

# Esprits Numeriques

Séance 3

Jonathan Simon

# Apercu

- 0) Clarification : computationalisme vs fonctionnalisme
- 1) Bostrom: les subtilités de l'implementation (de la semaine passée)
- 2) Putnam: L'argument de la réalisabilité multiple
- 3) Chalmers: L'argument de l'invariance organisationnelle (qualia absents, évanescents ou dansants)

# Computationalisme vs Fonctionnalisme

# Computationalisme vs Fonctionnalisme

- Fonctionnalisme: La vue métaphysique selon laquelle les états mentaux sont individualisés par leurs relations fonctionnelles avec les entrées et sorties mentales et les autres états mentaux.
- Computationalisme: l'hypothèse selon laquelle les relations fonctionnelles entre les entrées et sorties mentales et les états internes sont informatiques / computationnelle.

# Computationalisme vs Fonctionnalisme

- À proprement parler, l'hypothèse selon laquelle les états mentaux ne sont que des algorithmes pourrait être appelée «fonctionnalisme computationnel» - cependant, ces termes ne sont souvent pas soigneusement distingués dans la littérature - c'est juste un point à connaître.

Bostrom

# Bostrom

- La dernière fois, nous avons abordé plusieurs questions difficiles sur l'implémentation :
- 1) il ne suffit pas qu'un système se comporte temporairement d'une manière qui peut être décrite comme l'implémentation d'une machine de Turing donnée ( un rocher peut le faire) :
  - nous devons supposer que les mécanismes causaux / contrefactuels sous-jacents correspondent à la structure algorithmique

# Bostrom

- La dernière fois, nous avons abordé plusieurs questions difficiles sur l'implémentation :
- 2) Mais nous devons aussi admettre que les systèmes n'ont pas besoin de s'adapter «parfaitement» à la structure de l'algorithme à tout moment :
  - les systèmes peuvent être défaillants ou peu fiables mais «compter» quand même comme des implémentations d'un algorithme donné.

X39.81



# Bostrom

- La dernière fois, nous avons abordé plusieurs questions difficiles sur l'implémentation :
- 3) Bostrom observe que ces contraintes (en particulier celle qui autorise un certain manque de fiabilité) peuvent donner lieu à des cas limites délicats

# Bostrom

- Nous pouvons imaginer une série de cas, chacun légèrement moins fiable que le précédent : mais chacun (par coïncidence) continuant à agir en accord avec l'algorithme : lequel est le premier de la série à ne pas compter comme mettant en œuvre l'algorithme ?

# Bostrom

- Nous pouvons également noter que dans certains cas, le fait qu'il y ait un système ou deux peut dépendre de la fiabilité des circuits qui les relient : Bostrom suggère que cela signifie qu'il y a parfois un nombre fractionnaire d'expérimentateurs.

# Bostrom

- Ce sont des questions intéressantes : la morale générale pour nous est que lorsque nous parlons d'IA ou de robots conscients, nous devons nous rappeler que nous parlons toujours de systèmes physiques réels et concrets qui admettent une certaine description abstraite - et il peut être philosophiquement difficile de dire si cette description s'applique vraiment, dans un cas donné.

Putnam

# Putnam

- Maintenant : arguments selon lesquels les robots / IA peuvent être conscients
- À la lumière de ce que nous venons de voir, note que cette affirmation est vague. On pourrait peut-être décrire correctement un humain comme mettant en œuvre un algorithme "IA"... mais les autres choses mettant en œuvre cet algorithme ne sont pas conscientes.

# Putnam

- Notre question est la suivante : qu'est-ce qui est *nécessaire et suffisant* pour la conscience (par exemple pour ressentir de la douleur) ?
- ... n'oublie pas que nous pouvons prendre les termes «nécessaire» et «suffisant» dans le sens de «nécessaire et suffisant compte tenu des lois psychophysiques » (l'approche dualiste) ou «nécessaire et suffisant compte tenu des lois métaphysiques » (l'approche matérialiste)...

# Putnam

- Putnam a été l'un des premiers à proposer une conception explicitement fonctionnaliste des états conscients (en réponse aux alternatives behavioristes et physicalistes)
- Contre les béhavioristes : son argument le plus célèbre (dans un autre article) est l'argument des Superspartans : des agents qui ne manifestent pas de comportement douloureux et qui, en fait, s'entraînent à ne plus avoir de dispositions pour le comportement douloureux.
  - (une version de l'argument selon lequel les conséquences comportementales des états mentaux, y compris les états phénoménaux comme la douleur, sont interconnectées de manière holistique).

# Putnam

- Contre les physicalistes :
- Ici, les adversaires de Putnam sont ceux qui insistent sur le fait que le substrat compte, que ce n'est qu'un type spécifique d'état neuronal-physique-chimique qui peut provoquer l'expérience de la douleur....

# Putnam

- «Consider what the brain-state theorist has to do to make good his claims. He has to specify a physical-chemical state such that any organism (not just a mammal) is in pain if and only if (a) it possesses a brain of a suitable physical-chemical structure; and (b) its brain is in that physical-chemical state. This means that the physical-chemical state in question must be a possible state of a mammalian brain, a reptilian brain, a mollusc's brain (octopuses are mollusca, and certainly feel pain), etc. At the same time, it must not be a possible (physically possible) state of the brain of any physically possible creature that cannot feel pain. Even if such a state can be found, it must be nomologically certain that it will also be a state of the brain of any extra-terrestrial life that may be found that will be capable of feeling pain before we can even entertain the supposition that it may be pain.

# Putnam

- It is not altogether impossible that such a state will be found. Even though octopus and mammal are examples of parallel (rather than sequential) evolution, for example, virtually identical structures (physically speaking) have evolved in the eye of the octopus and in the eye of the mammal, notwithstanding the fact that this organ has evolved from different kinds of cells in the two cases. Thus it is at least possible that parallel evolution, all over the universe, might always lead to one and the same physical "correlate" of pain. But this is certainly an ambitious hypothesis.

# Putnam

- Finally, the hypothesis becomes still more ambitious when we realize that the brain-state theorist is not just saying that pain is a brain state; he is, of course, concerned to maintain that every psychological state is a brain state. Thus if we can find even one psychological predicate which can clearly be applied to both a mammal and an octopus (say "hungry"), but whose physical-chemical "correlate" is different in the two cases, the brain-state theory has collapsed. It seems to me overwhelmingly probable that we can do this. Granted, in such a case the brain-state theorist can save himself by ad hoc assumptions (e.g., defining the disjunction of two states to be a single "physical-chemical state"), but this does not have to be taken seriously. »

# Putnam

- 1) Nous pouvons trouver au moins un prédicat psychologique qui peut clairement être appliqué à la fois à un mammifère et à une pieuvre (disons «faim»), mais dont le «corrélat» physico-chimique est différent dans les deux cas.
- 2) Si 1), la théorie de l'état-cerveau est fausse
- 3) Par conséquent, la théorie de l'état-cerveau est fausse

# Putnam

- Comment justifier 1): ??
- 1) Nous pouvons trouver au moins un prédicat psychologique qui peut clairement être appliqué à la fois à un mammifère et à une pieuvre (disons «faim»), mais dont le «corrélat» physico-chimique est différent dans les deux cas.

# Putnam

- Comment justifier 1): ??
- En particulier, comment savons-nous que le même prédicat s'applique à la pieuvre?

# Putnam

- Putnam: «we identify organisms as in pain, or hungry, or angry, or in heat, etc., on the basis of their behavior...»
- (Les similitudes comportementales sont la preuve de similitudes mentales. (Mais le théoricien de l'état du cerveau ne peut-il pas l'accuser de contourner la question ici ?)

Chalmers

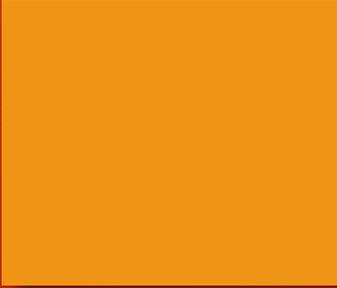
# Chalmers

- 1) Chalmers est un dualiste des propriétés : il soutient que des propriétés comme «avoir mal» sont irréductibles à des propriétés physiques ou fonctionnelles - et en particulier qu'il existe des mondes possibles où il y a de parfaits doubles physiques et fonctionnels de nous, qui n'instancient pas ces propriétés (ne sont pas conscients) - zombies.... !

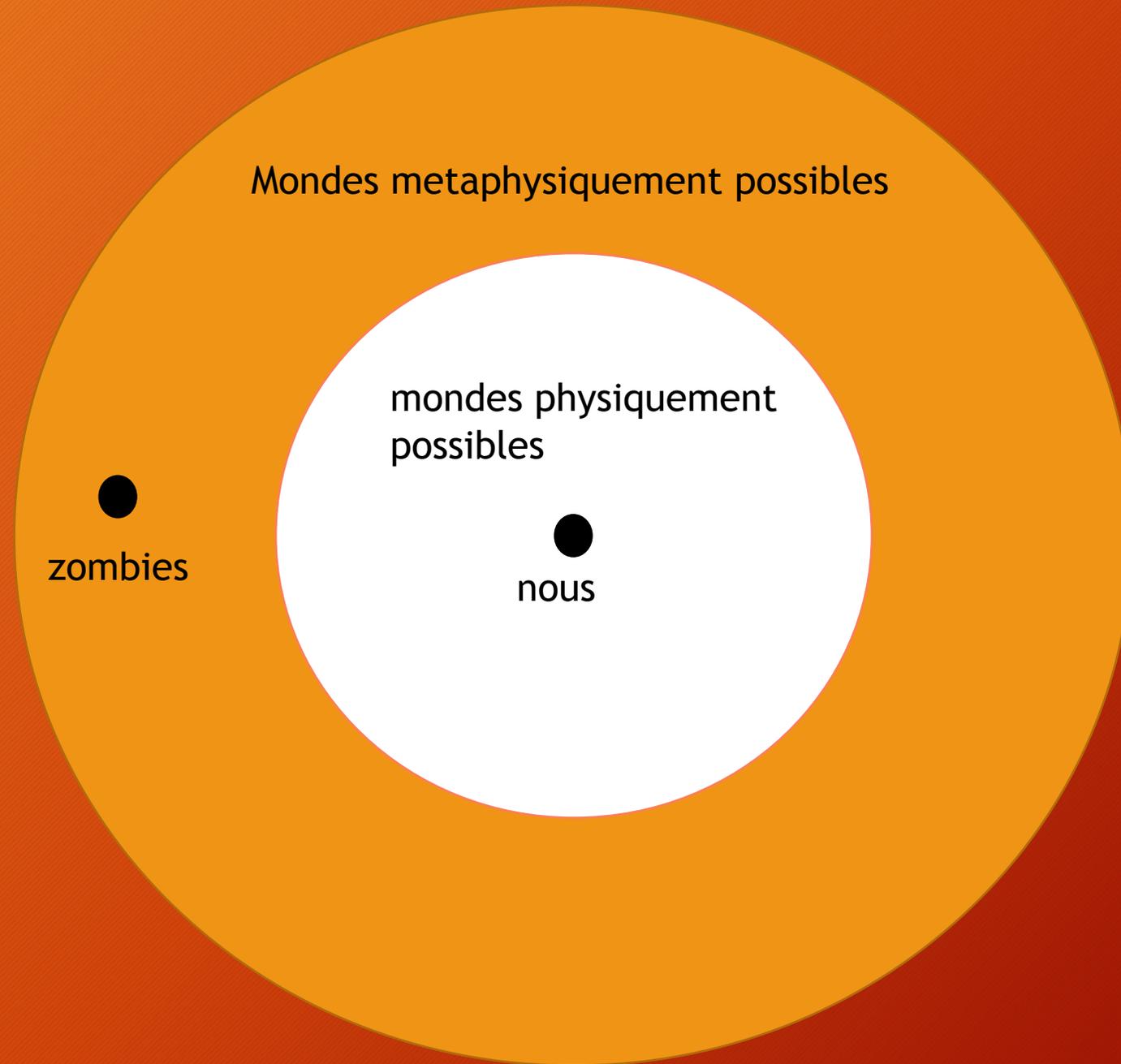
Mondes metaphysiquement possibles

mondes physiquement  
possibles

●  
nous



Dualistes:



Matérialistes:

●  
zombies

Mondes metaphysiquement possibles

mondes physiquement  
possibles

●  
nous



# Chalmers

- En physique (et dans d'autres sciences), on cherche des lois. Ces lois énoncent quelque chose de plus que la façon dont les choses se sont accidentellement produites dans les faits, mais quelque chose de moins que la façon dont les choses doivent être métaphysiques : elles énoncent la façon dont les choses doivent être dans les mondes naturellement possibles.

# Chalmers

- Beaucoup considèrent que les vertus théoriques telles que l'élégance et la simplicité sont des guides vers la vérité : face à deux théories candidates qui expliquent les faits, nous devrions préférer celle qui est la plus simple

# Chalmers

- Pour les dualistes, les principes qui relient les propriétés phénoménales (qualia) aux propriétés physiques vont être des lois (car il n'y a pas de connexions métaphysiquement nécessaires, ou du moins pas beaucoup).

# Chalmers

- Les arguments de Chalmers visent à montrer que tout ensemble de lois (psychophysiques) qui nie l'invariance organisationnelle, sera moins simple que les lois que l'on peut avoir si on la préserve (où OI est la thèse selon laquelle la duplication fonctionnelle signifie la duplication phénoménale, dans les mondes naturellement possibles).

# Chalmers

- Son argument (pour le fonctionnalisme) est intéressant en partie parce qu'il ne nécessite pas d'identité ou de réduction : au lieu de cela, il soutient que les lois psychophysiques (reliant les domaines distincts du physique et du mental) sont écrites au niveau des fonctions (computationnelles)

# Chalmers

- Pour la conversation : si nous rejetons le dualisme, pouvons-nous encore nous appuyer sur son argument pour établir le fonctionnalisme ?

# Chalmers

- (oui: en effet il *utilise* son dualisme pour établir la possibilité de cas de qualia absents ou inversés. (mais les théoriciens de l'identité cerveau-état devraient aussi l'accepter)
- (non: Tout cela est basé sur un appel à la simplicité des lois, mais si le réductionnisme est correct, nous ne cherchons pas des lois, mais des principes d'identité...)

# Chalmers

- Les arguments:
- Des qualia évanescents contre des qualia absents
- Des qualia dansants contre des qualia inversés

# Chalmers

- Qualia absents : doublons fonctionnels (artificiels) de nous qui n'ont pas d'expériences.
- Qualia inversés : doublons fonctionnels (artificiels) de nous qui ont des expériences différentes des nôtres.

# Qualia évanescents

- 1) Il est possible qu'il existe un robot, fonctionnellement isomorphe à moi, Z, sans qualia
- 2) Si 1), alors il est possible qu'il y ait une série de cas, chacun avec un neurone remplacé par une puce en silicium, entre moi et Z
- 3) soit le passage de mon expérience à aucune expérience est soudain, sur une seule étape, soit il est progressif (qualia évanescents)

# Qualia évanescents

- 3) soit le passage de mon expérience à aucune expérience est soudain, sur une seule étape, soit il est progressif (qualia évanescents)
  - 3a) si soudain: «Des discontinuités brutales dans les lois de la nature, contrairement à celles que l'on trouve partout ailleurs »
  - 3b) si progressif: alors le sujet se trompe sur (presque) tout ce qu'il vit, violant les principes naturels qui relient la conscience et la cognition (ce qui rend les lois qui les relient très compliquées).

# Qualia évanescents

- 4) Dans tous les cas, les lois de la nature seraient très laides : plus laides que s'il n'y avait pas de qualia effervescents.
- 5) Dans tous les cas, les lois de la nature seraient très laides : plus laides que si les qualia effervescents n'existaient pas.
- 6) Les lois ne sont pas aussi laides
- 7) Par conséquent, il n'y a pas de qualia effervescents - preuve par contradiction que 1) et faux

# Qualia dansants

- 1) Il est possible qu'il existe un robot, fonctionnellement isomorphe à moi, Z, avec qualia différents
- 2) Si 1), Il doit donc y avoir deux systems A et B qui diffèrent au maximum par un dixième de leur composition interne (neurone vs silicone), mais qui ont des expériences significativement différentes
- 3) Il est possible de prendre un circuit en silicone comme celui de B et de l'installer comme circuit de secours dans A (activé par un interrupteur).

# Qualia dansants

- 4) Le basculement de cet interrupteur ferait danser les qualia de A, sans que A s'en aperçoive (même si nous supposons que c'est l'aspect auquel A prête attention)
- 5) Si 4), alors les lois reliant la conscience et la cognition sont très complexes et inélégantes
- 6) les lois reliant la conscience et la cognition ne sont pas complexes et inélégantes.
- 7) Donc, ) Il n'est pas possible qu'il existe un robot, fonctionnellement isomorphe à moi, Z, avec qualia différents