

# Le problème des sages et des chapeaux

## 1 Énoncé

Voici l'énoncé du problème des sages et des chapeaux<sup>1</sup>.

Trois sages sont dans une pièce, chacun portant sur la tête un chapeau noir ou blanc. Ils savent qu'un chapeau au moins est blanc. Au bout de quelques secondes, le premier sage dit : "Je ne connais pas la couleur de mon chapeau". Puis le second sage dit : "Je ne connais pas non plus la couleur de mon chapeau". Alors le troisième sage peut conclure : "Mon chapeau est blanc".

## 2 Formalisation à l'aide de la logique modale

La formalisation et la résolution du problème utilise la logique modale, dont les notions essentielles sont décrites dans l'encart p. 2.

On peut ainsi reformuler les données de l'énoncé à l'aide des assertions logiques suivantes :

- il y a au moins un chapeau blanc :

$$\Box(a)((\Diamond(x)\text{blanc}(y) \wedge x \neq y) \rightarrow (\text{blanc}(z) \wedge z \neq x \wedge z \neq y))$$

i.e. tout le monde sait que si le chapeau de  $x$  est noir, que le chapeau de  $y$  est noir et que  $x$  et  $y$  sont des personnes différentes, alors le chapeau de  $z$  sera blanc et  $z$  sera une personne différente de  $x$  et  $y$ .

- chaque sage voit le chapeau des deux autres :

$$\Box(a)((\Diamond(x)\text{blanc}(y) \wedge x \neq y) \rightarrow (\Box(x)\text{blanc}(y)))$$

$$\Box(a)((\Diamond(x)\text{noir}(y) \wedge x \neq y) \rightarrow (\Box(x)\text{noir}(y)))$$

i.e. tout le monde sait que, s'il est compatible avec les connaissances de  $x$  que le chapeau de  $y$  soit blanc, et que l'on a  $x \neq y$ , alors  $x$  sait que le chapeau de  $y$  est blanc. En effet,  $x$  voyant le chapeau de  $y$ , s'il est compatible avec ses connaissances que le chapeau est blanc, c'est qu'il l'est. Il en va de même pour la couleur noire (seconde formule).

- le premier sage ne connaît pas la couleur de son chapeau :

$$\Box(x)(\Box(y)(\Diamond(\text{sage1})\text{noir}(\text{sage1})))$$

---

<sup>1</sup>[1] Alliot, J.-M., Schiex, T., Brisset, P. et Garcia, F. (2002). *Intelligence Artificielle et Informatique Théorique*. Cépaduès (Ed.)

i.e. toute personne  $x$  sait que toute personne  $y$  qu'il est compatible avec les connaissances du "sage1" qu'il ait un chapeau noir. En effet, le premier sage a été incapable de répondre, donc il sait qu'il est possible qu'il ait un chapeau noir, et chacun sait que les autres le savent.

– le deuxième sage ne connaît pas la couleur de son chapeau :

$$\Box(x)(\Box(y)(\Diamond(\text{sage2})\text{noir}(\text{sage2})))$$

i.e. cf. supra

L'ensemble de ces assertions permettent alors de démontrer :

$$\Box(\text{sage3})\text{blanc}(\text{sage3}) \quad \blacksquare$$

La logique modale permet d'étendre la logique du premier ordre (calcul des prédicats) en incluant les notions de possibilité/nécessité et de connaissance/croyance. Dans le premier cas, on parle de logique *aléthique*, Dans le second, il s'agit d'une logique *épistémique*, ou logique de la connaissance, développée par Hintikka(1963). On introduit deux nouveaux opérateurs, dits modaux, qui dans ce cadre de la logique de la connaissance sont interprétés comme<sup>a</sup> :

1.  $\Box \equiv \text{savoir}$

2.  $\Diamond \equiv \text{compatible avec mes connaissances}$

Ces opérateurs permettent d'exprimer aisément les propositions suivantes ([1]) :

– Tout ce que sait Paul, Eric le sait.

$$\forall x, \Box(\text{Paul})x \rightarrow \Box(\text{Eric})x$$

– Si quelqu'un sait qu'il fait beau, alors il fait beau.

$$\exists x, \Box(x)\text{FaireBeau} \rightarrow \text{FaireBeau}$$

De manière générale, on retiendra les deux axiomes suivants :

–  $\Box A \rightarrow A$  ("si je sais que  $A$  est vrai, alors  $A$  est vrai")

–  $\Diamond A \rightarrow \Box \Diamond A$  (si  $A$  est compatible avec mes connaissances, alors je sais que  $A$  est compatible avec mes connaissances")

et la propriété suivante :

–  $\Box A \rightarrow \Box \Box A$

Ces deux dernières propriétés sont appelées, respectivement, propriété d'*introspection négative* et propriété d'*introspection positive*. La première propriété  $\Box A \rightarrow A$  justifie quant à elle l'appellation de logique de la connaissance.

<sup>a</sup>L'interprétation de ces opérateurs est différente en logique aléthique. Celle-ci a été développée par Lewis pour remédier à certains paradoxes de la logique classique, comme les propriétés  $\neg A \rightarrow (A \rightarrow B)$  ("si  $A$  est faux, je peux déduire n'importe quoi de  $A$ ") ou  $A \rightarrow (B \rightarrow A)$  ("si  $A$  est vrai alors  $A$  se déduit de n'importe quoi"). Lewis a alors proposé une nouvelle implication, dite implication stricte et notée  $>$ . L'opérateur  $\Diamond$  est alors défini comme :

$$(A > B) =_{def} \neg \Diamond(A \wedge \neg B)$$

qui se lit : "il est impossible que  $A$  soit vrai et  $B$  faux lorsque  $A$  implique strictement  $B$ ".

La notion duale de nécessité est notée  $\Box$  et se définit comme :

$$\Box A =_{def} \neg \Diamond \neg A$$

qui se lit : " $A$  est nécessaire si non- $A$  est impossible".

individu 1	●	●	$C$	(1)
	○	○	$\bar{C}$	(2)
	○	●	$\bar{C}$	(3)
	●	○	$\bar{C}$	(4)
<hr style="border: 0.5px solid black;"/>				
individu 2		●	$C$	(5)
		○	$\bar{C}$	(6)
<hr style="border: 0.5px solid black;"/>				
individu 3			$C$	(7)

FIG. 1 – Illustration du degré de connaissance pour les 3 individus ( $C$  "je connais ma couleur",  $\bar{C}$  "je ne connais pas ma couleur"), en fonction des informations disponibles (ce qu'il voit dans le dos de la (les) personne(s) qui est (sont) devant lui).

### 3 Variante : le jeu du roi et des prisonniers

#### 3.1 Principe

Un roi propose à 3 prisonniers à vie de participer à un jeu à l'issue duquel ils peuvent être libérés s'ils trouvent la bonne réponse au problème suivant. Chacun des prisonniers est assis sur une des 3 chaises, disposées l'une derrière l'autre, de sorte que celui qui est assis en queue voit le dos des deux autres, tandis que celui qui est en position médiane ne voit que le dos de la personne qui est devant lui. L'individu en tête ne voit quant à lui personne. Trois cartons ont été tirés au hasard parmi un ensemble comprenant 3 cartons blancs et 2 cartons noirs, et ont été attachés au dos de chacun des prisonniers. Les prisonniers sont interrogés dans l'ordre, en commençant par celui qui est en queue, puis celui qui est au milieu, enfin celui qui est en tête. Ils doivent indiquer verbalement s'ils connaissent ou ne connaissent pas la couleur qui est dans leur dos, sans la mentionner.

#### 3.2 Solution

On peut schématiser la situation comme suit. En désignant par un disque noir et un disque blanc les indices vus par les individus, les informations de chacun sont résumées dans la figure 1, où les disques indiquent ce qui est vu par chacun des protagonistes. Les réponses "logiques" de chacun sont indiqués par les lettres  $C$  ("connaît") et  $\bar{C}$  ("ne connaît pas").

Le premier individu voit le dos de ses deux compagnons, ceux-ci pouvant présenter soit 2 disques blancs, soit 1 disque de chaque couleur (avec les 2 ordres possibles). Le deuxième individu ne voit quant à lui que le dos de la personne qui est devant lui, qui peut présenter un disque blanc ou noir. Si la première personne voit 2 disques noirs (cas 1), alors elle connaît nécessairement sa couleur (blanc), et les deux autres individus connaissent également, lorsqu'elle répond "je connais ma couleur", la couleur qui est dans leur dos (noir). Si la première personne répond "je ne connais pas ma couleur",

1	○	●	○	●	○	○	●
2	●	○	●	●	○	○	○
3	●	●	○	○	○	●	○

FIG. 2 – Ensemble des solutions possibles pour  $n = 3$  et  $k = 2$ .

alors on est dans le cas où elle ne voit pas 2 disques noirs, et il y a 3 cas possibles (cas 2, 3 et 4). Si la deuxième personne voit un disque noir (cas 5, découlant du cas 3), alors elle sait qu'elle porte la couleur blanche puisqu'autrement la première personne n'aurait pas répondu "je ne connais pas ma couleur". Elle répondra donc "je connais ma couleur", sauf si elle voit un disque blanc (cas 6, découlant des cas 2 et 4). La troisième personne connaît ainsi toujours sa couleur (cas 7) : si la première et la deuxième personne ont répondu "je ne connais pas ma couleur", c'est qu'il porte du blanc, et si la première ou la deuxième personne (le ou est ici exclusif) ont répondu "je connais ma couleur" (cas 1 ou 5), alors il porte du noir.

L'ensemble des situations possibles est résumé dans la figure 2. La difficulté dans cet exercice provient principalement de la nécessité d'adopter un raisonnement par l'absurde, c'est-à-dire de raisonner sur des hypothèses amenant à des contradictions, et qui après élimination, ou *exclusion*, permettent d'en déduire la solution unique.

Le problème se complique sensiblement si on augmente le nombre  $k$  de couleurs, ou le nombre  $n$  de joueurs. En effet, même si la méthode de résolution demeure identique, le nombre d'informations à retenir (i.e. les cas hypothétiques déduits des réponses précédentes) devient rapidement très important, ce qui a pour conséquence une saturation de la mémoire de travail.

### 3.3 Remarques

Il n'y a pas lieu ici de tenir de raisonnement probabiliste (malgré la notion de tirage au hasard des cartons). Le hasard n'intervient pas en effet dans les informations acquises par les participants. Par ailleurs, les individus n'ont aucun intérêt à mentir (sinon ils ne seraient pas libérés), contrairement à certains autres jeux (e.g. dilemme des prisonniers) ; par conséquent, les informations données par les deux premières personnes sont censées être correctes, à moins d'une faute de raisonnement.

Dans tous les cas, le dernier à parler connaît nécessairement sa couleur. Le premier ne connaît sa couleur que lorsque ses deux partenaires portent un carton noir ; dans tous les autres cas, il ne la connaît pas.

## 4 Encore des chapeaux et des couleurs

Voici le problème<sup>2</sup> :

<sup>2</sup>[2] [www.dma.ens.fr/benaych/benaych.chapeaux.pdf](http://www.dma.ens.fr/benaych/benaych.chapeaux.pdf)

100 personnes sont disposées en ronde (de façon à ce que chacun voie tous les autres), chacune ayant un chapeau, qui est soit jaune, soit bleu. Personne ne peut voir la couleur de son chapeau, mais tout le monde voit les chapeaux des autres. Le problème est de donner une stratégie qui permettra au plus grand nombre possible de personnes de connaître avec certitude la couleur de leur chapeau, les contraintes étant les suivantes :

- une personne est désignée pour parler en premier, puis c'est le tour de la personne située à sa gauche, puis la personne suivante à gauche, et ainsi de suite . . .
- chaque personne dit "jaune" ou "bleu", et aucune autre information n'est communiquée

Attention ! On demande de trouver la stratégie qui permette *au plus grand nombre* de personnes de trouver la couleur de leur chapeau, et pas une stratégie donnant le nombre moyen de personnes pouvant répondre correctement.

En fait, tous sauf un (eh oui, encore le premier !) peuvent connaître la couleur de leur chapeau. Pour cela, il faut que le premier compte le nombre de chapeaux bleus qu'il voit. Si ce nombre est pair, il annonce "jaune", et le cas échéant, il dit "bleu". Ensuite, chacun détermine la couleur de son chapeau en comptant le nombre de chapeaux bleus portés par les autres, sans compter celui a parlé en premier (ni se compter soi-même évidemment). Si ce nombre a la même parité que celle énoncée par celui a parlé en premier, alors son chapeau est jaune, sinon c'est qu'il est bleu.