

# L'esprit, 4

## La conscience, le fonctionnalisme et le physicalisme

### Philosophie des SCHS

### séance 4

M. Cozic

# 1. La conscience

# le mystère de la conscience

- ▷ D. Dennett, *Consciousness Explained*, 1991, trad.fr. P. Engel, *La conscience expliquée*, O. Jacob, Paris, 1993

“La conscience humaine est peut-être le dernier mystère qui reste. Un mystère est un phénomène dont on ne sait quoi penser - dont on ne sait pas encore quoi penser. D'autres grands mystères ont existé : celui de l'origine de l'univers, celui de la vie et de la reproduction, celui du dessein qu'on peut lire dans la nature, celui du temps, de l'espace et de la gravitation. Ce n'étaient pas seulement des zones d'ignorance scientifique ; c'étaient des sujets de perplexité extrême et d'étonnement.”

# le mystère de la conscience

- ▷ D. Dennett, *Consciousness Explained*, 1991, trad.fr. P. Engel, *La conscience expliquée*, O. Jacob, Paris, 1993

“ Nous n’avons toujours pas de réponse ultime aux questions que posent la cosmologie et la physique des particules, la génétique moléculaires et la théorie de l’évolution, mais nous savons au moins les aborder. Les mystères ne se sont pas évanouis, mais on les a apprivoisés. Mais avec la conscience, nous sommes toujours terriblement embourbés. La question de la nature de la conscience est aujourd’hui le seul sujet qui laisse souvent perplexes et muets les penseurs les plus subtils.”

## 2. fonctionnalisme et qualia

## le scénario de Block

- (1) un **corps artificiel** identique en apparence au nôtre, avec des organes sensoriels et des organes moteurs. Ce corps artificiel a des neurones sensoriels qui proviennent des organes sensoriels et des neurones moteurs qui vont vers les organes moteurs
- (2) chaque chinois dispose d'un émetteur/récepteur
- (3) un ensemble de satellites est capables de représenter des symboles dans le ciel, symboles que tous les Chinois peuvent voir
- (4) chaque chinois doit suivre *une* instruction qui est déclenchée par la réception d'un certain input neuronal et d'un certain symbole écrit par satellite ; en retour, il ordonne que les satellites représentent un nouveau symbole et déclenche un certain *output* neuronal

# le scénario de Block



© Jolyon Trusciano 2005 www.jolyon.cs.uk

## doubles fonctionnels

- ▶ Supposons qu'il soit possible de faire en sorte que le système  $S$  formé par (a) le peuple chinois, (b) le corps artificiel et (c) les satellites aient globalement la même organisation fonctionnelle (réalise la même machine de Turing) qu'une certaine personne, Pierre.
- ▶  $S$  est alors un **double fonctionnel** de Pierre.
- ▶ si le fonctionnalisme est vrai, alors toute la psychologie de Pierre est capturée par son organisation fonctionnelle:  
**Pierre et son double fonctionnel  $S$  ont la même psychologie** du point de vue fonctionnaliste.



## un contre-exemple au fonctionnalisme ?

- ▶ pour Block, la situation constitue un **contre-exemple** au fonctionnalisme :
- (P1) si le fonctionnalisme est vrai, Pierre et *S* ont la même psychologie (et notamment les mêmes états mentaux)
- (P2) (il est vraisemblable que) Pierre et *S* n'ont pas la même psychologie
- ▶ (par contraposition) le fonctionnalisme est faux

## un contre-exemple au fonctionnalisme ?

- ▶ (P1) ne pose pas de difficulté; tout l'argument repose sur (P2)
- ▶ L'argument de Block est destiné à montrer que le fonctionnalisme est **trop libéral** : il accorde la mentalité à des systèmes qui, selon lui, en sont dépourvus
- ▶ on montrerait (P2) si par exemple on montrait que *S* n'a pas de croyances ou n'a pas de désirs, bref n'a pas d'attitudes propositionnelles. Ce n'est pas quelque chose d'évident.

## ce qui manque à *S* selon Block

- ▷ N. Block, "Troubles with Functionalism", 1978

“Ce qui fait du système avec une tête d’homoncules qui vient d’être décrit un contre-exemple prima facie au fonctionnalisme (des machines) et qu’il existe un doute prima facie quant à la question de savoir s’il a un quelconque état mental - en particulier s’il a ce que les philosophes appellent des “états qualitatifs”, des “sentiments bruts” ou des “qualités phénoménologiques immédiates”.”

## ce qui manque à *S* selon Block, suite

- ▷ N. Block, "Troubles with Functionalism", 1978

“Je ne prétends pas qu’aucune sorte de chose qui vous simule fonctionnellement ne peut avoir de qualia...  
J’affirme plutôt que toute sorte de chose qui vous simule fonctionnellement avec une tête à homoncules n’a pas nécessairement des qualia. S’il existe ne serait-ce qu’une chose qui puisse vous simuler fonctionnellement sans avoir de qualia, le fonctionnalisme est faux.”

## les qualia (absent ou non)

- ▶ l'idée est donc qu'un double fonctionnel comme le système *S* ne semble pas avoir de **qualia**. D'où l'expression **argument des qualia absents**. Qu'est-ce que les qualia ?
- ▷ Shoemaker, S. (1982) "The Inverted Spectrum", *The Journal of Philosophy*, vol.79, pp. 357-81  
"les propriétés qualitatives ou phénoménales des expériences sensorielles, en vertu desquelles elles se ressemblent et diffèrent les unes par rapport aux autres, qualitativement, de la façon dont elles le font"

# les qualia



Figure: De Byrne (2006)

- ▶ l'**effet** que fait la vision du carré rouge diffère de l'effet que fait la vision du carré mauve qui diffère de l'effet que fait la vision du carré turquoise
- ▶ si un second carré rouge était ajouté, on dirait que l'effet que fait le premier carré rouge est **semblable** à celui que fait le second carré rouge
- ▶ on dirait aussi probablement que l'effet que fait le carré rouge est **plus semblable** à l'effet que fait le carré mauve qu'à l'effet que fait le carré turquoise

# les qualia



Figure: De Byrne (2006)

- ▶ les qualia désignent ces effets en question. Ce sont des propriétés d'états mentaux : la vision du carré rouge a le qualia  $Q_r$ , celle du carré mauve le qualia  $Q_m$  et celle du carré turquoise le qualia  $Q_t$ .

## retour à l'argument

- ▶ considérons la douleur. Ce que Block soutient, c'est que lorsque  $S$  est dans l'état fonctionnel qui correspond à la douleur -  $f_d$  - il ne ressent probablement pas ce que nous ressentons de manière caractéristique quand nous ressentons de la douleur.
- ▶ un événement qui cause chez nous de la douleur causera  $f_d$  chez  $S$  ; le même genre de réaction que la douleur cause chez nous sera causé chez  $S$  par  $f_d$ .
- ▶ mais il manquera vraisemblablement à  $S$  l'aspect qualitatif caractéristique de la douleur.



## réponses à l'argument des qualia absents

- ▶ 2 réponses radicales à cette façon de soutenir (P2) (donc à cette façon de soutenir l'argument anti-fonctionnaliste des qualia absents) :
  - (i) **libéralisme des qualia**: malgré les apparences, un système comme *S* a des qualia.
  - (ii) **scepticisme des qualia** : il n'existe rien de tel que les *qualia*

### 3. physicalisme et qualia

## 3.1 la chambre de Marie

## le scénario

- ▶ **argument antiphysicaliste** proposé par F. Jackson en 1982 dans "Epiphenomenal Qualia"
- Marie est une jeune femme qui a toujours vécu dans une pièce très particulière (la "pièce de Jackson"): tout y est en noir et blanc. Ses seules communications visuelles avec le monde extérieur se font par l'intermédiaire d'un écran de télévision en noir et blanc.
- Marie trouve des compensations intellectuelles : elle étudie la neurophysiologie de la vision au point de maîtriser **toutes les connaissances physiques** (ou "informations physiques") relatives à la vision. Pas seulement les connaissances dont nous disposons aujourd'hui, mais toutes les connaissances physiques pertinentes possibles . Ces connaissances incluent la connaissance des différentes relations fonctionnelles.

## le scénario, suite

- Le jour de son quarantième anniversaire, le jour J, Marie est libérée de la pièce de Jackson, et à l'instant de sa sortie, on lui montre une tomate bien mûre.



Figure by MIT OCW.

black and white  
Mary



Figure by MIT OCW.

# l'argument

- ▶ **L'argument** : le jour J, quand Marie voit la tomate qu'on lui présente, **elle apprend quelque chose** : elle apprend l'effet que cela fait de voir du rouge. Par conséquent, ses connaissances préalables sur la vision sont incomplètes. Les connaissances physiques sur la vision n'épuisent donc pas les connaissances sur la vision. Donc le physicalisme est faux.
- ▶ Décomposition de l'argument
  - (P1) Avant le jour J, Marie une connaissance physique complète de la vision
  - (P2) Le jour J, Marie apprend quelque chose de nouveau
  - (C1) Toutes les connaissances pertinentes pour la vision ne sont pas physiques
  - (C2) Le physicalisme est faux

## la leçon qu'en tire Jackson

- ▶ Jackson (1982) adopte une position **épiphénoménaliste** à l'égard des **qualia** : les **qualia** ne sont pas des propriétés physiques, sont (au moins parfois) causés par des événements physiques mais ne sont pas causalement efficaces vis-à-vis du monde physique.
- ▶ c'est une façon de concilier (1) le principe de complétude du monde physique et (2) l'argument de la connaissance.

## réponses physicalistes

- ▶ réponse #1 : l'hypothèse de capacité (Lewis)
- accepte que Marie apprend quelque chose, mais à rejette l'hypothèse selon laquelle Marie apprend une nouvelle information, ou encore qu'elle acquiert une nouvelle connaissance.
- Lewis ne soutient pas pour autant que Marie n'apprend rien : ce qu'elle acquiert, ce n'est pas un **savoir que** (*knowing that*) mais un **savoir faire** (*knowing how*) ou encore une capacité (*ability*).
- le type de savoir que Marie acquiert est le type de savoir qu'un enfant acquiert quand il apprend à faire du vélo, qu'un néophyte en oenologie apprend à reconnaître un cépage, etc. Pour cette raison, Lewis baptise son hypothèse *l'hypothèse de capacité*.



# réponses physicalistes

- ▶ réponse #1 : l'hypothèse de capacité (Lewis)
- ▶ Les capacités que Marie acquiert sont les suivantes:
  - capacité de se souvenir de l'expérience du rouge
  - capacité d'imaginer l'expérience du rouge
  - capacité de reconnaître l'expérience du rouge

## réponse # 2

- ▶ réponse # 2 : de nouvelles connaissances d'anciens faits
- une bonne partie des philosophes physicalistes ont soutenu que l'on pouvait maintenir le physicalisme et reconnaître que Marie acquiert d'authentiques nouvelles connaissances (pas seulement des savoirs-faire).
- Idée générale : Marie apprend quelque chose qu'en un certain sens elle connaissait déjà, mais qu'elle possède maintenant sous une nouvelle forme, à l'aide de **nouveaux concepts**
- Marie s'approprie avec de nouveaux concepts des faits qu'elle connaît déjà sous une autre conceptualisation.

## réponse # 2

- ▶ réponse # 2 : de nouvelles connaissances d'anciens faits
- ces nouveaux concepts dont Marie dispose sont souvent appelés des **concepts phénoménaux**.
- les *concepts* phénoménaux sont des concepts recognitionels qui dénotent des *propriétés* physico-fonctionnelles
- les *concepts* phénoménaux sont indépendants des *concepts* physiques

## réponse #3

- ▶ réponse # 3 : "il doit y avoir une réponse !"
- La réponse que privilégient Braddon-Mitchell et Jackson (1996) met en avant une sorte d'incompatibilité entre certaines considérations épistémologiques très générales et la conclusion de l'argument.
- Plus précisément, elle fait valoir qu'une épistémologie épiphénoménaliste n'est pas tenable
- idée de base: il doit y avoir une origine causale à tout changement des croyances d'un individu. Il doit donc y avoir une origine causale aux changements dans les croyances de Marie quand elle voit son premier objet "en couleur". Mais si les *qualia* sont épiphénoménaux, alors ils ne peuvent intervenir causalement dans l'apparition des croyances à propos des *qualia*.

## 3.2 l'argument des zombies



# reconstruction de l'argument

► l'argument:

(Def) un zombie = un être qui est physiquement identique à nous mais qui n'a pas d'expériences conscientes (d'états mentaux P-conscients)

(P1) un zombie est concevable

(C1) un zombie est possible

(C2) les phénomènes mentaux ne **surviennent** pas sur les phénomènes physiques

(C3) le physicalisme est faux

## qu'est-ce que la survenance ?

- ▶ (C2) les phénomènes mentaux ne **surviennent** pas sur les phénomènes physiques
- ▶ survenance du mental sur le physique:  
(S) deux choses qui sont identiques du point de vue physique sont identiques du point de vue mental  
(S') si deux choses diffèrent du point de vue mental, alors elles diffèrent du point de vue physique
- ▶ pas de différence mentale sans différence physique
- ▶ Chalmers (1996): “une fois que Dieu a déterminé les faits physiques du monde, il a tout déterminé”



# qu'est-ce que la survenance ?

- ▶ exemple: la propriété d'être chauve **survient** sur la distribution des cheveux
- si Pierre et Jean ont la même distribution de cheveux, alors ou bien ils sont tous les deux chauves, ou bien aucun des deux n'est chauve
- si l'un est chauve et l'autre pas, alors ils n'ont pas la même distribution de cheveux

## survenance et physicalisme

- ▶ on considère en général que la survenance du mental sur le physique est une exigence physicaliste minimale
- ▶ pour le dire autrement: si le physicalisme est vrai, alors la thèse de survenance (du mental sur le physique) est vraie
- ▶ ou encore: si la thèse de survenance est fausse, alors le physicalisme est faux
- ▶ c'est cette implication qui est utilisée dans l'argument des zombies

# réponses physicalistes

- ▶ deux façons principales de contrer l'argument des zombies:
- ▶ réponse #1: rejeter le passage de la caractérisation des zombies (Def) à leur concevabilité (P1)
- ▶ réponse #2: rejeter le passage de la concevabilité des zombies (P1) à leur possibilité (C1)

## 4. conclusion: quelle science de la conscience?

▷ Chalmers (2002)

“...pourquoi et comment les processus physiques donnent naissance à l’expérience ? Pourquoi ces processus n’ont pas lieu “dans le noir”, sans expérience qui les accompagne ? C’est le mystère central de la conscience.”

## structure des arguments anti-physicalistes

- ▶ quelques éléments de conclusion à partir de D. Chalmers (2002), “Consciousness and Its Place in Nature”
  - ▶ les arguments antiphysicalistes basés sur les *qualia* ou sur l’aspect phénoménal de la conscience ont la structure suivante:
    - (1) Il y a un **fossé épistémique** entre les vérités physiques et les vérités phénoménales
    - (2) S’il y a un fossé épistémique entre les vérités physiques et les vérités phénoménales, il y a un **fossé ontologique** entre les deux et le physicalisme est faux

---
- (C) Le physicalisme est faux

## commentaires

- (1) Il y a un fossé épistémique entre les vérités physiques et les vérités phénoménales
- ▶ fossé épistémique: nos connaissances des faits physiques (et fonctionnels) ne sont pas suffisantes pour nous faire connaître (à elles seules) les faits phénoménaux = les faits relatifs aux états mentaux conscients (l'effet que cela fait de voir telle chose, d'entendre telle autre, etc.)
  - ▶ exemple #1: Marie apprend quelque chose quand elle sort de la chambre
  - ▶ exemple #2: les zombies sont concevables
  - ▶ l'étape cruciale des arguments anti-physicalistes consiste à passer de ce fossé épistémique à **fossé ontologique**: les propriétés phénoménales ne sont pas des propriétés physiques (ni n'ont trait à l'organisation fonctionnelle des propriétés physiques)

# réponses

- ▶ les physicalistes peuvent
- (A) nier qu'il existe véritablement un fossé épistémique: en fait, les zombies ne sont pas concevables ; en fait, Marie n'apprend rien de véritablement nouveau quand elle sort de sa Chambre
- Typiquement, les tenants de cette position pensent que les aspects fonctionnels de la conscience épuisent ses propriétés. Par exemple, avoir un état mental conscient permet de l'exprimer, permet d'en faire usage pour différentes tâches cognitives, etc. Expliquer les fonctions de la conscience, c'est l'expliquer ("tout court").
- objection: (A) va à l'encontre de nos intuitions.



- ▶ les physicalistes peuvent
- (B) accepter l'existence d'un fossé épistémique, mais nier que l'on puisse en inférer un fossé ontologique: les zombies sont concevables, mais pas métaphysiquement possibles ; Marie n'apprend pas de nouveaux faits, mais des anciens faits d'une nouvelle manière.

Typiquement, les tenants de cette position pensent que les états phénoménaux s'identifient a posteriori (voir eau et  $H_2O$ ) à des états physiques ou fonctionnels.

objection: le fossé épistémique entre le phénoménal et le physico-fonctionnel semble différent du fossé phénoménal entre eau et  $H_2O$

- ▶ les physicalistes peuvent
- (C) accepter l'existence d'un fossé épistémique, mais soutenir qu'en principe il peut être surmonté. Nos limitations conceptuelles actuelles ne nous permettent pas de comprendre comment la conscience pourrait être un processus physique, mais peut-être dans le futur nous pourrions le comprendre, ou peut-être des êtres plus sophistiqués cognitivement peuvent le comprendre.  
objection: (C) est instable et tend à se confondre avec les positions précédentes ou suivantes

- ▶ si l'on accepte ce genre d'arguments, on peut
- (D) reconnaître que les propriétés phénoménales sont à part, ontologiquement, des propriétés physiques et qu'elles interagissent avec elles (dualisme interactionniste).  
objection: (D) viole le principe de complétude de la physique
- (E) reconnaître que les propriétés phénoménales sont à part, ontologiquement, des propriétés physiques et qu'elles peuvent être affectées par les propriétés physiques, mais pas réciproquement (dualisme épiphénoménaliste).  
objection: (E) heurte nos intuitions sur l'efficacité causale des états mentaux conscients.