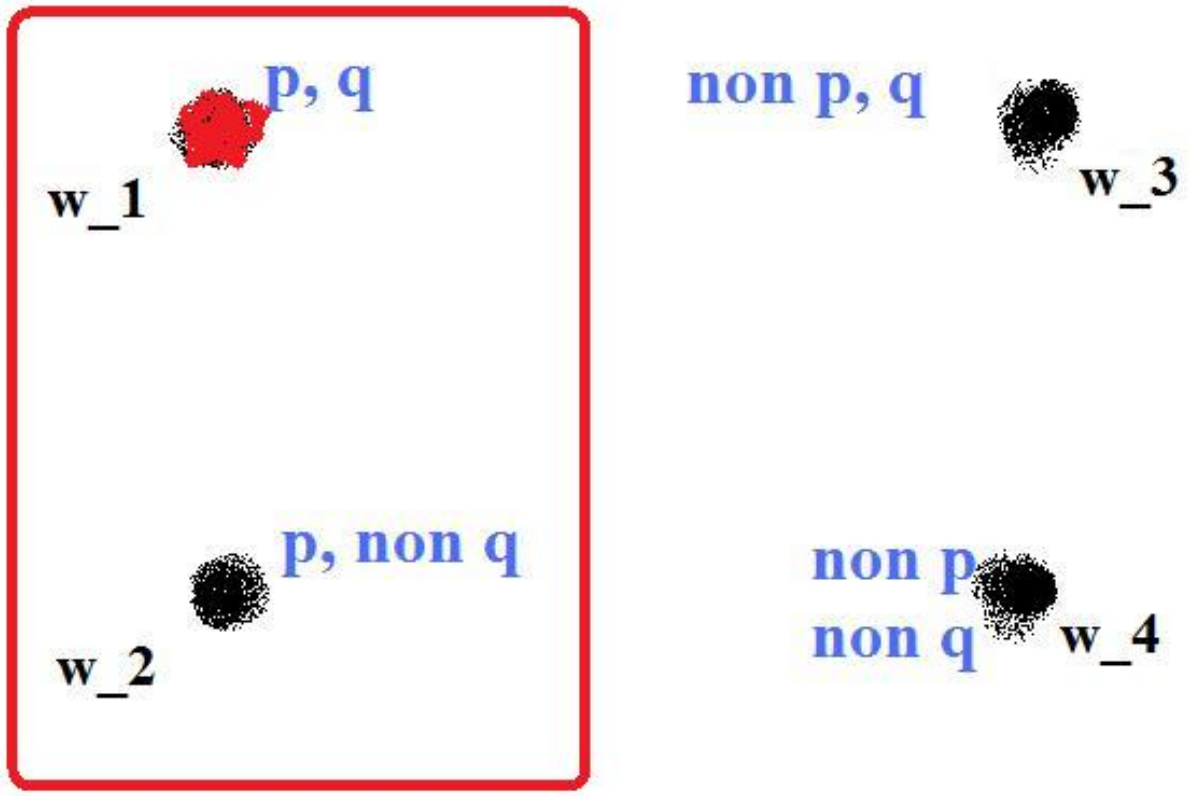
La croyance

L'état doxastique d'un agent est caractérisé par l'ensemble des mondes qu'il considère comme possibles.

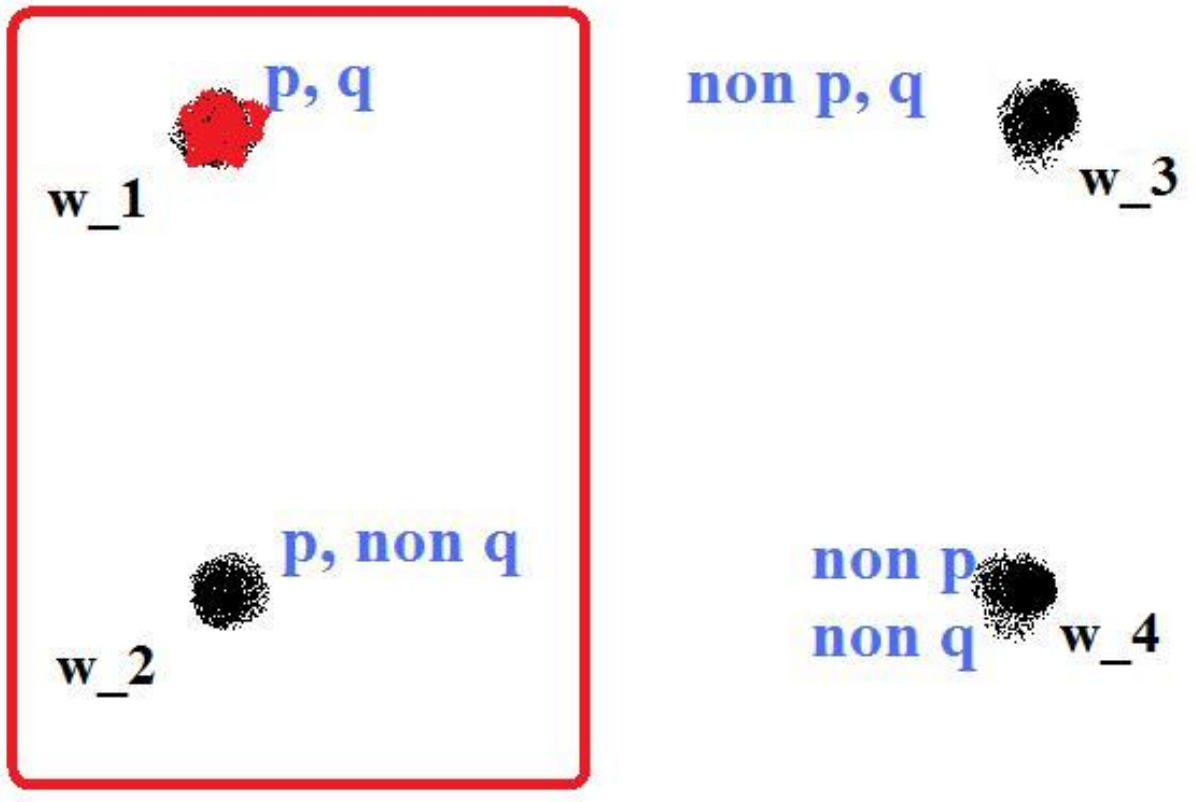


La croyance

L'état doxastique d'un agent est caractérisé par l'ensemble des mondes qu'il considère comme possibles.



Introducing *B*



Introduisons *B* pour “l’agent croit que”

- $w_1 \models Bp$
- $w_1 \not\models Bq$
- $w_1 \not\models B\neg q$

Sémantique de la croyance

Idée

Un agent croit que p
ssi p est vrai dans tous les mondes qu'il considère comme possibles.

Sémantique de la croyance

Idée

Un agent croit que p
ssi p est vrai dans tous les mondes qu'il considère comme possibles.

$w \longmapsto f(w)$
À un monde w, on associe un ensemble non vide de mondes, $f(w)$, l'ensemble des mondes que l'agent considère comme possibles en w.

Sémantique de la croyance

Idée

Un agent croit que p
ssi p est vrai dans tous les mondes qu'il considère comme possibles.

$w \longmapsto f(w)$
À un monde w, on associe un ensemble non vide de mondes, $f(w)$, l'ensemble des mondes que l'agent considère comme possibles en w.

Definition

$w \vDash B\phi$ ssi pour tout $w' \in f(w)$, $w' \vDash \phi$

Propriétés de B

- les croyances ne sont pas forcément véridiques,

On peut avoir $w \vDash B\phi$ et $w \nvDash \phi$
Pour que cela soit le cas, il faudrait exiger $w \in f(w)$.

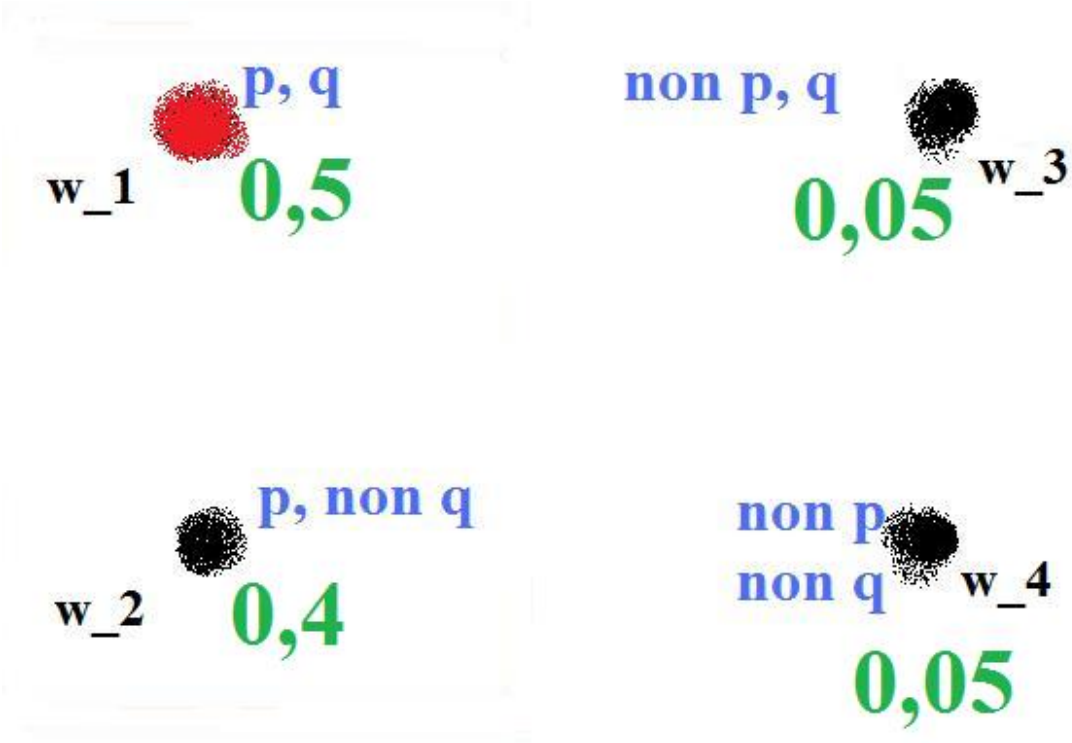
- les croyances sont cohérentes,

On n'a jamais $w \vDash B\phi$ et $w \vDash B\neg\phi$

- les croyances s'agrègent

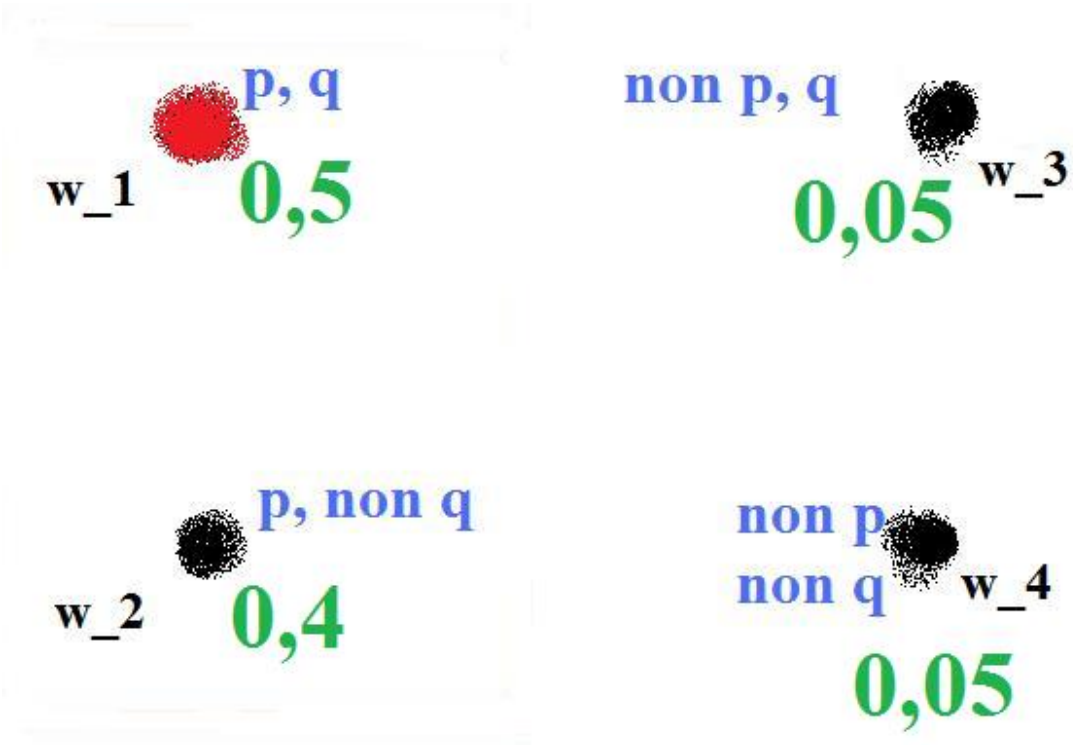
Si $w \vDash B\phi$ et $w \vDash B\psi$ alors $w \vDash B(\phi \wedge \psi)$

Probabilités



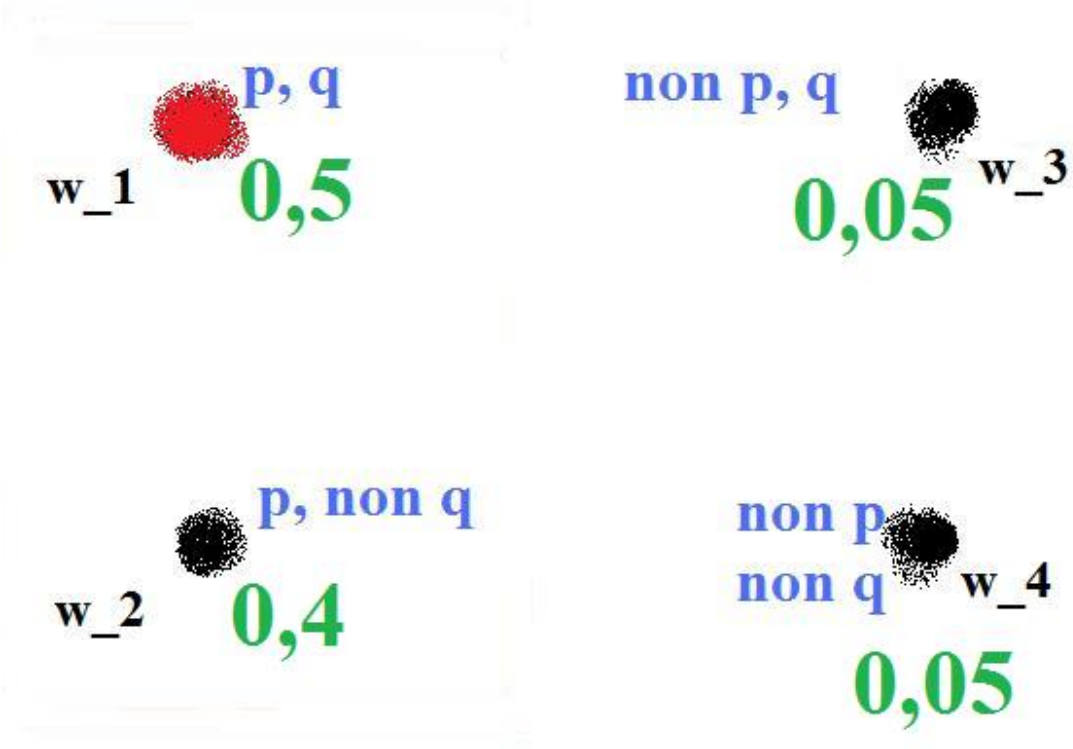
On peut associer à chaque monde w un poids $pr(w)$ correspondant à la probabilité subjective que ce monde soit le monde réel (la somme des $pr(w)$ doit être 1).

Probabilités



- $Pr(p) = 0,55$
- $Pr(\neg p) = 0,45$
- $Pr(q) = 0,8$

Probabilités



$$Pr(\phi) = \sum_{w_i \models \phi} pr(w_i)$$

est une mesure de probabilité.

La thèse de Locke

Quel lien y a-t-il entre

- la croyance B
- les probabilités subjectives Pr ?

La thèse de Locke

Quel lien y a-t-il entre

- la croyance B
- les probabilités subjectives Pr ?

Idée

Un agent croit que ϕ
ssi sa probabilité subjective que ϕ est suffisamment grande.

La thèse de Locke

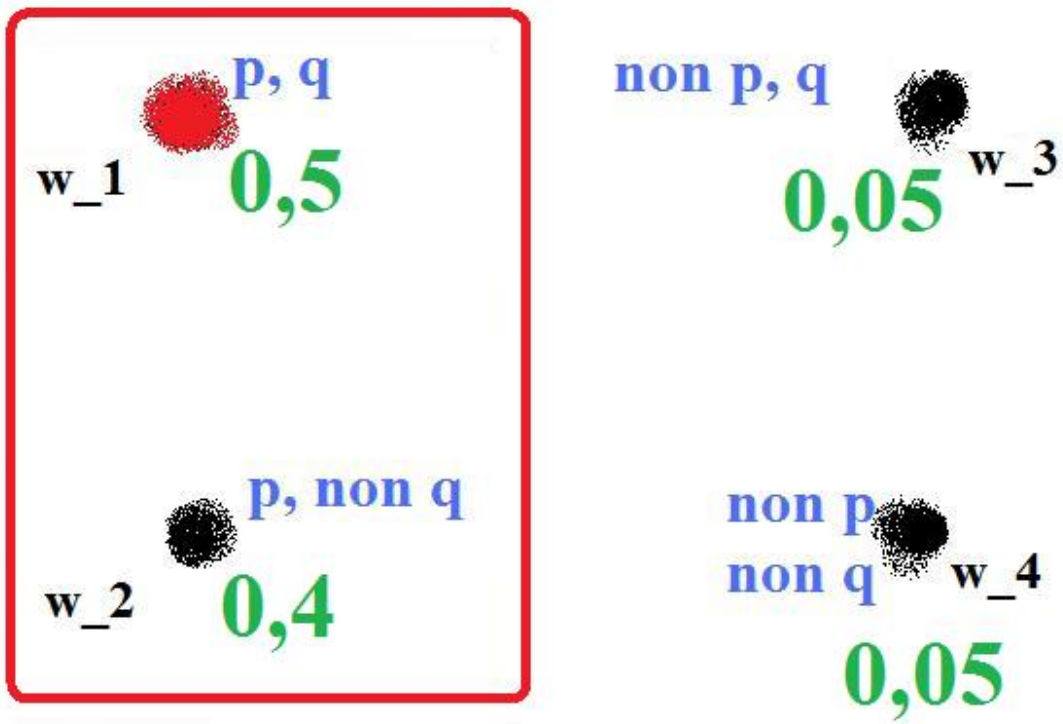
Quel lien y a-t-il entre

- la croyance B
- les probabilités subjectives Pr ?

Idée

Un agent croit que ϕ
ssi sa probabilité subjective que ϕ est suffisamment grande.

> *The mind, if it will proceed rationally, ought to examine all the grounds of probability for or against any probable proposition and upon due balancing of the whole reject or receive it.*
>
> John Locke
> *An Essay Concerning Human Understanding*, 1690



$B(\phi)$ ssi $Pr(\phi) \geq 0,8$

Une loterie

On considère une loterie avec 100 tickets. Il y a un et un seul ticket gagnant, et il est rationnel de penser que la loterie est équitable, tous les tickets ont autant de chances l'un que l'autre de gagner.

t_i = le ticket i va gagner,
$\neg(\neg t_1 \wedge \neg t_2 \wedge ... \wedge \neg t_1 00)$ = un des tickets est gagnant,

Imaginons que $B\phi$ ssi $Pr(\phi) \geq 0,9$ (TL)

- $B\neg t_i$ pour tout i *(par TL)*
- $B\neg(\neg t_1 \wedge \neg t_2 \wedge ... \wedge \neg t_1 00)$ *(par TL)*
- $B(\neg t_1 \wedge \neg t_2 \wedge ... \wedge \neg t_1 00)$ *(par agrégation)*

Le paradoxe de la loterie

Les trois affirmations suivantes sont incompatibles :

- les croyances sont cohérentes,
- les croyances s'agrègent,
- la croyance correspond à un degré de probabilité subjective élevé mais inférieur à un,

Exercise I.A

p = John is tall
q = Bob is tall
It is not the case that John is tall and Bob is tall = $\neg(p \wedge q)$
(here we need parentheses, the english sentence is in fact ambiguous)

p	q	$p \wedge q$	$\neg(p \wedge q)$
1	1	1	0
1	0	0	1
0	1	0	1
0	0	0	1

Exercise I.A

p = Mary comes
q = John is happy
If Mary comes, then John is not happy = $p \rightarrow \neg q$

p	q	$\neg q$	$p \rightarrow \neg q$
1	1	0	0
1	0	1	1
0	1	0	1
0	0	1	1

Exercise I.A

p = John is tall
John is tall and John is not tall = $p \wedge \neg p$

p	$\neg p$	$p \wedge \neg p$
1	0	0
0	1	0

Exercise I.A

p = John is tall
It is not the case that John is tall and John is not tall = $\neg(p \wedge \neg p)$

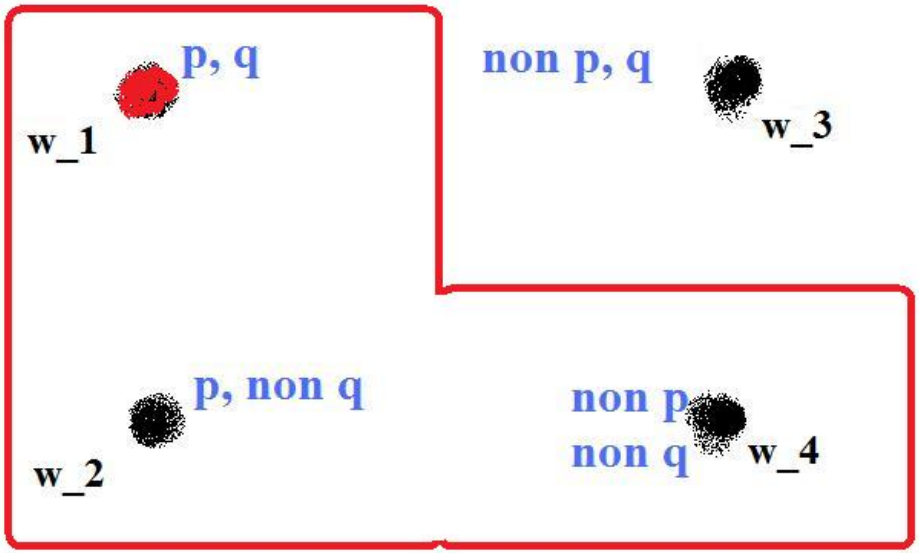
p	$\neg p$	$p \wedge \neg p$	$\neg(p \wedge \neg p)$
1	0	0	1
0	1	0	1

Exercise I.B

Three categories of sentences :

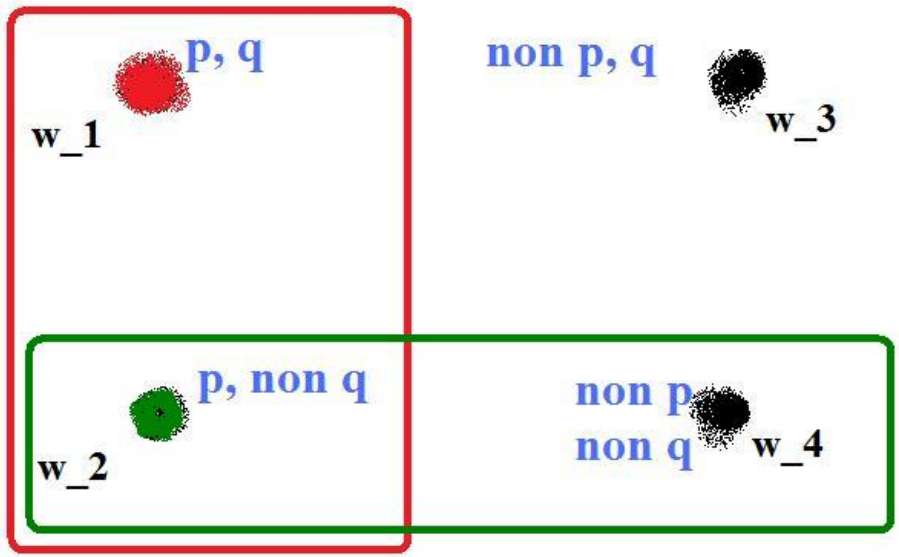
- *neutral sentences*
  sometimes true, sometimes false
  eg $\neg(p \wedge q)$
- *contradictions*
  always false
  $p \wedge \neg p$
- *tautologies*
  always true
  eg $\neg(p \wedge \neg p)$

Exercise II.A



- $w_1 \nVdash Bp$ (because $w_4 \in f(w_1)$)
- $w_1 \nVdash Bq$ (because $w_4 \in f(w_1)$)
- $w_1 \nVdash B(p \rightarrow q)$ (because $w_2 \in f(w_1)$)
- $w_1 \vDash B(q \rightarrow p)$

Exercise II.B



- $w_1 \models Bp$ (because $w_1 \models p$ and $w_2 \models p$)
- $w_2 \not\models Bp$ (because $w_4 \in f(w_2)$)
- $w_1 \not\models BBp$ (because $w_2 \in f(w_1)$)

Exercise II.C

$Bp \to BBp$ is true everywhere
iff for all w, w', if $w' \in f(w)$, then $f(w') = f(w)$.

- If for all w, w', if $w' \in f(w)$, then $f(w') = f(w)$, then $Bp \to BBp$ is true everywhere.

Assume $w \vDash Bp$. Consider $w' \in f(w)$. By definition of B, $w' \vDash p$. But $f(w') = f(w)$. Hence for all $w'' \in f(w')$, $w'' \vDash p$ since $f(w') = f(w)$. By definition of B, for all $w' \in f(w)$, $w' \vDash Bp$, hence by definition of B again $w \vDash BBp$.

- If $Bp \to BBp$ is true everywhere, then for all w, w', if $w' \in f(w)$, then $f(w') = f(w)$.

Assume that there are w, w' with $w' \in f(w)$ and $f(w') \neq f(w)$. Wlg, say that there is w'' with $w'' \in f(w')$ but $w'' \notin f(w)$.
Set p such that $z \vDash p$ for all $z \in f(w)$ but $w'' \nvDash p$. Check that $w \vDash Bp$ but $w \nvDash BBp$.

Exercise III.A

You write book, say, a cognitive science book. In this book, you make many assertions, each of which you can adequately defend. In particular, suppose that it is rational for you to have a degree of confidence x or greater in each of these propositions, where x is sufficient for belief but less than 1. Nonetheless, you admit in the preface that you are not so naïve as to think that your book contains no mistakes. You understand that any book as ambitious as yours is likely to contain at least a few errors.

Let p_i be the i^{th} sentence in the book. Following the scenario, Bp_i for each i. Hence, by aggregation, $B(p_1 \wedge ... \wedge p_n)$ where n is the total number of sentences in the book. However, according to what the preface says, $B\neg(p_1 \wedge ... \wedge p_n)$. Hence we have B_ϕ and $B\neg\phi$.

Exercise III.B

Let n be the total number of sentences, for $k \subseteq \{1, ..., n\}$ let w_k be the world in which sentences i with $i \in k$ are false. For each w_k, set $pr(w_k) = \frac{1}{2^n}$.
Check that $Pr(\neg p_i) = \frac{1}{2}$ and $Pr(\neg(p_1 \wedge ... \wedge p_n)) = 1 - \frac{1}{2^n}$.

III.D

Les croyances doivent correspondre à des probabilités subjectives élevées *et stables*

Definition (résistance)

La résistance $R(\phi)$ est le minimum des $Pr(\phi|\psi)$ où ψ est compatible avec ϕ.

Theorem (Skyrms, 1980)

L'ensemble des propositions dont la résistance est supérieur à n pour $n \geq 0,5$ *est cohérent et s'agrège.*