

L'accès à la conscience

Stanislas Dehaene
Chaire de Psychologie Cognitive Expérimentale

Cours

Théories de la conscience d'accès

Résumé du cours précédent

- La conscience est un objet légitime de recherches expérimentales, à condition
 - De prendre au sérieux l'introspection des sujets
 - D'identifier des contrastes minimaux entre traitements conscient et non-conscient
- Distinctions entre
 - Conscience d'accès à une information
 - Conscience au sens intransitif (niveaux de vigilance)
 - Conscience réflexive (*higher-order thoughts*)
 - Conscience phénoménale (expérience subjective, *qualia*)
 - Conscience de soi
 - Différence soi/non-soi (Jeannerod, Frith/Blakemore)
 - Point de vue subjectif et illusions de déplacement du soi (Olaf Blanke)
 - Activité de repos et référence au soi (Buckner)

La synthèse moderne des idées sur l'accès à la conscience

Il existe un début de convergence, sinon vers un « modèle standard », du moins en faveur d'un ensemble d'idées proposées depuis les années 1950-1960, et de mieux en mieux acceptées:

La conscience serait associée à

- Un système de supervision centrale
- A capacité limitée
- Nécessitant des boucles lentes, réentrantes et descendantes (« top-down ») par opposition au traitement ascendant (« bottom-up ») rapide et non-conscient des données sensorielles
- Un espace interne de synthèse, de maintien et de partage des données (« théâtre », « tableau noir », « espace de travail neuronal global »)

Première idée: un système de supervision centrale

Idée ancienne, présente par exemple chez William James (1890) :

- « L'étude de la distribution de la conscience montre qu'elle est exactement ce qu'on attendrait d'un organe ajouté afin de **diriger un système nerveux** devenu trop complexe pour se réguler lui-même »
- La conscience n'est intense que lorsque les processus nerveux sont hésitants. Au cours des actions rapides, automatiques, habituelles, elle se réduit à un minimum.

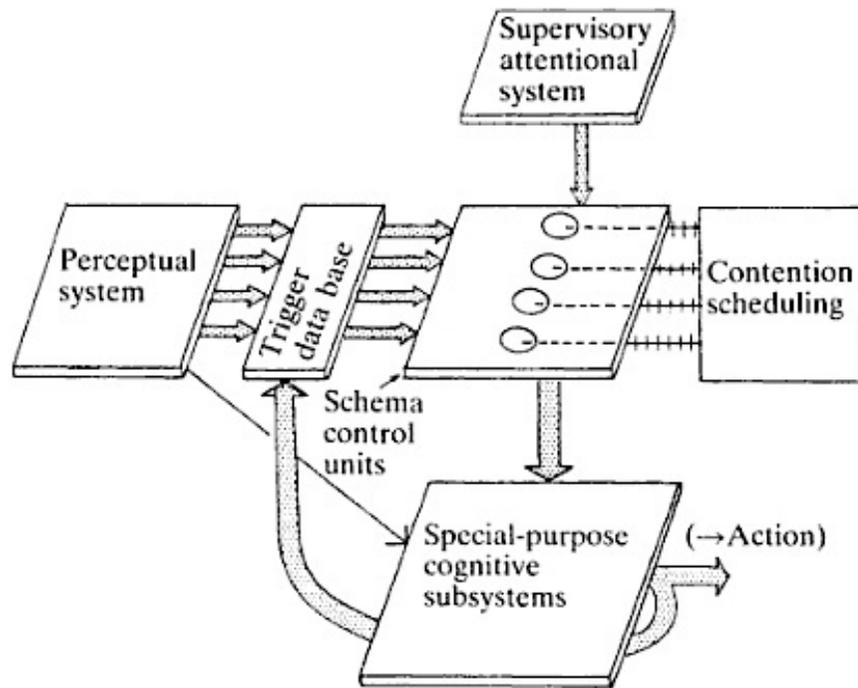
Idée reprise par Michael Posner dans sa distinction entre processus automatiques et processus contrôlés (Posner & Klein, 1973; Posner & Snyder, 1975)

- Certains processus mentaux, dits « automatiques », démarrent sans intention, n'interfèrent pas avec les autres, et échappent à la conscience
- D'autres, dits « contrôlés », font appel à un système central à capacité limitée, qui ne permet pas l'exécution de plusieurs opérations simultanées sans interférence; ils sont dépendants de nos intentions et conduisent à une expérience consciente.

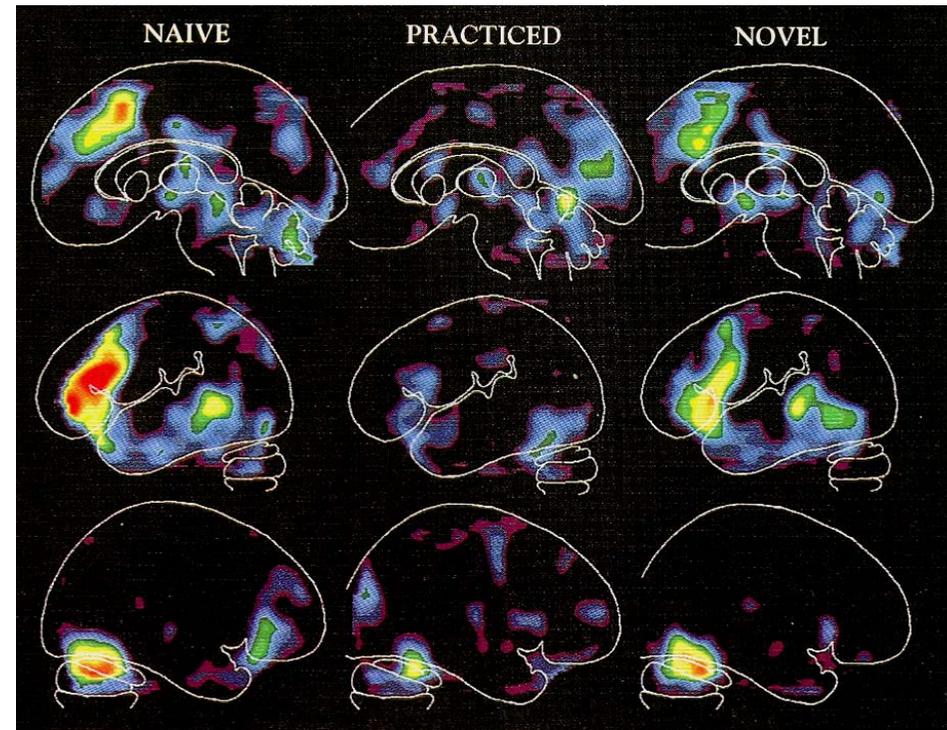
Selon Shallice (1978, Norman & Shallice, 1980), le comportement volontaire conscient résulterait d'un système de supervision, hiérarchiquement supérieur aux processeurs automatiques, et chargé de leur contrôle et de leur inhibition

Le superviseur attentionnel de Shallice

Un grand nombre de données neuropsychologiques et d'imagerie cérébrales associe la supervision attentionnelle ou « contrôle exécutif » aux cortex préfrontaux et cingulaire antérieur.



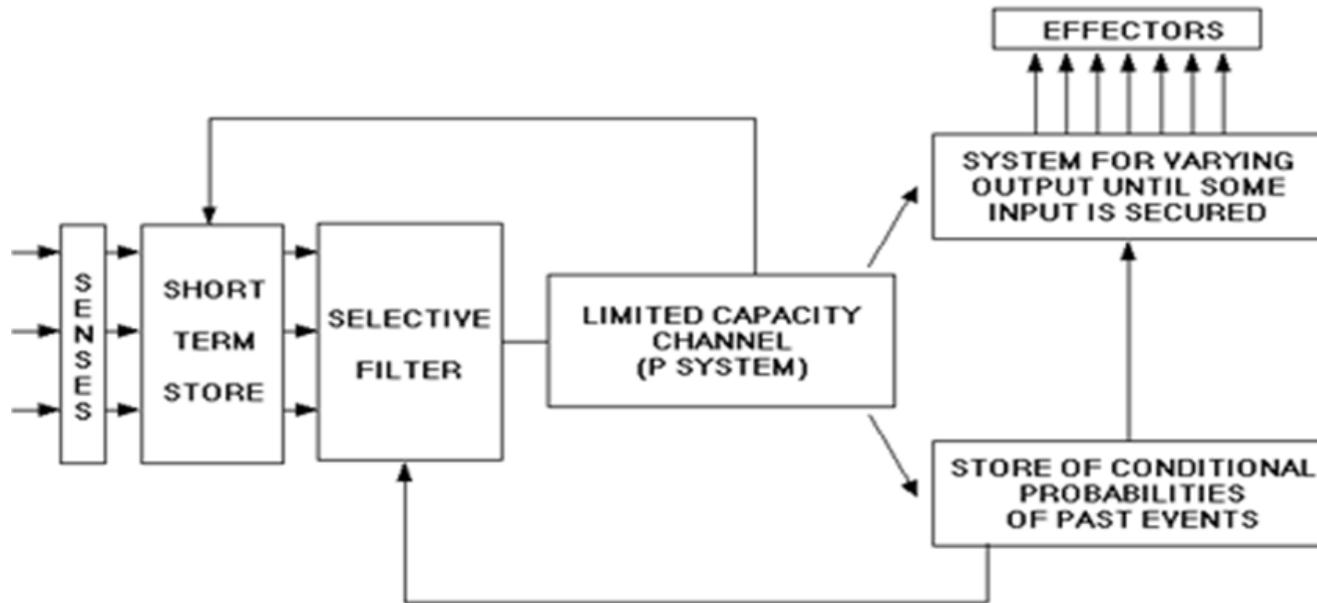
Norman-Shallice 1980



M. Raichle, J. Fiez et al., *Cerebral Cortex*, 1994

Seconde idée: une étape sélective à capacité limitée

Broadbent (1958): La perception comprend un filtre qui ne laisse entrer qu'une fraction des informations dans un « canal à capacité limitée »

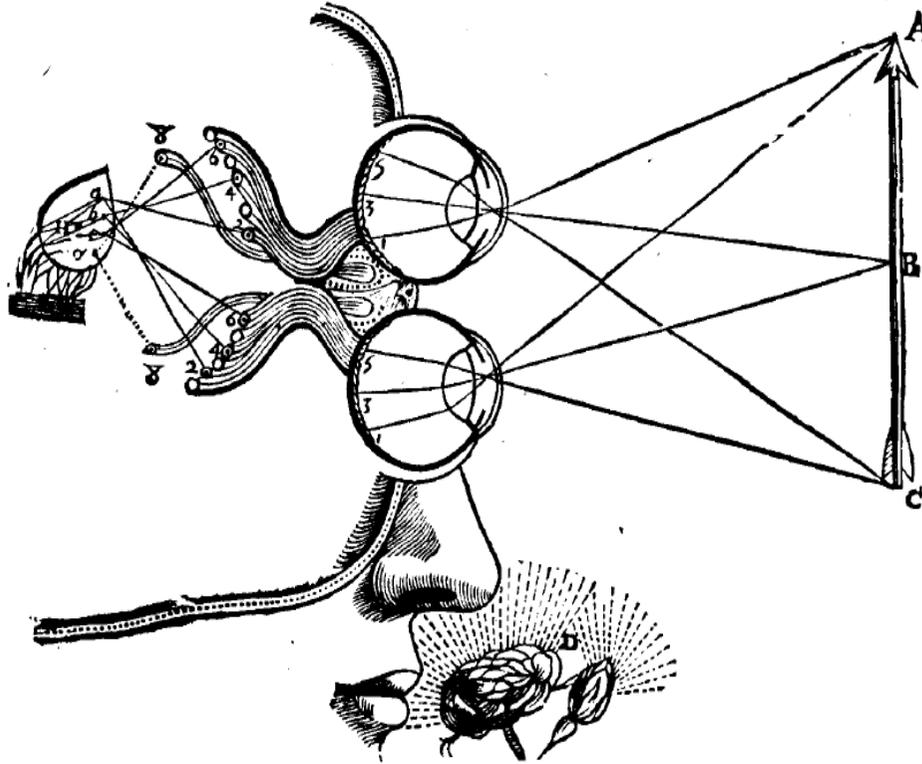


Cette théorie répond à une nécessité algorithmique: tout organisme est bombardé en permanence de stimuli sensoriels qui excèdent sa capacité de traitement et surtout d'action.

Il faut donc *sélectionner* une fraction de ces entrées (rôle de l'attention sélective)

La conscience serait associée au traitement plus approfondi de la fraction des données sélectionnée pour sa pertinence vis-à-vis des buts de l'organisme

« Les idées s'empêchent l'une l'autre »... Une idée très ancienne



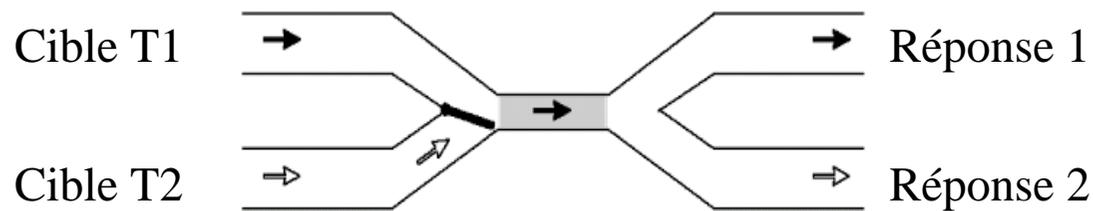
« Pendant que cette glande est retenue ainsi penchée vers quelque côté, cela l'empêche de pouvoir si aisément recevoir les idées des objets qui agissent contre les organes des autres sens. (...) D'où vous voyez comment les idées s'empêchent l'une l'autre, et d'où vient qu'on ne peut être fort attentif à plusieurs choses en même temps. »

René Descartes, *Traité de l'homme*, 1648

Plusieurs modèles récents en deux étapes

Pashler, H. (1994). Dual-task interference in simple tasks: data and theory. *Psychol Bull*, 116(2), 220-244.

Hypothèse du goulot d'étranglement central



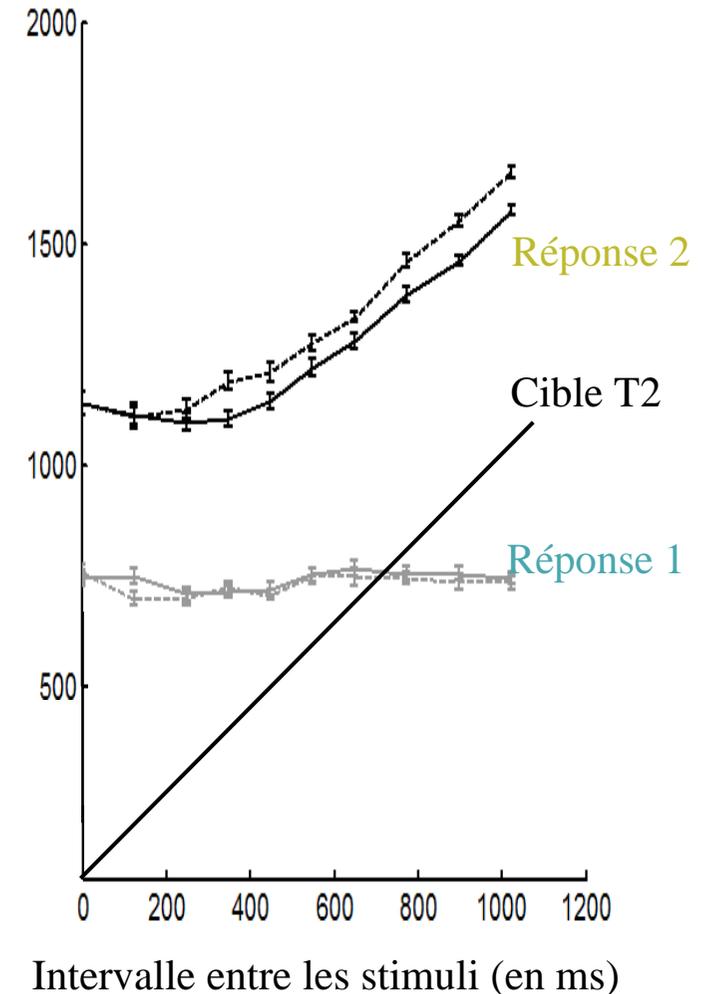
Observation empirique pertinente:
La « période psychologique réfractaire »
(Welford, 1952; Pashler, 1984)

Lors de l'exécution d'une tâche double, si les stimuli sont très proches, la première tâche ralentit la seconde.

En réalité, seule l'étape *centrale* de la seconde tâche est ralentie.

Lien avec la conscience: le sujet n'a pas conscience que T2 est mise en attente (Corallo et al, *Psych Science* 2009; Marti et al., *Cognition*, sous presse)

Temps de réponse



Plusieurs modèles récents en deux étapes

Chun, M. M., & Potter, M. C. (1995). A two-stage model for multiple target detection in rapid serial visual presentation. *J Exp Psychol Hum Percept Perform*, 21(1), 109-127.

1^{ère} étape: détection et sélection rapide des candidats (au moins 10 items/seconde)

2^{ème} étape: transfert en mémoire à court-terme, identification et consolidation (2 ou 3/s)

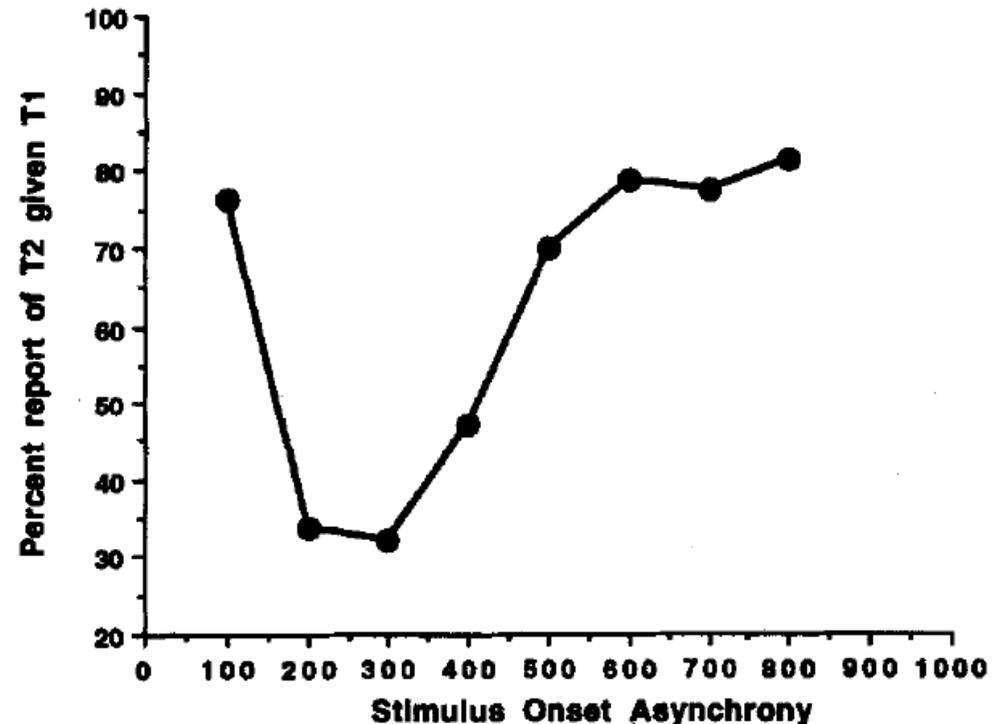
Observation empirique pertinente: le clignement attentionnel (*attentional blink*; Broadbent & Broadbent, 1987; Raymond, Shapiro et Arnell, 1992):

Dans une série de stimuli successifs, la détection d'une cible empêche la détection ultérieure d'une seconde cible présentée dans les ~400 ms suivantes.

Liens avec la conscience:

- Le sujet perd toute conscience de la présence même de T2
- Lorsque T1 est au seuil de conscience, l'effet n'existe que si T1 a été consciemment détectée

(Nieuwenstein., Van der Burg, Theeuwes, Wyble & Potter, 2009)



Plusieurs modèles récents en deux étapes

Huang, L., Treisman, A., & Pashler, H. (2007). Characterizing the limits of human visual awareness. *Science*, 317(5839), 823-825; Voir également Huang & Pashler, *Psychological Review*, 2007

Selon cette théorie spéculative, la conscience visuelle serait limitée, à un instant donné, à une carte booléenne étiquetée (*labeled Boolean map*) qui lie un seul trait donné par dimension (par exemple une couleur *et* une orientation) avec une carte spatiale. Nous ne pourrions donc pas connaître, en un seul acte de conscience, la présence de plus d'une couleur.

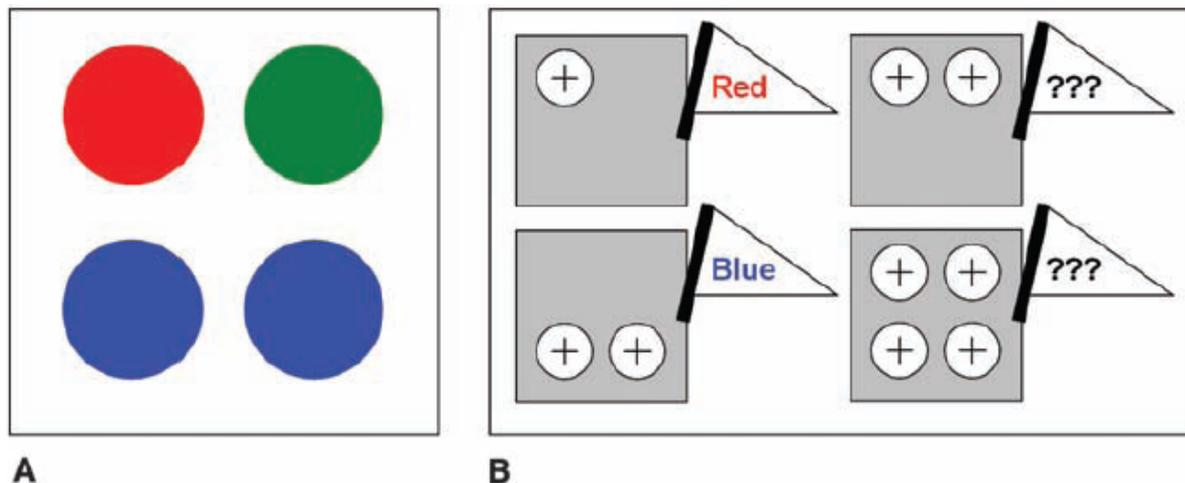
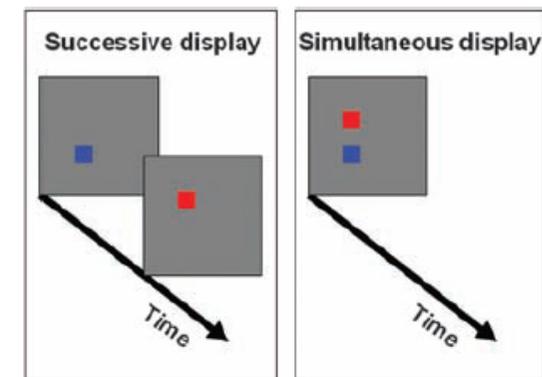
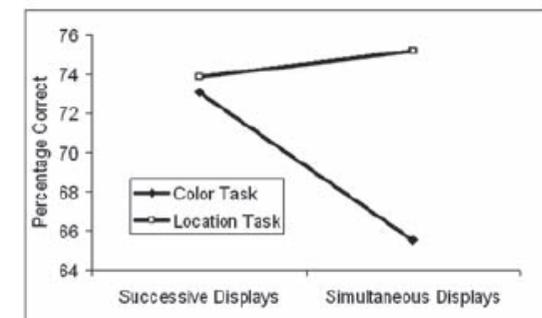


Fig. 1. (A) A sample display. (B) Some possible states of awareness (possible percepts) when viewing the sample display of (A) in a single brief exposure. When only disks of one color are selected (left two examples), both the locations of the disks and that color value can be consciously accessed. When more than one color is present in the selected disks (right two examples), only the locations of the disks, but not their colors, can be consciously represented. In (B), the regions marked with plus signs stand for the selected regions of the Boolean map, and the flags stand for the color information that is associated with the selected region as a label. The selected region in the map and the color label constitute the only visual information that is consciously accessible at one time.



A Method of Experiment 1



B Results of Experiment 1

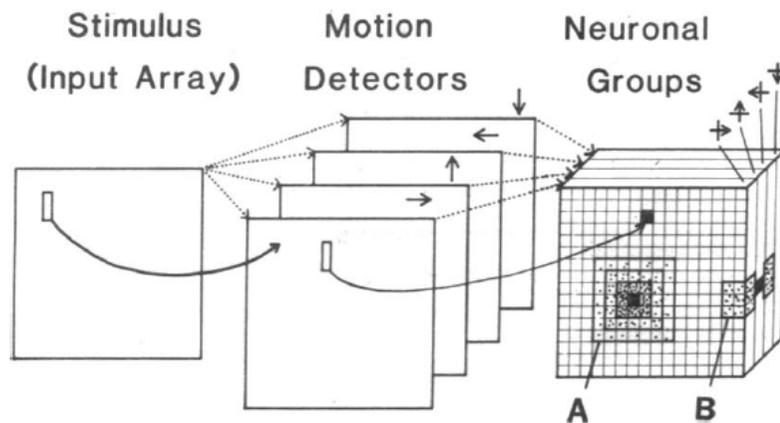
La distinction entre conscience et attention

- Les modèles précédents donnent l'impression que la conscience et l'attention sont indistingables ("My experience is what I agree to attend to", William James)
- Effectivement, si l'on définit l'attention par sa seule nature *sélective* ("the taking possession by the mind, in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought", James 1890) alors la conscience, de par sa capacité limitée, requiert une sélection, donc l'attention.
- Cependant, l'attention possède aujourd'hui de nombreuses acceptions. Si l'on s'en tient à la sélection d'un objet ou d'une position spatiale, et l'amplification de ses attributs sensoriels, alors il ne fait aucun doute que l'attention et la conscience soient dissociables (Koch and Tsuchiya, 2007; voir cours 2009).
 - Des effets d'alerte et de réorientation de l'attention peuvent être induits par des indices subliminaux (Bressan and Pizzighello, 2008; McCormick, 1997)
 - Le déplacement attentionnel possède une signature spécifique en potentiels évoqués, l'onde N2pc, qui est facilement observée dans des conditions subliminales (Robitaille and Jolicoeur, 2006; Woodman and Luck, 2003).
 - L'attention peut amplifier le traitement de stimuli subliminaux (Kentridge et al., 2008; Kiefer and Brendel, 2006; Naccache et al., 2002).
 - Les corrélats en MEG de l'attention et de la conscience de stimuli présentés au seuil de conscience sont radicalement différents (Wyart and Tallon-Baudry, 2008).

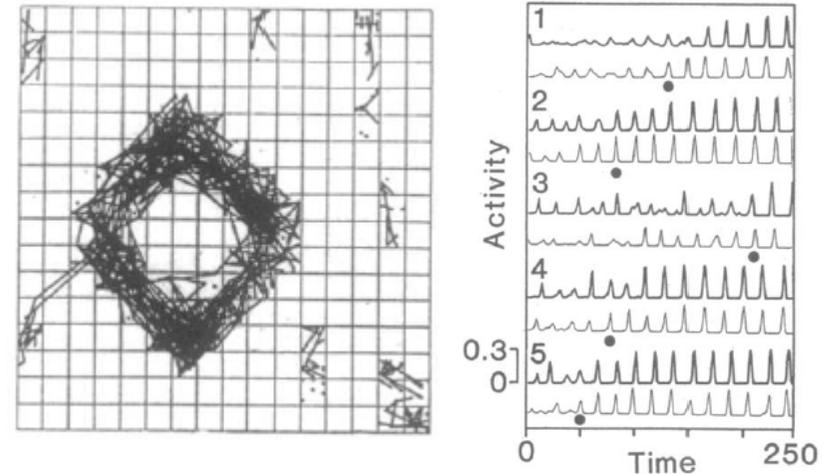
Troisième idée: la distinction entre traitement perceptif précoce et ascendant (*feedforward*) et amplification tardive réentrante et descendante (*top-down*)

Gerald Edelman (*Neuronal Darwinism*, 1987) voit dans la « **ré-entrée** » un mécanisme indispensable de la perception unifiée: L'échange permanent, dynamique et récursif de signaux parallèles entre cartes cérébrales

Exemple: modèle de groupement perceptif de Sporns, Tononi et Edelman, *PNAS* 1991



« liage » de neurones codant pour un déplacement global cohérent



Toutefois, dans cette simulation, la ré-entrée survient à des niveaux perceptifs extrêmement précoces et vraisemblablement non-conscients

Pour Edelman, ce sont certaines formes de ré-entrée qui conduisent à la conscience.

Un modèle de conscience par ré-entrée massive

Edelman, *The remembered present*, 1989

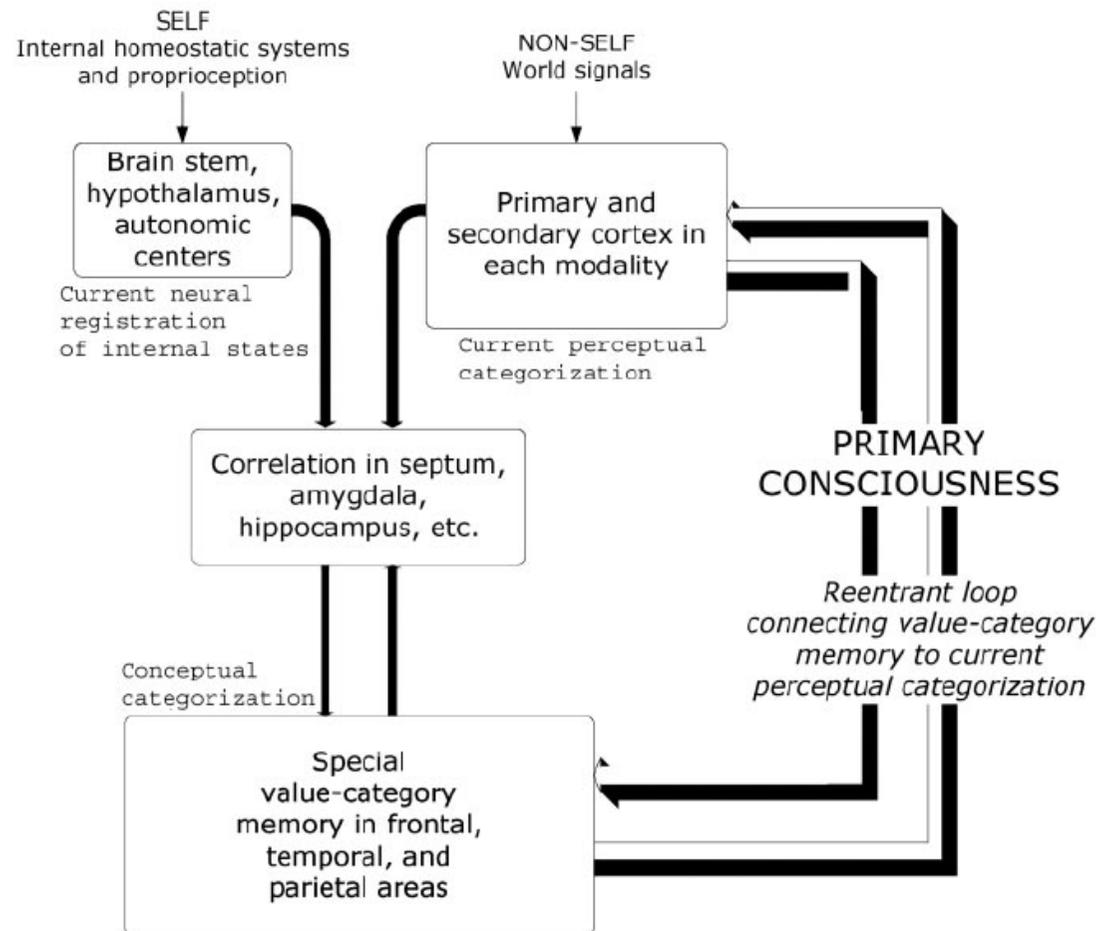


Fig. 1. Reentrant pathways leading to primary consciousness. Two main kinds of signals are critical: those from self, constituting value systems and regulatory elements of the brain and body, and those from nonself, signals from the world that are transformed through global mappings. Signals related to value and categorized signals from the outside world are correlated and lead to memory, which is capable of conceptual categorization. This value-category memory is linked by reentrant paths (heavy lines) to the current perceptual categorization of world signals. This reentrant linkage is the critical evolutionary development that results in primary consciousness. When it occurs across many modalities (sight, touch, and so forth), primary consciousness can connect objects and events through the memory of previous value-laden experiences. The activity of the underlying reentrant neural systems results in the ability to carry out high-level discriminations. This ability enhances survival value.

Les idées plus récentes d'Edelman et Tononi: Différentiation et Intégration (*Science*, 1998)

L'information consciente est à la fois

- **intégrée**: il est impossible de concevoir une scène consciente qui ne soit pas unifiée (cf rivalité binoculaire, blink)
- **différenciée**: le nombre de contenus de conscience est gigantesque, et chacun d'eux est très informatif

La **différentiation** des états neuronaux résulte de la théorie des groupes neuronaux. Elle se traduit par un EEG désynchronisé, qui disparaît dans les crises épileptiques généralisées ou dans le sommeil à ondes lentes)

L'**intégration** de l'information résulte des nombreuses boucles ré-entrantes. Elle peut être mesurée mathématiquement:

Phi = information mutuelle minimale entre deux sous-parties du système complet

« Consciousness is information integration »

Conséquences pour le cerveau humain:

- pas de localisation neurale précise, mais un « noyau dynamique » (*dynamic core*) thalamo-cortical distribué, dont les contours ne cessent de changer, et qui varie considérablement d'une personne à l'autre
- Exemple: travaux de Srinivasan, Russell, Edelman et Tononi (*J. Neuroscience*, 1999)

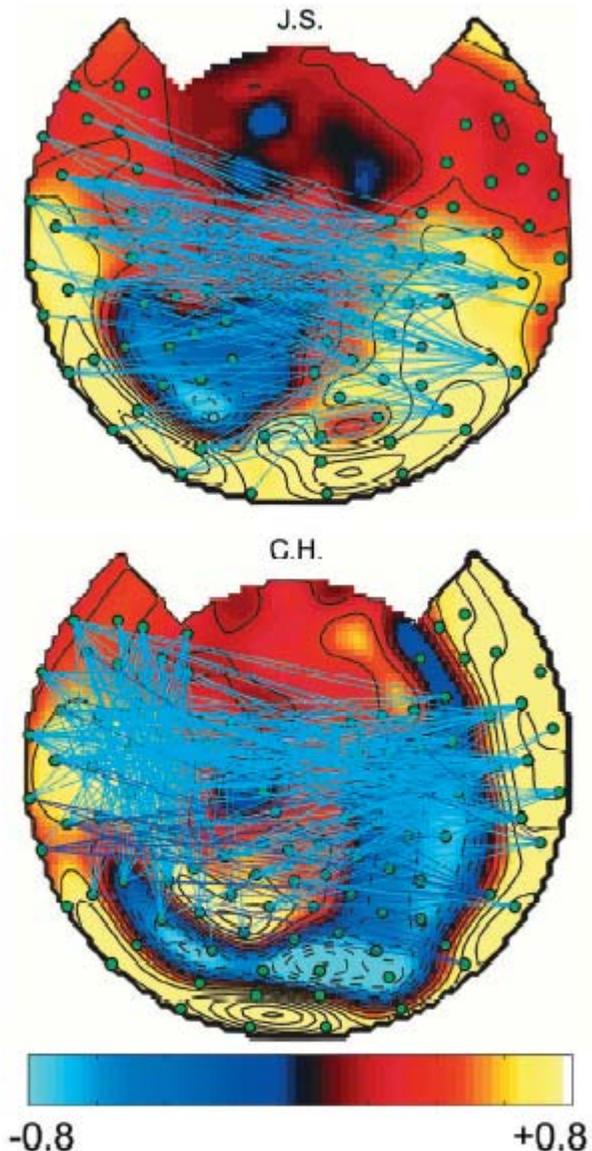
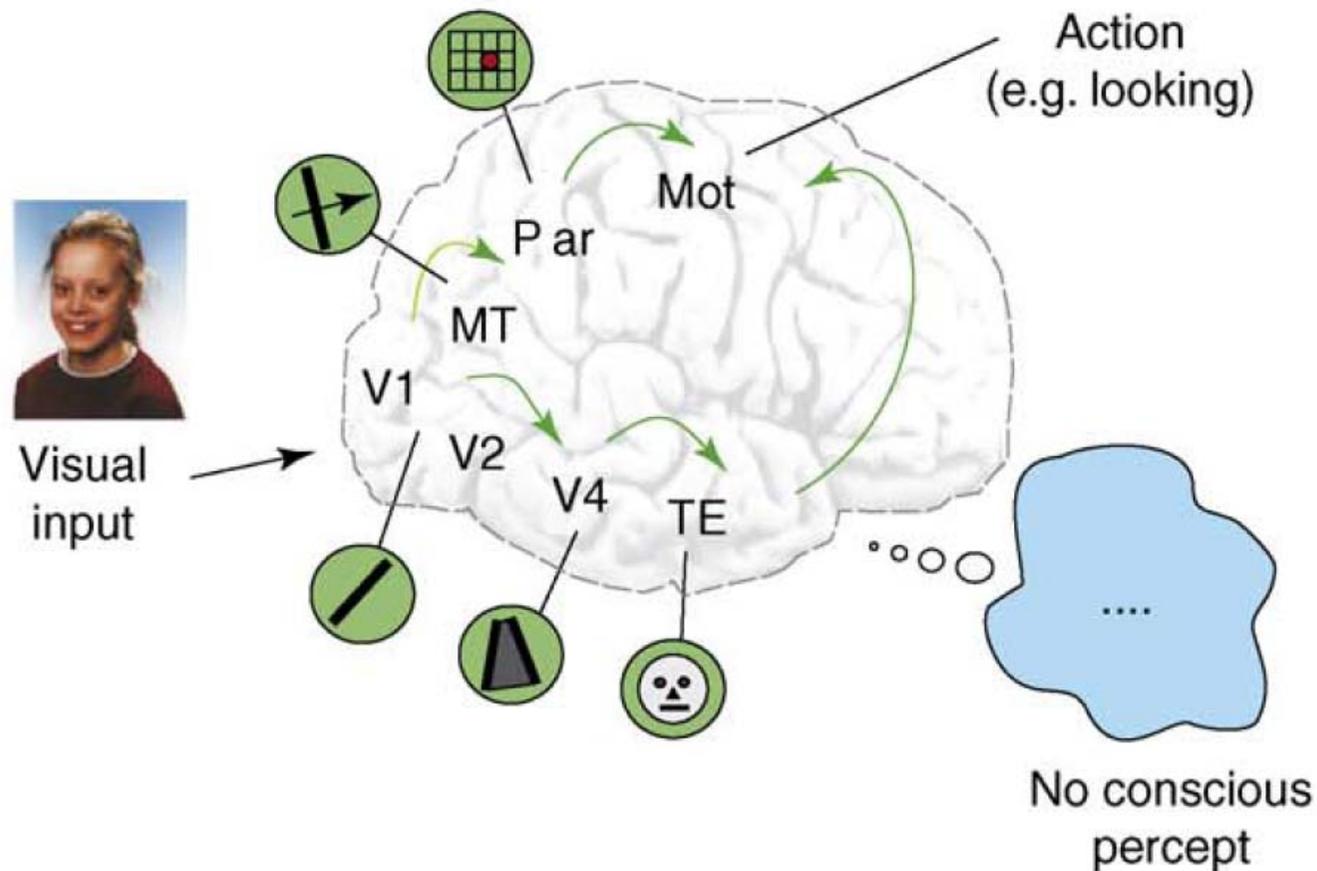


Fig. 1. Amplitude and coherence differences between the steady-state neuromagnetic responses during binocular rivalry when subjects were conscious of a stimulus and when they were not.

Le rôle essentiel des connexions récurrentes dans le modèle de Victor Lamme (*TICS*, 2006)

Pour Victor Lamme, trois étapes conduisent à la perception consciente

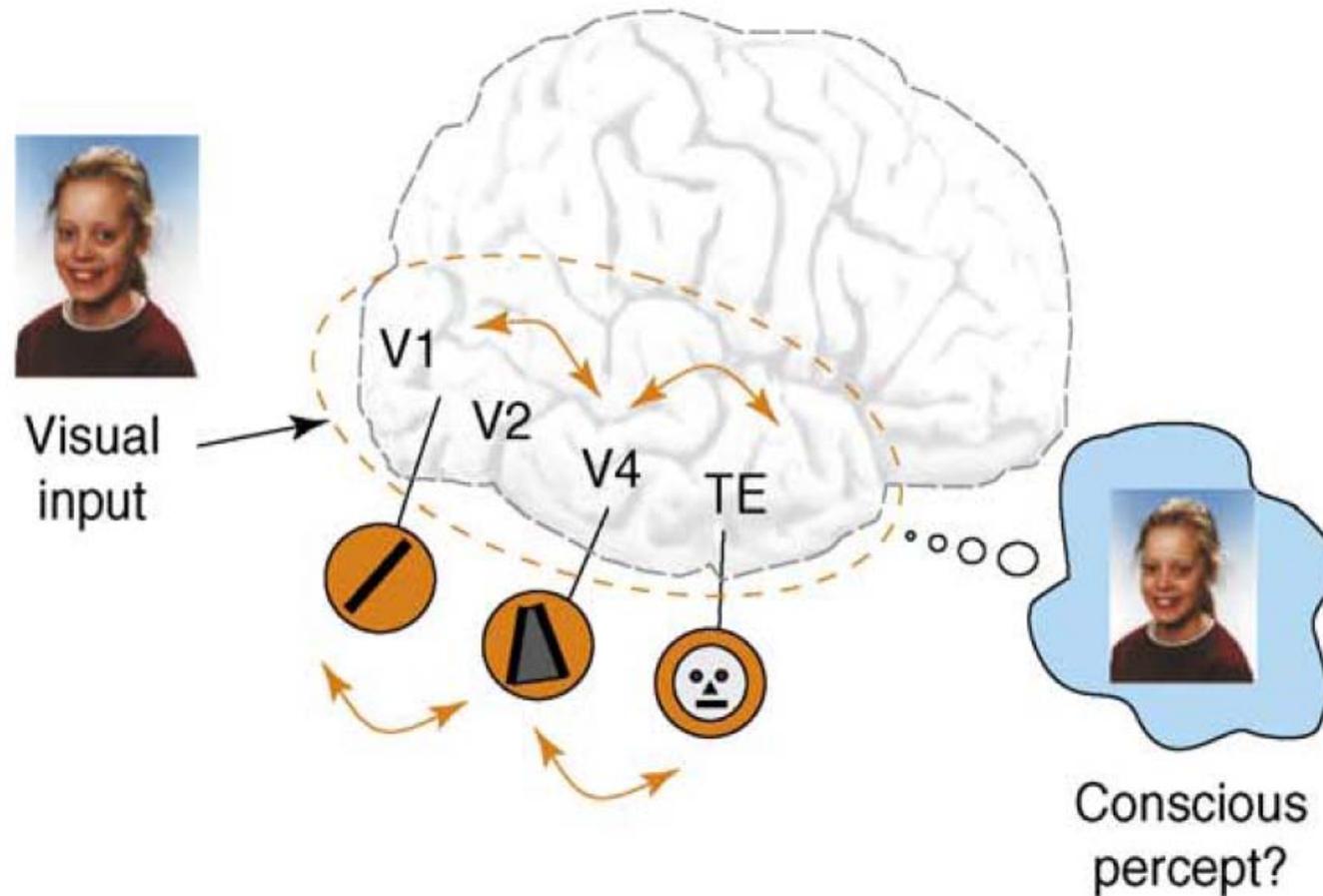
(a) The feedforward sweep



Le rôle essentiel des connexions récurrentes dans le modèle de Victor Lamme (*TICS*, 2006)

Pour Victor Lamme, trois étapes conduisent à la perception consciente

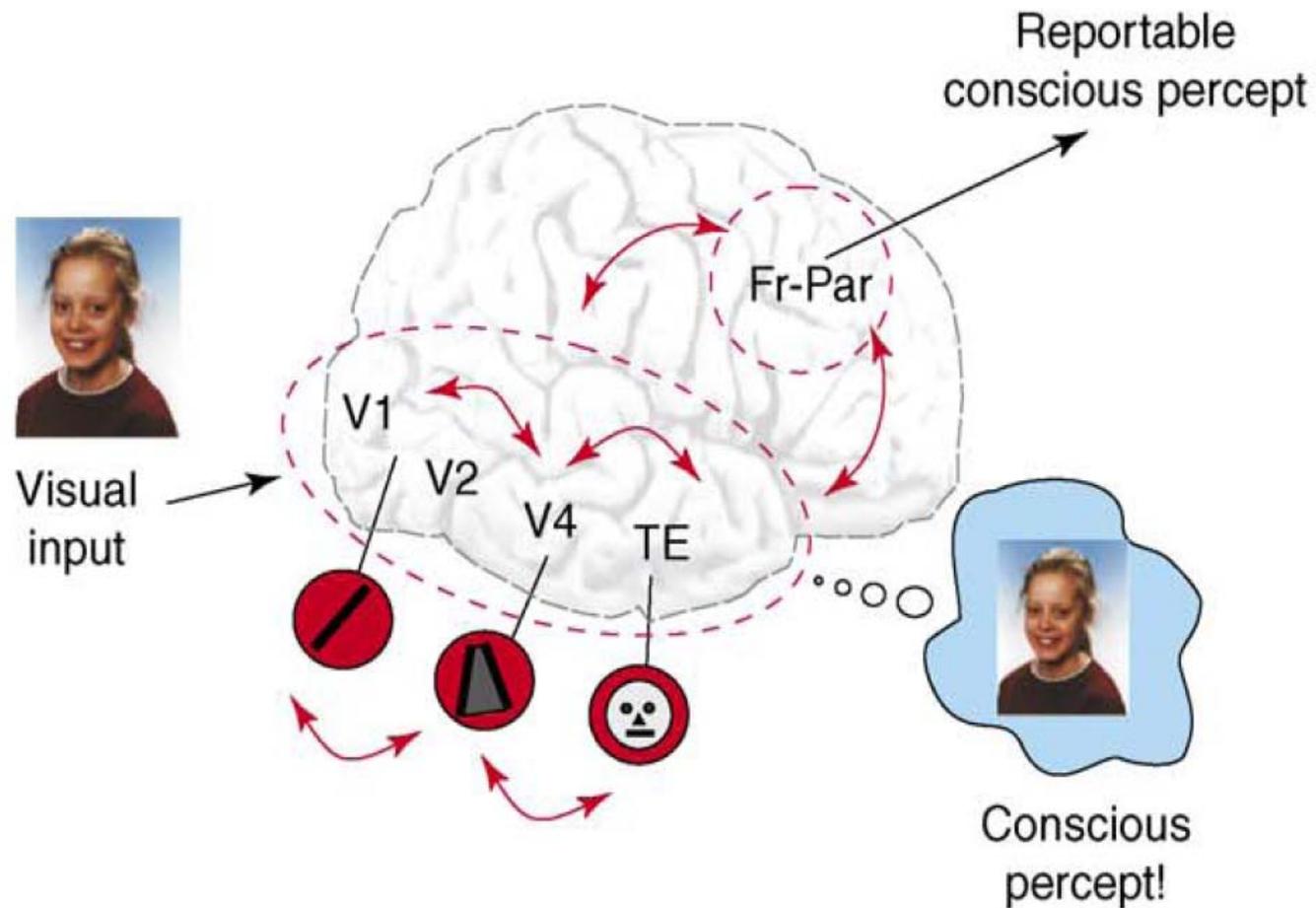
(b) Localized recurrent processing



Le rôle essentiel des connexions récurrentes dans le modèle de Victor Lamme (*TICS*, 2006)

Pour Victor Lamme, trois étapes conduisent à la perception consciente

(c) Widespread recurrent processing



Quatrième idée: un « théâtre » de synthèse mentale qui permet la diffusion des informations pertinentes

Hippolyte Taine (*De l'intelligence*, 1870)
Métaphore du « théâtre mental »

On peut comparer l'esprit d'un homme à un théâtre d'une profondeur indéfinie, dont la rampe est très étroite, mais dont la scène va s'élargissant à partir de la rampe. Devant cette rampe éclairée, il n'y a guère de place que pour un seul acteur... Au delà, sur les divers plans de la scène, sont d'autres groupes d'autant moins distincts qu'ils sont plus loin de la rampe. Au delà de ces groupes, dans les coulisses et l'arrière-fond lointain, se trouve une multitude de formes obscures qu'un appel soudain amène parfois sur la scène ou même jusque sous les feux de la rampe, et des évolutions inconnues s'opèrent incessamment dans cette fourmilière d'acteurs de tous ordres pour fournir les coryphées qui tour à tour, comme en une lanterne magique, viennent défiler devant nos yeux.

Le modèle de l'espace de travail global conscient

Bernard Baars, *A cognitive theory of consciousness*, 1989

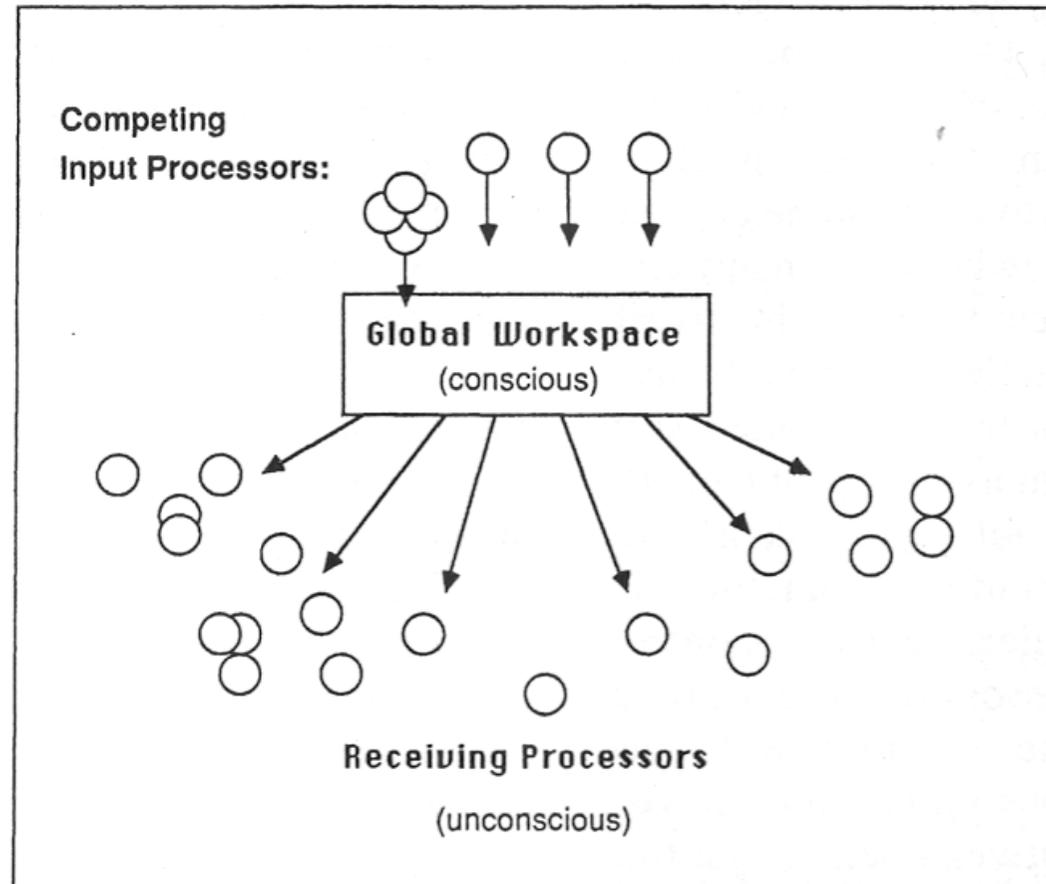
De très nombreux processeurs spécialisés opèrent en parallèle.

A un instant donné, une seule coalition de processeurs dominants envoie son résultat dans l'**espace de travail global**.

L'accès conscient est contraint par une hiérarchie de contextes enchâssés et non-conscients.

Ce résultat est alors diffusé (*broadcasted*) à l'ensemble des autres processeurs, qui peuvent ou non l'utiliser.

Le processus se répète, selon une alternance de traitements sériels/parallèles qui, selon Shanahan & Baars (2005), offre une solution au problème du cadre (*frame problem*).



Model 1. A global workspace in a distributed system.

Multiples distinctions entre les processeurs et l'espace de travail global

(Baars, 1989)

Table 2.1. *Capabilities of comparable conscious and unconscious processes*

Capabilities of conscious processes	Capabilities of unconscious processors
1 Computationally inefficient: High number of errors, low speed, and mutual interference between conscious computations	1 Highly efficient in their own tasks: Low number of errors, high speed, and little mutual interference
2 Great range of different contents over time; great ability to relate different con- scious contents to each other; great ability to relate conscious events to their unconscious contexts	2 Each specialized processor has limited range over time; each one is relatively isolated and autonomous
3 Have internal consistency, seriality, and limited capacity	3 Diverse, can operate in parallel, and together have great capacity

Les mécanismes cérébraux de l'espace de travail global

(Selon Baars, 1989)

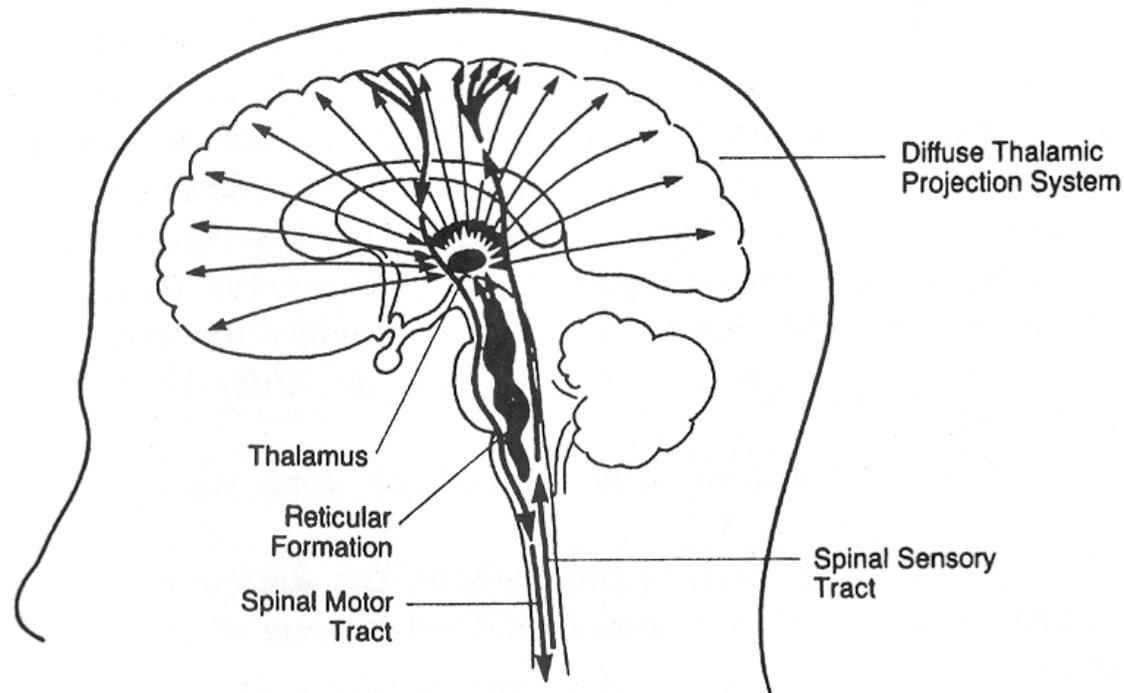
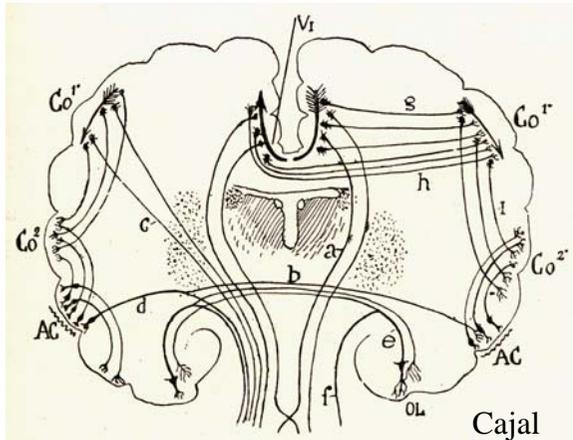


Figure 3.1. The ERTAS: A neural global workspace? Brain structures most closely associated with conscious experience include the Reticular Formation of the brain stem and midbrain, the outer shell of the thalamus, and the set of neurons projecting upward diffusely from the thalamus to the cerebral cortex. Together these structures can be labeled the "Extended Reticular-Thalamic Activating System" (ERTAS), since stimulation of a number of neurons throughout this system causes cortical activation. The ERTAS as a whole can be interpreted as a functional global workspace

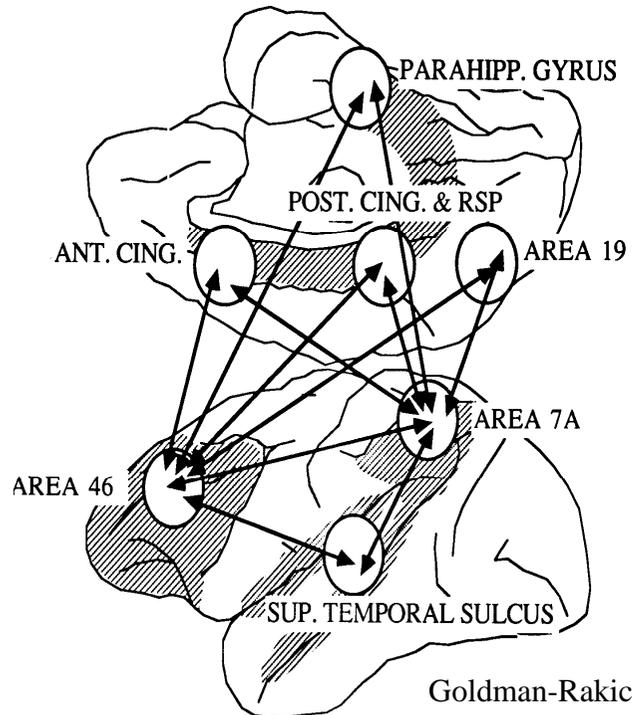
L'espace de travail *neuronal* global (GNW)

Dehaene, Kerszberg & Changeux, *PNAS* 1998

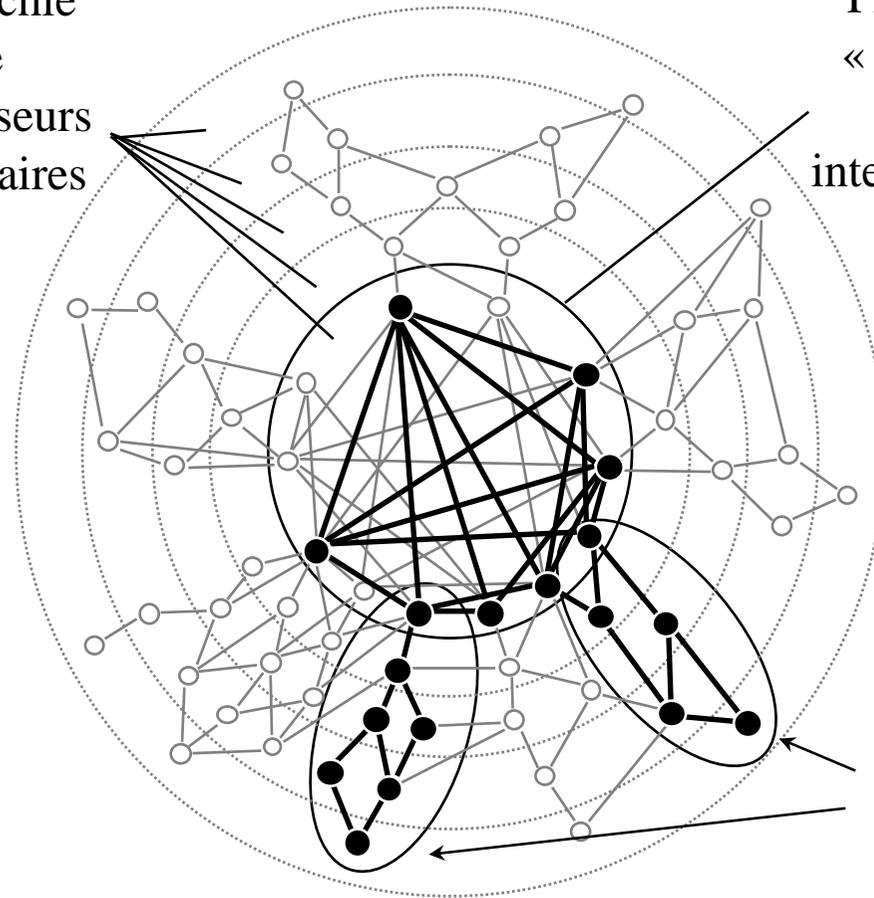


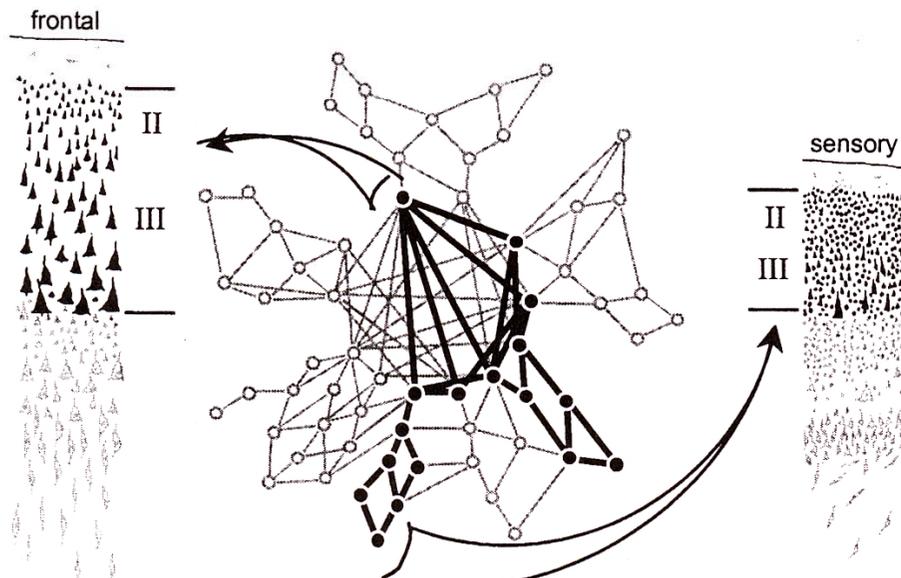
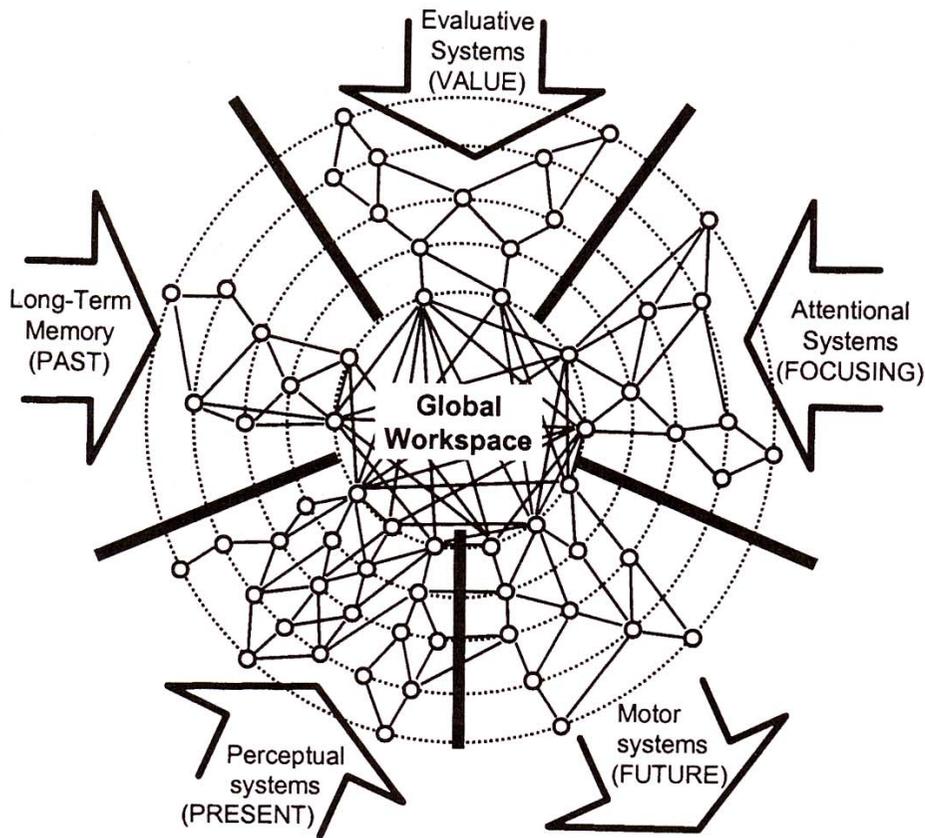
hiérarchie
de
processeurs
modulaires

Processeurs
« centraux »
à forte
interconnection



Processeurs
périphériques
mobilisés
par la tâche





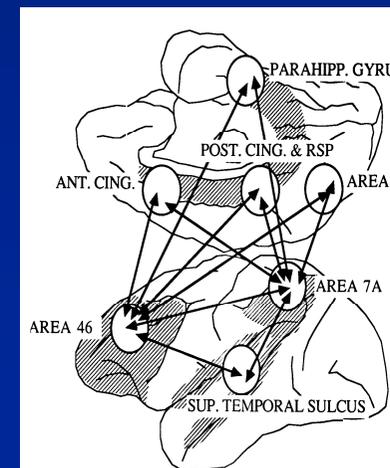
Le modèle de l'espace de travail neuronal global

Dehaene, Kerszberg & Changeux, PNAS 1998

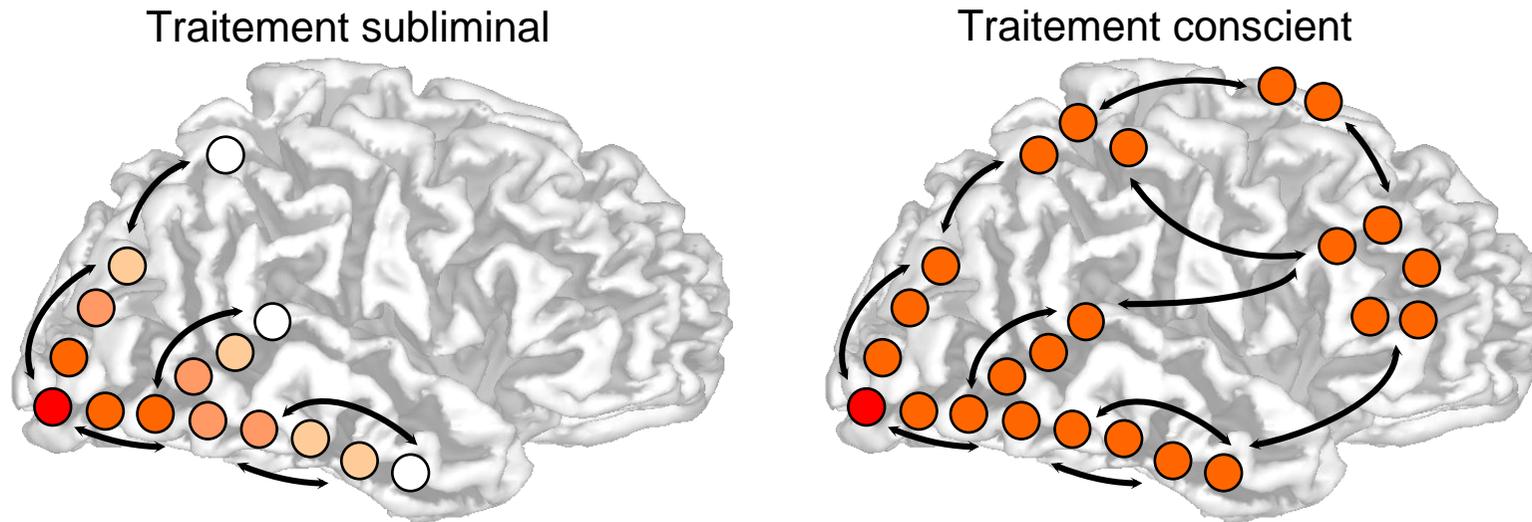
Deux espaces computationnels

1) *Les processeurs spécialisés* modulaires, encapsulés et automatiques

2) *un espace global* formé de neurones à axones longs capables de diffuser l'information à des aires très distantes, ce qui conduit à une expérience subjective synthétique et rapportable



L'accès à la conscience selon le modèle GNW



- Ce que nous appelons “l'accès à la conscience” est la **disponibilité globale d'une information**, qui la rend **accessible** à tous les processeurs et donc **rapportable** (à nous même ou aux autres)
- L'accès à la conscience correspond à l'**ignition soudaine** de neurones à axones longs, distribués mais particulièrement denses dans les cortex préfrontaux et pariétaux (*global neuronal workspace*, GNW)
- Durant la perception conscience, le GNW se stabilise, pour une durée minimale, dans un **état réverbérant et métastable** et distribue cette information au reste du cortex
- Le même processeur peut participer à des traitements conscients et non-conscients, le mode conscient se caractérisant par une **amplification** descendante et une **mise en synchronie** avec les neurones du GNW.
- Les neurones du GNW sont le siège d'une **activité spontanée permanente**, une série temporelle d'états métastables, modulée par (1) des signaux ascendants d'éveil; (2) des signaux sélectifs de renforcement

L'espace de travail et les *qualia* de l'expérience phénoménale

Dehaene & Naccache, 2001

- « Le flux permanent des états neuronaux du GNW associés à n'importe quelle expérience perceptive dépasse très largement ce qui peut être rapporté verbalement ou stocké en mémoire à long terme.
- De plus, bien que les traits majeurs de ce répertoire soient partagés par tous les membres de notre espèce, ses détails résultent d'un processus développemental d'épigénèse et sont donc propres à chaque individu.
- En bref, les contenus de la conscience perceptive sont des états neuronaux complexes, dynamiques, et multidimensionnels qui ne peuvent pas être totalement mémorisés ou transmis à d'autres personnes.
- Ces propriétés biologiques semblent susceptibles de substantier l'intuition philosophique des *qualia* de l'expérience sensible »
- ... sans qu'il soit nécessaire de postuler aucune autre propriété ou substance nouvelle.

Un désaccord superficiel avec le *global workspace*: la notion d'ébauches multiples (*multiple drafts*)

Dennett, *Consciousness explained*, 1991

Dans un effort louable de lutter contre l'idée d'un *homunculus*, Dennett milite fermement contre la notion d'un *théâtre Cartésien* responsable de la conscience.

A la place existerait un processus délocalisé de révision permanente:

« all varieties of thought and mental activity are accomplished in the brain by parallel, multi track processes of interpretation and elaboration of sensory inputs. Information entering the nervous system is under continuous 'editorial revision'. »

Cependant, cette idée n'est pas nécessairement incompatible avec la théorie de l'espace global. Dans le numéro spécial de *Cognition* (2001), Dennett se rallie à la proposition d'un espace de travail cortical global, aux conditions suivantes

- il s'agit d'un système distribué -- il n'existe pas de spectateur subjectif du théâtre conscient, mais la subjectivité émerge d'une collection de processeurs « stupides »
- il se suffit à lui-même: l'accès à ce système ne cause pas la conscience mais *est* la conscience

Dennett développe la métaphore de la *célébrité cérébrale*.

Un désaccord plus profond: La non-unité de la conscience

Zeki, *TICS* 2003

- L'unité de la conscience est une illusion: il existe de multiples consciences qui forment une hiérarchie
- Chaque qualia est associé à une région cérébrale qui génère une "micro-conscience".
 - Couleur dans V4
 - Mouvement dans V5, etc.
- La *microconscience* des différents traits d'un objet peut survenir à des moments différents
- La *macro-conscience* résulte du liage de ces différents traits en un seul percept
- La *conscience unifiée* résulte de l'intégration du percept avec la représentation du sujet en train de percevoir

(similarité avec l'idée de Damasio d'une carte de second ordre, qui représente la manière dont l'organisme est modifié par le stimulus perçu: "a sense of self in the act of knowing")

- Critiques:
 - La proposition des "micro-consciences" reste extrêmement floue: par exemple "It is heightened activity within a specialized cortical area that leads to conscious vision (...) Of course, processing-perceptual sites are not sufficient on their own in generating a conscious correlate but depend upon enabling systems in the brain stem and possibly additional uncharted cortical systems" (Zeki, 2003).
 - Elle ne permet pas d'expliquer les phénomènes de compétition pour l'accès à la conscience, tels que le clignement attentionnel
 - Elle est incompatible avec la découverte d'activations subliminales des aires perceptives

Conclusion:

La synthèse moderne des idées sur l'accès à la conscience

Il existe un début de convergence, sinon vers un « modèle standard », du moins en faveur d'un ensemble d'idées proposées depuis les années 1950-1960, et de mieux en mieux acceptées:

La conscience serait associée à

- Un système de **supervision centrale**
- A **capacité limitée**
- Nécessitant des **boucles lentes**, réentrantes et descendantes (« top-down ») par opposition au traitement ascendant (« bottom-up ») rapide et non-conscient des données sensorielles
- Un **espace interne** de synthèse, de maintien et de partage des données (« théâtre », « tableau noir », « espace de travail neuronal global »)