

La vision active

Le système visuel ne traite pas de façon équivalente l'ensemble des informations lumineuses qui lui parviennent : elles sont activement sélectionnées.

> PAR LAURENT MADELAIN, PROFESSEUR À L'UNIVERSITÉ CHARLES-DE-GAULLE-LILLE-III, ET ANNA MONTAGNINI, CHARGÉE DE RECHERCHE AU CNRS, INSTITUT DE NEUROSCIENCES DE LA TIMONE (INT)

J'ai rendez-vous à la gare avec un ami qui m'a dit qu'il portait du vert aujourd'hui. En arrivant je me rends compte que nous n'allons probablement pas nous trouver : nous sommes à Dublin et aujourd'hui c'est la Saint-Patrick. Je me sens perdu en observant des centaines de personnes habillées en vert, qui se déplacent dans toutes les directions... Je pense que je n'arriverai pas à le repérer avant que le train parte.

La situation ci-dessus illustre l'impossibilité, pour notre système visuel, de traiter en détail toute l'information présente dans notre environnement visuel. Il est tout aussi vrai que des caméras statiques à très haute résolution ne peuvent pas nous aider dans cette tâche, du moins si derrière l'étape d'acquisition de l'image il n'y a pas une unité de calcul parallèle de grande puissance. De la même façon, seul un énorme cerveau, qui pèserait une dizaine de tonnes, pourrait garantir une puissance suffisante pour analyser rapidement les caractéristiques des centaines de personnes qui affolent la gare Connolly à l'heure de pointe. Opérer une sélection dynamique de l'information visuelle est donc indispensable : c'est le rôle de la vision active.

Des bases anatomiques et fonctionnelles

Le terme « vision active » désigne le processus dynamique qui nous permet de sélectionner une partie de l'information visuelle pour la traiter en détail et, en contrepartie, nous amène à traiter de manière plus sommaire voire à négliger le reste de l'information. Quels sont les mécanismes qui nous permettent de mettre en place cette stratégie ? En premier lieu, il y a l'anatomie de notre œil. La lumière provenant des objets présents dans la scène visuelle est projetée sur la rétine, l'organe du système nerveux central qui traduit la lumière en signaux neuronaux, grâce aux cellules

photoréceptrices. La portion centrale de la rétine, la fovéa, est riche de récepteurs (les cônes) sensibles à la couleur et aux fins détails spatiaux, ce qui rend l'image des objets placés dans l'axe du regard, au centre de notre champ visuel, beaucoup plus nette que celle des objets périphériques. Il faut noter que notre vision périphérique (assurée par les bâtonnets), bien que ne permettant pas d'accéder aux détails fins, possède de nombreuses qualités offrant notamment une excellente sensibilité aux mouvements ou une bonne vision scotopique (c'est-à-dire à faible luminosité).

Conjointement à cette différence de résolution de notre rétine, intervient la possibilité d'orienter notre regard, avec une rotation de l'œil, vers différentes parties de la scène. Ces saccades oculaires sont très rapides, fréquentes (nous faisons en moyenne trois saccades par seconde) et ne demandent pas d'effort. Ces mouvements, par exemple pendant la lecture, ont pour conséquence de placer au centre de la rétine la portion du champ visuel contenant le mot que nous lisons : ce mot est vu en détail mais le reste de la page sera flou. L'orientation du regard permet donc à la fois de traiter l'ensemble de l'espace visuel relativement sommairement et en détail une portion particulière de notre champ visuel.

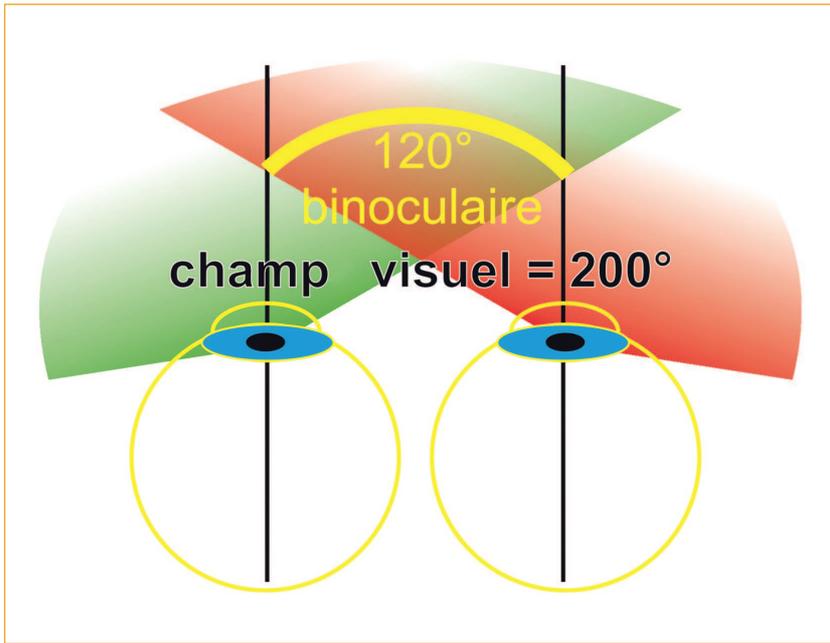
Une attention sélective

Cependant une sélection, que l'on appelle attention sélective, peut être faite même sans déplacer le regard. Pour ces deux types de sélection on utilise en fait les mêmes termes : on parle de mécanismes d'orientation, et on utilise respectivement les termes d'orientation implicite (si les yeux ne bougent pas) et orientation explicite (si une saccade est effectuée) de l'attention. L'hypothèse selon laquelle ces deux mécanismes seraient fortement liés (on parle de théorie prémotrice de l'attention) est de plus en plus largement acceptée. Chaque fois qu'une saccade oculaire est programmée

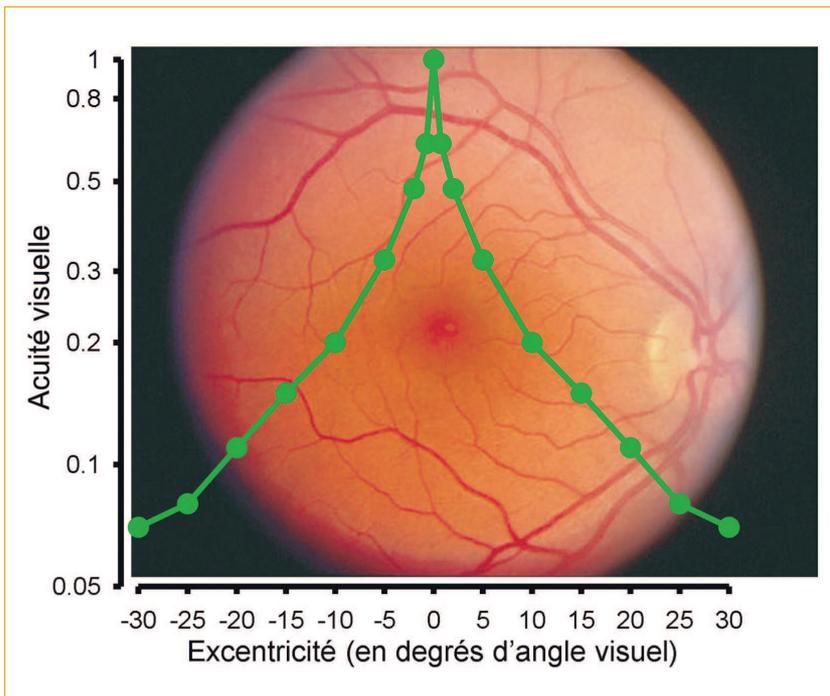
> Les bases anatomiques et fonctionnelles de la vision active.

L'acuité visuelle décroît très rapidement avec l'excentricité : ce qui n'est pas dans l'axe du regard apparaît flou comme sur l'image ci-dessous, filtrée avec un algorithme modélisant la vision humaine.





© LAURENT MADELAÏN



© LAURENT MADELAÏN



vers une cible périphérique, notre perception visuelle en résulte modifiée de manière importante avant même que ce mouvement ne soit commencé. Pendant un bref intervalle de temps (environ 200 millisecondes) précédant la saccade, notre performance visuelle à la future position du regard s'améliore considérablement relativement aux autres régions de l'espace : un déploiement de l'attention visuelle sélective précède celui du regard. Il s'agit d'un effet automatique, que l'on ne peut que très partiellement influencer volontairement et dont les bases neuronales ont été beaucoup étudiées. D'autres effets d'interférence ont été observés et les études d'imagerie ont confirmé la superposition des cartes d'activation cérébrale pendant des tâches d'orientation implicite de l'attention visuelle et pendant des saccades. Mais l'histoire est encore plus compliquée et intéressante : juste avant l'exécution d'une saccade, toute la perception de l'espace change et les objets sont perçus comme s'ils étaient plus proches de la cible de la saccade que ce qu'ils ne le sont vraiment. Récemment, l'interaction complexe entre les mécanismes de sélection oculomotrice et attentionnelle a été étudiée non plus pour des portions de l'espace mais pour des caractéristiques des objets : par exemple il a été montré que la programmation de saccades vers une cible d'une certaine couleur entraîne une facilitation dans le traitement visuel des objets de cette même couleur indépendamment de leur position spatiale.

Il reste à comprendre pourquoi exactement ces modulations complexes sont mises en place au niveau perceptif autour de l'exécution d'une saccade. Une possibilité est que les changements physiologiques et perceptifs qui ont été observés servent à garantir la stabilité du monde visuel qui serait autrement mise en cause par le déplacement continu du regard. Notre cerveau anticiperait et adoucirait en quelque sorte les changements abrupts que chaque saccade implique : une façon de faire cela est de « réorganiser » (*remapping* en anglais) le monde perçu ou au moins les informations visuelles les plus saillantes. Ces considérations nous laissent voir à quel point une caméra vidéo, même à résolution variable et équipée d'un mécanisme d'orientation intelligent et rapide, n'est pas un bon modèle de notre vision active, qui, elle, présente une complexité bien plus importante.

Il est toutefois remarquable que, dans un certain nombre de situations, ces mécanismes d'orientation ne soient pas nécessaires à la vision. C'est le cas, par exemple, lorsqu'un objet se distingue des autres par une seule propriété telle que la couleur ou la forme. S'il est effectivement difficile de trouver une aiguille dans une botte de foin, l'aiguille partageant avec le foin un grand nombre de caractéristiques, il est aisé de distinguer une fleur de coquelicot dans un champ de blé, la couleur rouge de la fleur étant une propriété visuelle qui ne se trouve pas dans le blé. Cet effet connu sous le nom de « saute aux yeux » ●●●

●●● Un phénomène d'aveuglement inattentionnel

(on parle d'effet *pop-out* en anglais) a permis de mettre en évidence une vision préattentionnelle qui ne nécessite pas de focalisation pour la détection d'un élément : la présence d'un cercle bleu est immédiatement détectée parmi des cercles verts. Si, par contre, l'élément cible et les items qui l'entourent partagent deux propriétés ou plus, il devient nécessaire d'orienter l'attention, et bien souvent le regard, sur chaque élément présent : le temps de détection de la présence d'un cercle bleu entouré de cercles verts et de carrés bleus dépendra ainsi du nombre total d'éléments présents.

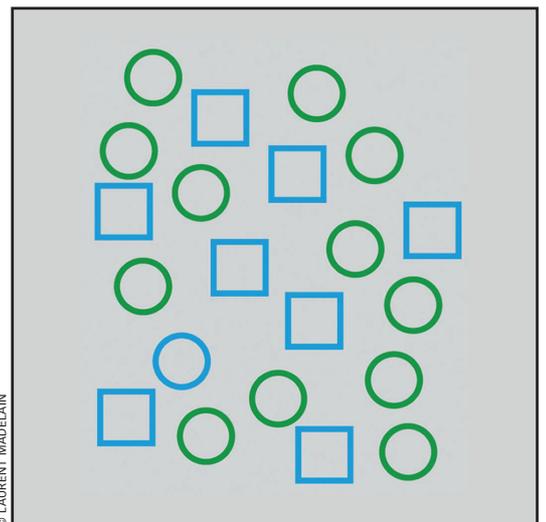
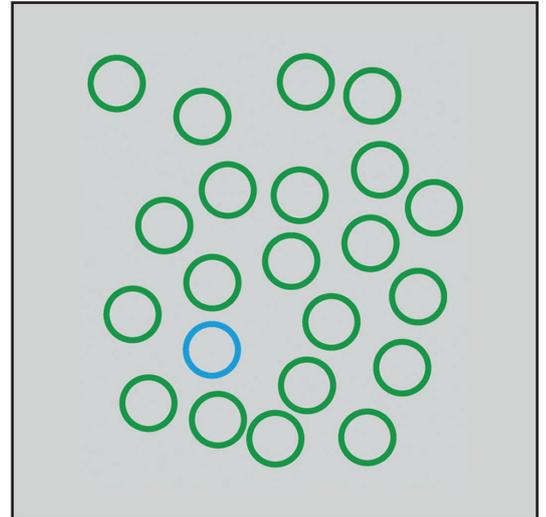
A contrario, il est possible de démontrer que l'orientation de l'attention rend « aveugle » à certains événements visuels. Ainsi le phénomène bien établi d'aveuglement inattentionnel consiste à focaliser explicitement l'attention sur des éléments d'une scène visuelle et à en ignorer d'autres. Dans cette situation, les sujets ne perçoivent pas un événement inattendu de la scène alors même que si l'on ne leur demande pas de focaliser leur attention sur une tâche particulière ils sont parfaitement capables de le voir. Dans une expérience célèbre, le professeur de psychologie Daniel Simons demande aux sujets de compter le nombre de fois où des personnes habillées de blanc se passent des ballons tout en ignorant des personnes habillées de noir se passant également des ballons. Lorsqu'au beau milieu de la tâche un étudiant déguisé en gorille traverse la pièce, la majeure partie des sujets ne le voit pas. On peut citer un autre type d'aveuglement : l'aveuglement au changement, qui se manifeste par exemple lorsque deux images présentées séquentiellement diffèrent par un détail. Il est souvent très difficile de déterminer quel est le détail en question, alors même que, si les sujets sont informés de la nature du changement, celui-ci est immédiatement détectable. Ce type d'aveuglement est bien connu des magiciens...

Bénéfices de la vision active

Les mécanismes d'orientation et de sélection sont au cœur de la vision active, mais peut-on considérer qu'ils sont opérants ? Jusqu'à quel point les utilisons-nous de manière efficace pour optimiser le traitement de l'information visuelle ? Dans certaines situations nous sommes « victimes » d'une orientation involontaire : c'est le cas de la capture attentionnelle que l'on observe quand un stimulus se distingue de manière particulière du reste de la scène ou alternativement quand il apparaît soudainement dans une scène par ailleurs statique. Dans ce cas, pendant un court laps de temps, nos mouvements des yeux (et/ou notre attention) ont tendance à être dirigés vers ce stimulus, ce qui implique une baisse de performance dans le traitement d'autres objets ou régions de l'image.

> La vision préattentionnelle.

Sur l'image du haut, le rond bleu saute aux yeux, alors que sur l'image du bas il faut regarder chaque objet avant de décider si un rond bleu est effectivement présent.



Le phénomène de capture attentionnelle peut être surprenant par sa force

© LAURENT MADELAIN

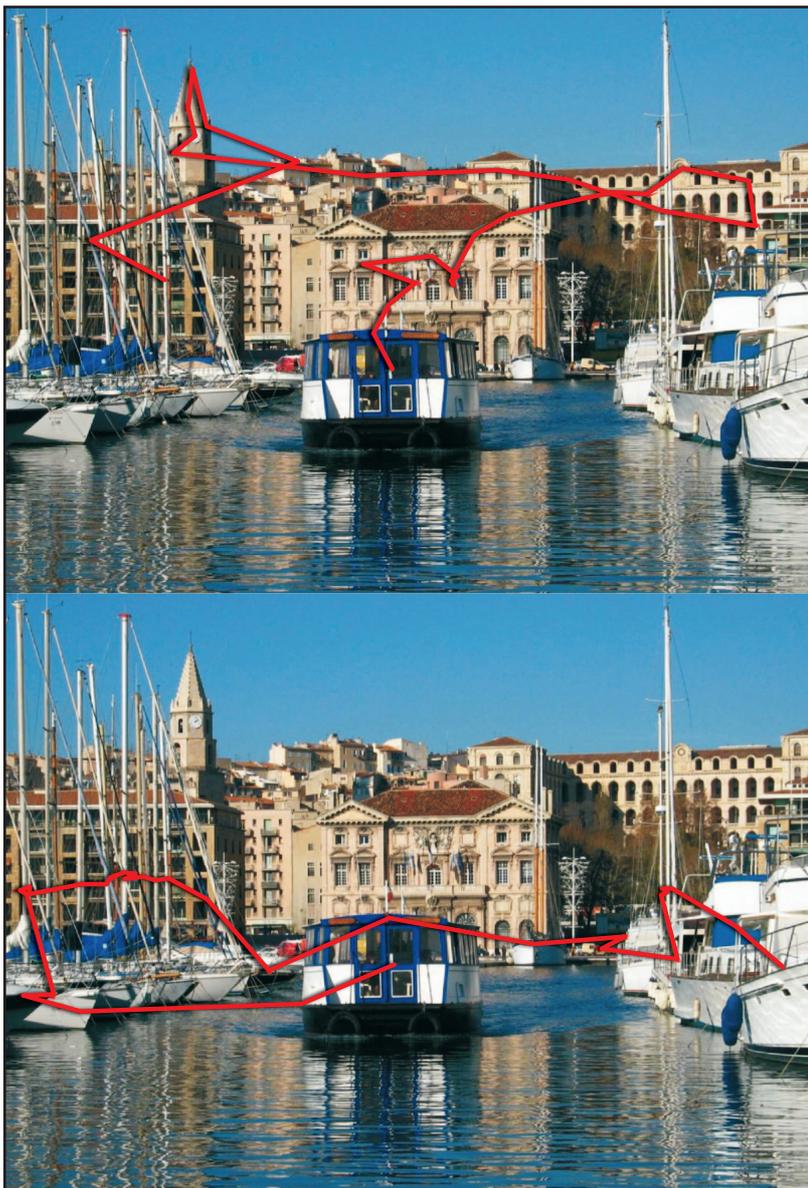
Le phénomène de capture peut être surprenant par sa force et la difficulté à l'inhiber. On pourrait d'ailleurs avoir l'impression que cette orientation automatique n'est pas du tout adéquate et qu'elle nous distrait des objets importants à traiter. Il faut cependant considérer que dans nombre de situations les objets saillants et la nouveauté, par exemple l'apparition soudaine d'un camion qui se dirige vers nous, constituent effectivement des événements informatifs qui méritent d'être analysés de manière sélective. Certains auteurs ont montré que notre regard et notre attention peuvent aussi très bien être capturés par un objet dont les caractéristiques physiques ne sont pas forcément saillantes s'il est porteur d'un message incongru au niveau sémantique avec le reste de la scène : l'image d'un iguane dans un salon de la même couleur pourrait donc attirer notre regard plus rapidement que d'autres objets plus « attendus » dans ce contexte.

La recherche visuelle

Comment les mécanismes actifs de notre vision se mettent-ils en place lors de l'exploration d'une scène complexe contenant plusieurs éléments qui diffèrent dans leur apparence ainsi que dans leurs fonctions, attractivité, intérêt ? Cela reste une des questions majeures du domaine. Si le rôle de l'information cognitive dans le contrôle de l'orientation du regard n'est désormais plus en discussion, il reste que nous ne sommes toujours pas en mesure de prédire comment l'information visuelle et l'information cognitive interagissent dynamiquement dans la sélection active des points de fixation (voir ci-dessous). Le concept de cartes de saillance, issu du domaine de la vision artificielle, postule que les différentes propriétés visuelles à travers toute l'image (telles que la couleur, la luminosité, l'orientation des bords, le mouvement) sont codées indépendamment, cartographiées et normalisées, puis combinées

▼ L'exploration de scènes visuelles.

Deux séquences de fixations différentes, après l'instruction « Identifier la ville où la photo a été prise » (en haut) et « Rapporter le nombre de bateaux dans l'image » (en bas).



en une carte unique à partir de laquelle on extrait une série de points en ordre décroissant de saillance. D'autres propriétés peuvent enrichir ce modèle assez simpliste de cartes de saillance, telles que les caractéristiques anatomiques du système visuel (la résolution hétérogène de la rétine) ou encore des informations de plus haut niveau sur le contexte (par exemple « c'est l'image d'une cuisine »). L'hypothèse qui guide ces travaux est qu'il doit être possible ainsi de prédire l'exploration d'une scène visuelle par un être humain en se basant principalement sur les caractéristiques physiques de la scène.

Certaines études suggèrent plutôt que la séquence de mouvements oculaires lors de l'exploration d'une scène visuelle correspond à une séquence de « questionnements » (tels que « S'agit-il de l'image d'un port ? ») et de tentatives de réponse partielle (une fixation sur un objet vertical pour valider l'hypothèse que ce soit un mât de voilier). Dans des situations moins naturelles, typiques de l'expérimentation dans les laboratoires, il a été montré que notre cerveau peut coordonner le système visuel-oculomoteur de manière proche de l'optimal (dans son acception mathématique) dans le but de maximiser après chaque mouvement la probabilité d'acquiescer de l'information importante sur la présence et la localisation de l'objet recherché. Cependant la variabilité des enregistrements oculaires obtenus au cours de l'exploration visuelle de sujets différents qui observent la même scène, ou d'un observateur unique qui répète l'exploration de la même scène plusieurs fois, révèle que nous sommes loin d'être des robots qui exécutent une série de fixations strictement déterminée par des critères quantitatifs et « optimaux ».

Le comportement de vision active constitue un élément essentiel de nos capacités visuo-motrices et présente une grande sensibilité aux caractéristiques de l'environnement. En outre, les comportements visibles associés à l'orientation de l'attention chez une personne que l'on observe, tels que les mouvements des yeux, de la tête ou du tronc, voire les mouvements du bras tel le pointage en direction d'un objet, constituent autant de stimulus que nous sommes à même d'exploiter pour orienter notre propre attention en direction du même objet. Cette capacité, nommée attention conjointe, dépasse le cadre strict de la vision active mais constitue un exemple frappant de l'importance des comportements sociaux pour la perception visuelle (pp. 26-27). ●

SAVOIR +

- L'expérience de Simons portant sur l'aveuglement inattentionnel : www.theinvisiblegorilla.com/gorilla_experiment.html
- Exemples de stimulus illustrant l'aveuglement au changement : www2.psych.ubc.ca/~remsink/flicker/index.html