

MEMOIRES ORIGINAUX

*Laboratoire de Psychologie Expérimentale
Université René-Descartes, Paris 5
CNRS UMR 8581¹*

L'influence des facteurs sémantiques sur la cécité aux changements progressifs dans les scènes visuelles

Malika Auvray² et J. Kevin O'Regan

SUMMARY : Influence of semantic factors on blindness to progressive changes in visual scenes.

Several studies have shown that under some circumstances observers have difficulty detecting changes that occur in successive views of a visual scene. Recent work has confirmed this finding in situations where the change, instead of occurring abruptly, is made gradually but in full view.

This work suggests that observers only have access to a small portion of the information available in a scene. The question arises, what determines which information will be accessed?

The experiments reported here first confirm that change blindness can be obtained using progressive changes in full view. Second, they investigate the influence on change detection of two semantic factors: "interest" of scene elements and their "consistency" within the scene.

We show that change detection improves as a function of the "interest" of an element and as a function of its degree of "inconsistency" within the scene.

The general discussion provides possible explanations of these results and their implications to the study of the nature of our representations.

Key words : vision, change blindness, semantic factors.

¹ Institut de Psychologie, Centre Universitaire de Boulogne, 71 Avenue Edouard Vaillant, 92724 Boulogne-billancourt Cedex.

² E-mail : malika.auvray@univ-paris5.fr

Introduction

Notre expérience visuelle immédiate semble être celle d'un monde riche, détaillé et cohérent dans lequel tous les objets se présentent à nous de manière synchrone. De la vivacité et de l'indubitabilité supposée de cette expérience, nous inférons les propriétés de richesse et de cohérence, que nous attribuons non seulement au monde externe mais aussi à un monde interne, dans lequel sont censées se constituer des répliques plus ou moins identiques aux objets perçus. Au vu de ce qu'est notre perception consciente du monde visuel, nous pouvons alors penser que la perception d'une scène consiste en la possession d'un modèle interne détaillé de celle-ci. Cette idée trouve son expression prototypique dans le travail de Marr (1982) qui considère qu'un des problèmes de la vision est de reconstituer à partir des images rétiniennes bi-dimensionnelles un « modèle interne » tri-dimensionnel. Un certain nombre de travaux semblent suggérer que nous formons rapidement une représentation précise des scènes perçues. Il est ainsi possible de reconnaître un grand nombre d'images même si elles sont vues une seule fois (Shepard, 1967 ; Standing, 1973 ; Standing, Conezio et Haber, 1970).

Cependant, un ensemble de travaux sur la mémoire visuelle à court terme tend à remettre en question cette conception de la perception. Cette ligne de recherche, entamée dans les années 70, montre que les capacités de la mémoire visuelle à court terme sont limitées : seule une petite partie des informations présentes dans la scène visuelle est stockée et utilisable dans des tâches de restitution ou de comparaison (Sperling, 1960 ; Phillips, 1974 ; Coltheart, 1980, 1983 ; Haber, 1983 ; Pashler, 1995). Un ensemble de travaux suggère en complément que ces informations seraient stockées sous forme abstraite, selon un code non visuel (pour une revue, voir Irwin et Andrews, 1996).

Ceci remet en cause l'idée que notre perception d'une scène visuelle s'effectue par la médiation de l'intégration de tous les éléments de la scène dans un hypothétique « buffer visuel », ou modèle interne (Rayner et Pollatsek, 1983). Mais si nous ne stockons pas

l'intégralité des détails d'une scène visuelle, comment expliquer alors notre impression phénoménologique d'un monde visuel interne si riche ?

Une possibilité serait de considérer notre impression d'une présence visuelle riche comme une sorte d'illusion. Cette illusion s'expliquerait par l'existence d'un mécanisme spécial dans le système visuel, qui nous rend particulièrement sensibles aux signaux transitoires visuels (O'Regan, 1992 ; O'Regan et Noë, 2001). Ces signaux sont des changements abrupts de luminance ou de couleur qui accompagnent toute modification au sein du champ visuel. Ils sont immédiatement détectés et traités par des mécanismes de bas niveau du système visuel. Les signaux transitoires peuvent capter automatiquement l'attention du sujet et l'attirer sur le lieu du changement et ce, de manière exogène, c'est-à-dire indépendamment des buts ou intentions du sujet (Folk, Remington et Johnston, 1992 ; Yantis, 1993 ; Yantis et Egeth, 1999 ; Yantis et Jonides, 1990).

La cécité aux changements.

Un moyen de vérifier que les changements abrupts de luminance ont pour fonction d'attirer l'attention à l'emplacement du changement et d'en favoriser ainsi la détection a été d'utiliser, au cours d'expériences dites de « cécité aux changements », des protocoles qui permettent de noyer les signaux transitoires accompagnant le changement.

Une cécité aux changements a pu ainsi être induite lorsque le changement survient durant une saccade (Rayner et Pollatsek, 1983 ; Irwin, 1991 ; Henderson, 1997) ou un clignement de paupières (O'Regan, Deubel, Clark et Rensink, 2000), lorsqu'un écran blanc (« flicker ») est inséré entre l'image originale et l'image modifiée (Pashler, 1988 ; Simons, 1996 ; Rensink, O'Regan et Clark, 1997), lorsque des perturbations locales consistant en 6 petits rectangles ou ovales noirs et blancs (« mudsplashes ») sont dispersés à travers l'image au moment du changement sans pour autant recouvrir le lieu du changement (O'Regan, Rensink

et Clark, 1999), lors d'un changement de plan dans une séquence de film (Hochberg, 1986 ; Levin et Simons, 1997 ; Simons, 1996) ou encore lors d'une «interruption dans le monde réel» (Simons et Levin, 1998).

Ceci permet de déterminer que, lorsque les changements dans une scène visuelle ne peuvent plus capter automatiquement l'attention, les sujets ont beaucoup de difficultés à les détecter.

La cécité aux changements lents.

L'ensemble de ces études semble donc montrer une cécité aux changements lorsque ceux-ci surviennent entre deux vues successives de la même scène. Il est néanmoins possible de soulever une objection. Toutes ces études sur la cécité aux changements impliquent une interruption de la vision de la scène lors du changement. La cécité obtenue ne pourrait-elle pas être due à cette rupture dans la continuité de la vision de la scène ? L'interruption ne pourrait-elle pas engendrer un effacement ou une réinitialisation des informations contenues dans une représentation interne du monde visuel externe ? Pour évaluer cette possibilité, un nouveau protocole de recherches sur le sujet a été récemment mis en place : on y effectue un changement progressif dans un élément de la scène afin de produire une cécité sans aucune interruption de la vision de la scène.

Dans une étude exploratoire, Chabrier (1999) utilise sept images dessinées dans lesquelles il modifie très lentement un aspect de l'image. Le changement peut concerner les propriétés d'un objet, sa structure, sa couleur, sa dimension, son orientation ou bien l'objet lui-même. Précisons que le changement apparaît évident aux yeux du sujet si on lui présente successivement l'image originale et l'image finale. Les résultats montrent que le changement progressif est identifié entre 10% et 27% des cas. Ces expériences sont donc révélatrices d'une forte cécité aux changements progressifs dans les images dessinées. Simons, Franconeri

et Reimer (2000) confirment ces résultats avec des photographies de scènes naturelles : 61% des changements sont identifiés s'ils consistent en une apparition ou une disparition et 29 % des changements sont identifiés s'ils consistent en un changement de couleur. Ces résultats préliminaires semblent donc confirmer une cécité aux changements progressifs.

Facteurs déterminant la performance.

A travers l'ensemble des résultats de ces différents paradigmes, nous constatons que les observateurs parviennent difficilement à identifier un changement, même si ce dernier occupe une grande partie de l'image et que les observateurs savent qu'il va se produire. Ceci suggère qu'en l'absence d'une attention soutenue, les représentations mises en œuvre aux étapes précoces du traitement de l'information sont éphémères et insuffisantes pour identifier un changement (Rensink et al., 1997). Si l'attention est amenée sur le lieu du changement par un signal transitoire généré par une transformation dans l'image, le changement sera vu, sinon il demeurera inaperçu. En effet, en l'absence de signaux transitoires de mouvement, l'observateur n'a aucun indice provenant de mécanismes de bas niveau sur le changement. L'identification de ce changement, puisqu'elle ne peut plus être ascendante, va par conséquent requérir un examen sériel de l'image, élément par élément. Le temps moyen d'identification sera alors nécessairement long. L'identification plus rapide de tel ou tel élément résulterait de l'attraction de l'attention par des mécanismes de traitement de l'information descendants. Les différents facteurs favorisant la perception et la mémorisation de certains aspects de l'image peuvent être des facteurs physiques, comme la saillance visuelle (taille, couleur, emplacement, contraste...), sémantiques (intérêt, cohérence...), contextuels ou encore liés aux préférences subjectives des sujets (O'Regan et al. 2000).

Parmi ces différents facteurs, les expériences proposées ici se focaliseront sur les facteurs sémantiques. L'objectif de ces expériences va être d'étudier quelle peut être

l'influence de deux facteurs sémantiques particuliers sur la perception. Nous reprendrons la variable appelée «**intérêt**», déjà explorée dans le paradigme des « Flickers » (Rensink et al., 1997), puis nous étudierons l'effet de la «**cohérence**», qui n'a pas été approfondi jusqu'ici dans le cadre de la cécité aux changements. Cette étude s'effectuera au moyen des techniques propres à la cécité aux changements progressifs, qui présente l'intérêt de n'impliquer aucune perturbation lors de la vision de la scène.

Le but de ces expériences est double : à la fois confirmer une cécité aux changements progressifs et déterminer si des facteurs sémantiques comme l'intérêt et la cohérence des éléments influent sur la détection des changements. Ceci a pour enjeu de mieux cerner la nature de nos représentations. En effet, des résultats confirmant une cécité aux changements progressifs et une différence véritable d'identification des changements en fonction de facteurs sémantiques cautionneront une théorie de la perception dépendante de l'attention. Si nous trouvons une telle différence, cela peut impliquer l'idée que toutes les informations présentes dans une scène visuelle ne sont pas encodées de manière équivalente et ne se présentent pas à nous de manière synchrone. Mais au contraire que nous ne percevons que le contenu des éléments sur lesquels se porte notre attention et ce, en fonction de leur intérêt et de leur cohérence.

Expérience 1 : le facteur intérêt.

Les expériences de cécité aux changements utilisant des « flickers », des clignements de paupières ou des « mudsplashes » ont établi qu'un changement concernant un élément «d'intérêt central» se détecte plus facilement qu'un changement concernant un élément «d'intérêt marginal» et ce, même si les deux éléments sont de saillance visuelle comparable. Rensink et al. (1997) définissent de manière opérationnelle la notion d'intérêt. L'intérêt d'un élément est déterminé lors d'une expérience indépendante et préalable. Il est demandé à

plusieurs observateurs de décrire brièvement chaque scène. Les éléments mentionnés par plus de la moitié des observateurs sont considérés comme étant d'intérêt central et ceux qui ne sont mentionnés par aucun sont considérés comme étant d'intérêt marginal.

L'expérience présentée ici a pour but de vérifier les résultats de Rensink et al. avec la technique des changements progressifs et concernant des éléments qui occupent une plus grande proportion de la surface de l'image. Pour pouvoir effectuer une comparaison entre des éléments de taille importante et équivalente, la définition d'intérêt central et d'intérêt marginal que nous avons utilisée diffère quelque peu de celle de Rensink et al. L'intérêt d'un élément a été déterminé lors d'une expérience préalable. Il a été demandé à cinq observateurs de décrire brièvement chaque scène. Un élément est considéré comme étant d'intérêt central si, lorsque les sujets décrivent l'image, ces derniers citent cet élément en premier. Un élément est considéré comme étant d'intérêt marginal si aucun sujet ne le cite en première position.

Méthode.

Matériel. Toutes les séquences de changements progressifs sont composées d'une image originale A et d'une image modifiée A'. L'image A' consiste en la modification d'un élément de l'image A. Les images sont présentées selon la séquence A, A', A : l'image A se transforme progressivement en l'image A' puis revient progressivement à l'image initiale. Afin d'éviter tout signal transitoire pouvant attirer l'œil, cette lente modification s'effectue sur plus d'une minute. Cette séquence aller-retour, que nous appellerons séquence ou animation, est présentée une seule fois. Nous avons choisi de réaliser uniquement des changements de couleur. Les modifications consistent donc en un changement de couleur qui peut concerner aussi bien une petite partie de l'image que sa totalité.

Afin de tester l'influence de la rapidité de la modification, les mêmes séries d'animations sont présentées à deux ensembles de sujets à deux vitesses différentes. Pour la

première moitié des sujets, la séquence AA'A a une durée de 400 s et pour la seconde moitié des sujets, la même séquence a une durée de 100 s.

Les changements dans les images ont été réalisés avec le logiciel Adobe Photoshop et animés avec le logiciel Macromedia Flash selon la technique de la transparence de calque. La transition visuelle s'opère donc comme un fondu. Les images, présentées sur un écran d'ordinateur, mesurent 20 cm sur 27 cm et ont une résolution de 72 pixels/pouce.

Procédure. Les sujets sont informés des types de changements possibles. Ils savent que ces derniers consistent en un changement de couleur et peuvent se produire dans n'importe quelle partie de l'image. Les sujets doivent localiser le changement aussi rapidement que possible. Ils doivent identifier l'élément qui a subi la modification et l'indiquer aussitôt. Ils ajustent la position de l'écran de la manière qu'ils considèrent la plus appropriée et se tiennent à la distance de l'écran qu'ils choisissent, en moyenne entre 30 et 60 cm. L'ordre de présentation des animations est aléatoire. Le temps entre le début de la présentation et l'identification de l'élément modifié est chronométré. Lors de chaque animation, tous les commentaires du sujet sont notés. Si, à la fin de la présentation, le sujet n'a pas découvert le changement, la succession de l'image originale et de l'image modifiée lui est montrée.

Participants. 24 sujets issus de milieux divers et âgés de 18 à 55 ans passent les deux séries d'expériences. 12 sujets passent les ensembles d'expériences de la série lente, et 12 autres sujets passent la même série d'expériences mais de la série rapide.

Images. Le matériel de l'expérience est constitué de 12 animations, divisées en 6 paires d'animations. Chaque paire d'animation est divisée en deux sous-ensembles. Dans le premier de ces sous ensembles, un élément apparaît comme étant d'intérêt central ; dans le second de

ces sous-ensembles, le même élément apparaît comme étant d'intérêt marginal. Pour ce faire, la seconde animation est simplement constituée de la première animation sur laquelle on a ajouté un ou plusieurs éléments qui constituent des centres d'intérêt plus fort (voir annexe 1). L'élément considéré est ainsi, autant que possible, physiquement identique dans chaque paire. Par conséquent, on ne pourra objecter que les performances du sujet sont influencées par des facteurs visuels de bas niveaux comme la taille ou l'emplacement (central ou excentré) de l'élément qui subit la modification (Hollingworth et Henderson, 2000).

Les images sont présentées aux sujets de telle manière que chaque sujet voie une seule image de chaque paire, mais de telle sorte que toutes les images aient été vues, dans toutes les conditions, une fois que tous les sujets ont passé le test.

Résultats et discussion

Si le sujet n'a pas identifié le changement à la fin de la présentation, ce dernier se voit attribuer comme temps la durée totale de la présentation, à savoir 100 s pour la version rapide et 400 s pour la version lente.

La figure 1 présente les moyennes globales de la proportion du temps total pris pour identifier le changement à travers l'ensemble des sujets et des images.

Pour rendre compatibles les temps d'identification pour les deux conditions de vitesse, nous avons renormalisé les temps de la manière suivante : les temps pour la condition de présentation rapide ont été divisés par 100 et les temps pour la condition de présentation lente ont été divisés par 400.

L'influence du facteur intérêt est présente pour les deux conditions de vitesse et pour la majorité des images, comme le confirme la figure 2. L'effet du facteur intérêt est significatif dans l'analyse de variance par sujet [$F(1,22) = 12.94, p = 0.002$] et est marginalement significatif dans l'analyse de variance par images [$F(1,4) = 7.47, p = 0.052$].

Le facteur vitesse semble ne pas avoir d'influence sur l'effet du facteur intérêt. Ce fait peut signifier que la vitesse rapide ne laisse subsister aucun signal transitoire et qu'au-delà de cette limite, la durée de l'animation n'a pas d'influence sur les autres facteurs.

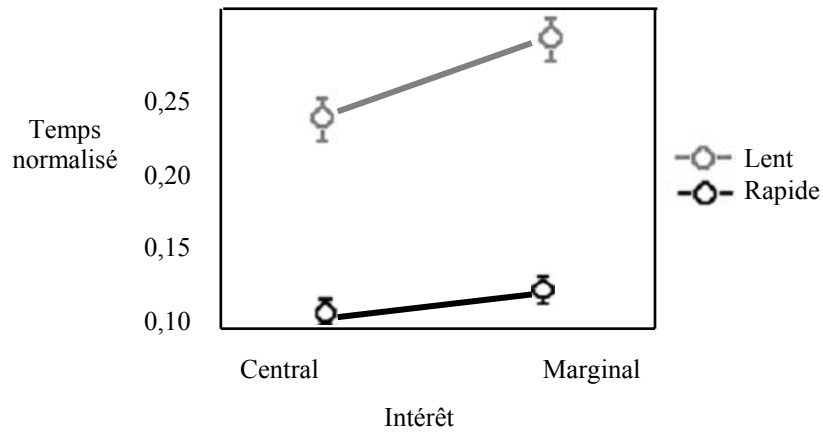


Fig 1- Temps moyen, exprimé comme proportion de la longueur de la séquence, pris par les sujets avant d'identifier le changement selon son intérêt, dans la condition de présentation lente (en bleu) et dans la condition de présentation rapide (en rouge).

La normalisation a consisté en la division du résultat par le temps total de la présentation, c'est-à-dire 400 s pour la condition de présentation lente et 100 s pour la condition de présentation rapide.

Les moyennes sont prises sur l'ensemble des sujets et des images.

Les barres d'erreur indiquent un écart type au dessus et en dessous de la moyenne.

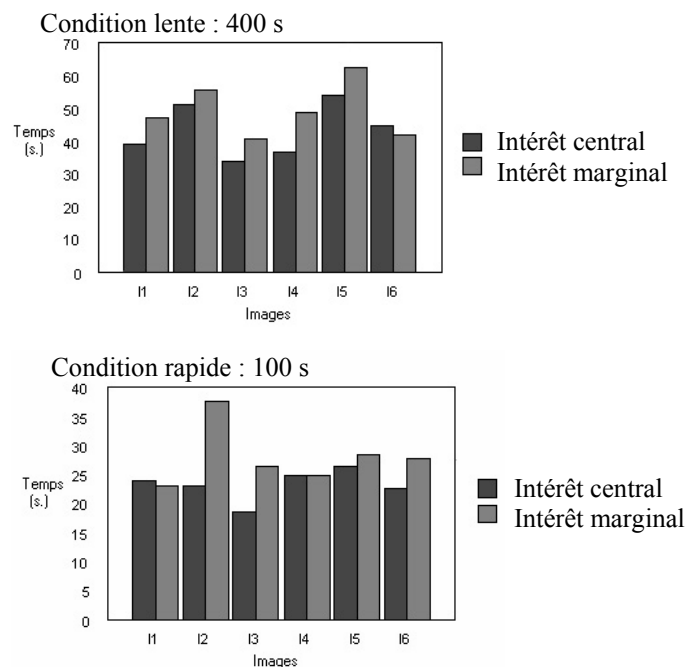


Fig 2. Temps moyen pris par les sujets avant de détecter le changement, selon son intérêt, en fonction des images, dans la condition de présentation lente et dans la condition de présentation rapide, pour les éléments d'intérêt central (en bleu) et pour les éléments d'intérêt marginal (en rouge). Les moyennes sont prises sur l'ensemble des sujets.

Dans le cadre d'un changement progressif, un changement survenant dans un aspect d'intérêt central de la scène est donc identifié plus rapidement que s'il survient dans un aspect d'intérêt marginal et ce, même si les deux éléments sont de saillance visuelle identique et importante. Ce résultat est en conformité avec ceux obtenus par les paradigmes des « flickers » et des « mudsplashes ».

Expérience 2 : le facteur cohérence.

Il est établi que le contexte, nos connaissances et nos attentes ont une influence sémantique sur la perception des objets dans une scène du monde réel. La cohérence fait partie des facteurs sémantiques influant sur la perception. Comme définition opérationnelle du concept de cohérence, nous pourrions dire qu'un objet est en cohérence avec son contexte lorsqu'il est habituel ou fréquent pour le contexte considéré, c'est à dire qu'il est enclin à apparaître dans ce contexte. Un objet incohérent avec son contexte, à l'inverse, sera un objet inhabituel pour le contexte considéré, soit parce qu'il est peu enclin à apparaître habituellement dans la scène (Hollingworth et Anderson, 2000, placent par exemple une bouche d'incendie au milieu d'un salon), soit parce que l'objet considéré est peu habituel en soi (par exemple, une rose bleue). La question de l'influence de la cohérence d'un objet sur les performances du sujet se pose dans les termes suivants : est-ce qu'un objet inhabituel dans une scène naturelle fera l'effet d'un «pop-out» en attirant immédiatement l'attention du spectateur (Christie et Klein, 1995 ; Hollingworth et Henderson, 2000) ou, à l'inverse, est-ce que nous sommes davantage sensibles à la cohérence entre les traits d'un objet et son contexte ? Selon cette dernière hypothèse, une scène visuelle est rapidement associée à une catégorie sémantique et les prédictions découlant de cette catégorie permettent d'identifier l'objet par la suite (Friedman, 1979 ; Antes et Peland, 1981 ; Biederman, Mezzanotte, Rabinowitz, Francolini et Plude, 1981). Deux thèses différentes sont donc possibles : d'un

côté une facilitation de la perception amenée par un schéma des objets compatibles avec le contexte et de l'autre un pop-out perceptuel des objets incompatibles avec la scène.

Il semble cependant que ces deux approches ne soient pas mutuellement exclusives mais que leur pertinence dépende de la tâche à accomplir par le sujet. L'influence de la cohérence d'un objet peut varier selon qu'il s'agit de détection d'un objet, d'attention portée à l'objet, d'identification et de catégorisation de l'objet ou de reconnaissance mémorielle de l'objet. Friedman (1979) distingue le processus de détection de traits du processus d'analyse de traits. Le processus de détection de traits est facilité pour les objets en cohérence avec la scène par l'activation du cadre de traitement spécifique à la scène. En revanche, le processus d'analyse de traits, parce qu'il inclut un processus qui distingue la scène présentée des autres scènes, opère selon un schéma d'items incohérents. Ceci a pour conséquence que l'apparence physique des objets compatibles avec la scène est encodée plus en détail que l'apparence physique des objets incompatibles avec la scène. Cette distinction peut expliquer que, lors d'une tâche de reconnaissance mémorielle, les performances sont meilleures pour les objets incompatibles avec la scène alors que la perception et la compréhension d'une scène, à un niveau pré-attentionnel tout du moins, est largement guidée par un schéma lié à la familiarité relative des éléments composant habituellement la scène.

L'expérience proposée explore la nature de l'interface entre les connaissances et l'attention dans le cadre restreint de la cécité aux changements lents. La scène de départ ne comportant pas d'objets incohérents mais seulement susceptibles de le devenir, la question ne se pose pas de savoir quelle est l'influence de la cohérence sur l'encodage des éléments mais de déterminer si le processus attentionnel est influencé par ce facteur.

Méthode

L'appareillage, la procédure et les participants sont identiques à l'expérience précédente.

Images. Cette expérience est constituée de 12 animations, divisées en 6 paires d'animations. Dans une même image de départ, un élément a été modifié de deux manières différentes. L'une des modifications donne lieu à une couleur habituelle pour l'objet en question ; l'autre est considérée comme donnant lieu à une couleur inhabituelle, par exemple des yeux rouges (voir annexe 2). Afin de ne pas nous exposer à d'éventuelles objections, à savoir que l'identification plus rapide des objets qui ne sont pas en cohérence avec leur contexte n'est jamais attribuable qu'à une saillance visuelle plus importante de ces objets (Henderson et Holingworth, 1998), nous avons tenté de faire en sorte que la saillance visuelle des objets cohérents et incohérents soient comparables ⁽¹⁾.

Les animations sont présentées aux sujets de telle manière que chaque sujet voie une seule animation de chaque paire, mais de telle sorte que toutes les animations aient été vues, dans toutes les conditions, une fois que tous les sujets ont passé le test.

Résultats et discussion.

De même que pour l'expérience précédente, si le sujet n'a pas identifié la modification à la fin de la présentation, ce dernier se voit attribuer comme temps la durée totale de la présentation à savoir, 100 s pour la version rapide et 400 s pour la version lente.

La figure 3 présente les moyennes globales de la proportion du temps total pris pour identifier le changement à travers l'ensemble des sujets et des images.

Comme pour la première expérience, pour rendre compatibles les temps d'identification pour les deux conditions de vitesse, nous avons renormalisé les temps de la manière suivante : les temps pour la condition de présentation rapide ont été divisés par 100 et les temps pour la condition de présentation lente ont été divisés par 400.

⁽¹⁾ Ceci a été fait de façon approximative, le logiciel Adobe Photoshop ne permettant pas de quantifier précisément la saillance des objets.

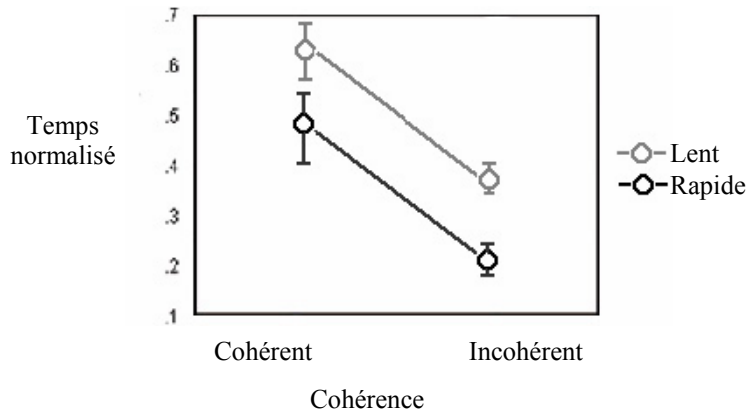


Fig 3- Temps moyen, exprimé comme proportion de la longueur de la séquence, pris par les sujets avant d'identifier le changement selon sa cohérence, dans la condition de présentation lente (en bleu) et dans la condition de présentation rapide (en rouge).

La normalisation a consisté en la division du résultat par le temps total de la présentation, c'est-à-dire 400 s pour la condition de présentation lente et 100 s pour la condition de présentation rapide.

Les moyennes sont prises sur l'ensemble des sujets et des images.

Les barres d'erreur indiquent un écart type au dessus et en dessous de la moyenne.

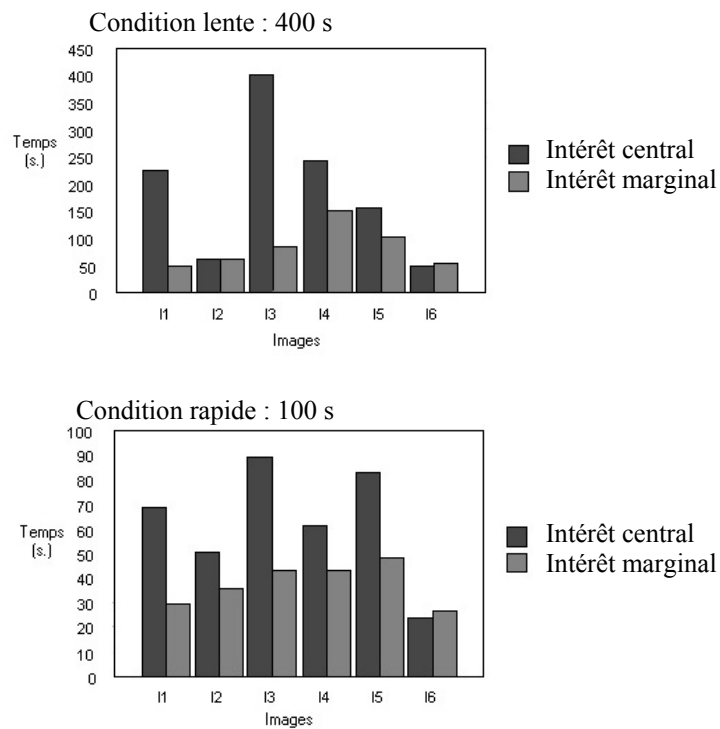


Fig 4. Temps moyen pris par les sujets avant de détecter le changement, selon sa cohérence, en fonction des images, dans la condition de présentation lente et dans la condition de présentation rapide, pour les éléments cohérents (en bleu) et pour les éléments Incohérents (en rouge). Les moyennes sont prises sur l'ensemble des sujets.

La détection d'un changement en cohérence avec l'ensemble de la scène est moins rapide que celle d'un changement qui ne l'est pas. Cette influence du facteur de cohérence est présente pour les deux conditions de vitesse et pour la majorité des images. La figure 4 le confirme. L'effet du facteur cohérence est significatif dans l'analyse de variance par sujets [$F(1,11) = 35.80, p < 0.0001$] et par images [$F(1,5) = 7.26, p = 0.043$].

Comme pour l'expérience précédente, le facteur vitesse semble ne pas avoir d'influence sur l'effet du facteur intérêt.

Les résultats montrent qu'un changement incohérent se voit plus rapidement qu'un changement cohérent. Le rapport verbal des sujets au cours de l'expérience permet d'appuyer ce résultat. Un nombre important de sujets a fait remarquer que dans le cas d'un changement non cohérent, ils ont été d'abord attirés sur le lieu du changement par l'incongruité de l'objet modifié. Puis, ils se sont dits que ce devait être cet objet qui a subi une modification sinon ils en auraient remarqué la couleur dès le départ. Et c'est dans un dernier temps seulement qu'ils essaient de se remémorer la couleur de départ. Il est intéressant de noter que dans les deux cas, les sujets parviennent difficilement à se souvenir de la couleur exacte de l'objet au début de la présentation. Ainsi, dans le cadre de l'identification d'un changement lent, le sujet remarque plus facilement un changement incohérent parce que l'objet ressort et capte son attention.

Ces résultats semblent contredire l'hypothèse du schéma selon laquelle la détermination de la catégorie d'une scène active un schéma qui fournit des informations descendantes pour faciliter l'identification des objets individuels de cette scène. Selon cette théorie, la connaissance des objets présents habituellement dans une scène donnée facilite la perception des objets en cohérence avec la scène. Mais cette catégorisation de la scène a lieu très tôt dans le traitement d'une nouvelle scène (Henderson, 1992) et la facilitation de l'identification des objets cohérents peut se restreindre à cette étape précoce d'identification de l'objet. Dans le

cadre de l'expérience effectuée, le changement d'un élément vers un élément cohérent ou incohérent avec la scène intervient une fois l'étape d'identification de l'objet achevée. Le but de notre expérience n'était donc pas de déceler dans quelle mesure l'encodage des éléments est influencé par leur cohérence au sein de la scène mais de déterminer si, une fois que les objets constitutifs de la scène sont identifiés, les processus attentionnels sont influencés par ce facteur sémantique. Les résultats confirment «l'hypothèse de l'attention» proposée par Henderson et Hollingworth et (1998) selon laquelle, une fois que l'objet a été identifié, l'attention se focalise préférentiellement sur les objets qui transgressent les contraintes imposées par le sens de la scène.

Discussion générale

Les résultats de ces deux expériences permettent à la fois de confirmer une cécité aux changements progressifs sur un échantillon important d'images et de sujets et de montrer l'influence des facteurs sémantiques sur l'identification des changements. La première expérience montre qu'un changement concernant un élément d'intérêt central est identifié plus rapidement qu'un changement concernant un élément d'intérêt marginal. Elle permet ainsi de confirmer que le résultat obtenu avec par exemple la technique des « flickers » (Rensink et al., 1997) et des « mudsplashes » (O'Regan, et al., 1999), se retrouve avec les changements progressifs. La seconde expérience montre qu'un changement donnant lieu à un élément incohérent est identifié plus rapidement qu'un changement donnant lieu à un élément cohérent. L'ensemble de ces résultats indique que deux facteurs sémantiques particuliers : l'intérêt des éléments et leur cohérence au sein de la scène, ont une influence sur l'identification des changements.

Les explications habituellement données sur la cécité aux changements impliquent des considérations sur la nature de nos représentations. Nous allons tenter de montrer quel peut être l'apport des résultats obtenus sur ces explications.

Le premier apport de ces expériences est de confirmer une cécité aux changements progressifs, ce qui permettra d'envisager les conséquences d'une telle cécité sans donner prise aux objections portées aux techniques de la cécité aux changements impliquant une interruption de la continuité de la vision de la scène. L'influence des facteurs sémantiques sur la cécité aux changements permettra d'appuyer la position théorique défendue.

Parmi les études menées sur la cécité aux changements, le phénomène de cécité obtenu est expliqué majoritairement par trois positions théoriques différentes. Dans ces explications, l'idée d'une représentation mentale de nature analogique, c'est à dire identique au percept, est remise en cause, soit en en faisant une entité incapable d'expliquer l'intégralité de notre expérience visuelle, soit en l'affaiblissant, soit en l'éliminant complètement.

Une représentation de nature analogique.

On conserve l'idée d'une représentation analogique ou iconique, mais insuffisante à l'explication du caractère continu et détaillé de notre expérience visuelle. Cette insuffisance peut avoir quatre origines possibles. Toute détection d'un changement semble être inférée à partir de tests de comparaison entre deux représentations d'un même objet. Il ne peut y avoir comparaison si l'un des termes ou les deux manquent. Suivant la classification de Simons (2000), on peut ainsi envisager que :

- Soit la première image manque : on considère que le stimulus qui précède le changement est simplement remplacé par le stimulus suivant.

- Soit la seconde image manque, auquel cas seuls les traits de l'objet initial ont été encodés et on échoue à encoder les traits de la scène modifiée.
- Soit les deux images sont retenues mais recombinaées, le produit du rapprochement n'est alors nullement la détection d'un changement mais seulement la création d'une nouvelle représentation qui combine différents traits de chacune des deux images originaires distinctes.
- Soit les deux images sont retenues mais maintenues séparées : une représentation de chaque vue est effectuée mais elles sont maintenues séparées de telle sorte que l'observateur ne peut avoir conscience de leurs différences.

Qu'il s'agisse de l'une ou l'autre de ces quatre alternatives, une explication de la cécité aux changements en termes de représentation analogique stricte, c'est-à-dire contenant l'intégralité des éléments de la scène perçue, pose problème lorsqu'il s'agit de rendre compte de l'influence des facteurs sémantiques sur la détection des changements. Si nous concevons la perception comme accès immédiat à l'intégralité des éléments extérieurs, comme si nous photographions la scène se présentant sous nos yeux, il devient difficile d'expliquer pourquoi nous percevons certains détails mieux que d'autres. Autrement dit, si les éléments sont encodés de manière équivalente et synchrone dans une représentation interne, comment se fait-il que nous ne détectons pas de manière équivalente des modifications physiquement comparables ? Pourquoi voyons nous plus facilement une rose blanche devenir bleue qu'une rose blanche devenir rouge ? Ou encore pourquoi ne voyons nous plus la même voiture changer de couleur alors que nous ajoutons simplement un nouvel élément dans l'image ? Cette explication de la cécité aux changements est donc problématique lorsque l'on introduit la question de l'influence des facteurs sémantiques. En effet, si pour l'une des quatre raisons décrites ci-dessus, la comparaison entre deux images, une image originale et une image modifiée, est impossible, il devient difficile d'expliquer comment l'identification d'un

changement est alors possible lorsque le changement concerne un élément d'intérêt central ou donne lieu à un élément incohérent. De même, comment expliquer qu'une identification de changements reste possible, même si c'est dans de moindres proportions, concernant des éléments cohérents ou d'intérêt marginal⁽²⁾.

Une représentation interne limitée.

On prend acte du fait que nos représentations internes sont volatiles et beaucoup moins détaillées qu'on peut le penser. Dans cette «théorie de la cohérence» (Rensink, 2000) le cerveau construit en un premier temps un arrangement sommaire et temporaire de la scène visuelle. Le rôle de l'attention sera de guider la construction d'un percept en sélectionnant un certain nombre de caractéristiques pertinentes et de stabiliser cette représentation afin qu'elle puisse former un objet individuel continu à travers l'espace et le temps. L'attention ne peut se focaliser que sur un petit nombre d'items à la fois (Pashler, 1988 ; Pylyshyn et Storm, 1988) et donc, à chaque instant, seuls un petit nombre d'items dans la scène auront une représentation stable. Ainsi nous ne formons pas une représentation complète et détaillée de l'ensemble des objets qui nous environnent mais, à chaque instant, nous ne formons de représentation détaillée seulement de l'objet de nos préoccupations immédiates et par conséquent, ne pouvons détecter de changement que concernant cet objet.

Cette théorie de la cohérence est assez proche de celle proposée par Kahneman, Treisman et Gibbs (1992). Selon ces auteurs, lorsque l'attention se focalise sur un objet, les différentes caractéristiques de cet objet constituent un « fichier objet » (object file), dans

⁽²⁾ En plus de ces quatre hypothèses, il serait possible d'envisager que l'information est préservée de manière détaillée mais que nous manquons d'accès conscient à la représentation ou au changement lui-même. Ceci empêche un report conscient du changement. Des travaux suggèrent ainsi l'existence d'une représentation implicite préservée sans conscience visuelle (Schacter, 1987), voir par exemple les travaux récents sur le « clignement attentionnel » (attentional blink) (Shapiro et al., 1997). Mais nous nous intéressons dans le cadre de cet article uniquement aux informations préservées consciemment et il faudrait, dans le cadre de cette dernière théorie, rendre compte de pourquoi nous avons un accès conscient dans certains cas (éléments d'intérêt central) et non dans d'autres (éléments d'intérêt marginal).

lequel, les différentes propriétés ou structures spatio-temporelles sont liées ensemble pour former, durant un instant, un objet unifié. A la différence de la théorie de la cohérence, ces « fichiers objet » peuvent contenir des informations sur des propriétés non-visuelles des objets et plusieurs « fichiers objet » peuvent se maintenir à la fois.

Elimination de l'idée d'image interne.

On fait passer au tranchant rasoir d'Occam l'idée même d'image interne en posant que la vision peut s'expliquer sans en passer par ces stimuli intermédiaires. Selon cette approche, le monde visible est une «mémoire externe» (O'Regan, 1992), mémoire puisque réserve d'objets subsistants et disponibles, à laquelle il est possible d'accéder instantanément et au moment voulu, via un mouvement des yeux. Il n'est ainsi nul besoin de reconstruire intérieurement le monde extérieur, il est déjà continuellement disponible à l'extérieur.

Dans une telle perspective, il est nécessaire d'expliquer comment nous avons accès à la scène visuelle externe. Si la perception n'est pas la constitution d'une réplique terme à terme de la chose, qu'est-elle alors ? Si nous n'encodons pas l'intégralité des détails d'une scène, qu'encodons-nous alors ?

Nous n'encodons, c'est-à-dire ne voyons, que les traits principaux d'une scène : sa substance sémantique. Par substance sémantique, nous voulons dire que nous ne formons que des représentations digitales ou propositionnelles de la scène visuelle. Et il n'est en aucun cas utile de postuler l'existence de représentation de nature analogique. La substance sémantique de la scène, son sens se délivre, ne serait-ce qu'approximativement, dès la première fixation. Aux fixations suivantes, pour peu que la substance soit semblable, nous ferons comme si les détails étaient les mêmes. Simons et Levin (1997) proposent l'exemple d'une rue très animée. Dans une telle situation, beaucoup de changements peuvent avoir lieu durant une saccade ou entre deux regards successifs. Un traitement visuel trop précis, traquant le moindre détail, se

retrouverait devoir faire face à une grande confusion. Au contraire, le fait que notre système visuel intègre seulement l'essentiel d'une scène d'une vue sur l'autre donne «l'impression d'une stabilité plutôt que d'un chaos». Autrement dit, c'est parce que notre système visuel n'intègre que l'essentiel de ce que nous regardons que le monde visuel nous paraît stable.

Les deux dernières théories présentées permettent à la fois de rendre compte de la cécité aux changements et de rendre compte de la différence d'identification des changements en fonction de facteurs sémantiques. Cette différence peut vouloir dire que toutes les informations présentes dans une scène visuelle ne sont pas encodées de manière équivalente et ne se présentent pas à nous de manière synchrone. Le système visuel ne forme pas de représentation interne persistante et détaillée du monde visuel externe, représentation qui constituerait ce qui est perçu. Nous ne percevons dans la scène que les éléments sur lesquels nous portons notre attention. Par conséquent, dans les conditions expérimentales empêchant une capture automatique de l'attention par l'apparition de signaux transitoires accompagnant les changements, nous ne percevons que les changements concernant les éléments sur lesquels s'était portée notre attention. Différents facteurs : physiques, contextuels ou sémantiques favorisent la perception et la remémoration de certains aspects de l'image.

Une autre question se pose à présent : si nos représentations ne sont pas complètes et détaillées, pourquoi avons-nous néanmoins la forte impression que c'est ce genre de représentation qui sous-tend notre expérience visuelle ? En d'autres termes, si notre vision du monde est une vision parcellaire de seulement un certain nombre de détails des scènes, comment se fait-il que nous ayons l'impression subjective de la richesse, de la précision et de la continuité du monde ?

Nous suggérons que si notre expérience visuelle est celle d'un percept détaillé du monde visuel, c'est à la fois parce que, dans des conditions normales de perception, notre sensibilité aux changements abrupts de luminance nous donne l'impression de pouvoir voir tous les changements qui se produisent dans la scène visuelle et donc d'être visuellement conscient à chaque instant de l'intégralité des informations présentes dans cette scène ; et parce que nous avons continuellement accès à l'intégralité des données qui nous sont offertes à l'extérieur. Ainsi notre impression d'un monde détaillé et cohérent est basée sur les propriétés du monde visuel externe et non sur les propriétés de représentations supposées qui sous-tendent notre expérience visuelle.

Bibliographie

- ANTES J.R., PELAND J.G. - (1981) Picture context effects on eye movement patterns, in D.F. FISHER, R.A. MONTY et J.W. SENDERS (Edit.) Eye movements : Cognition and visual perception, Hillsdale (NJ), Lawrence Erlbaum Associate, 157-170.
- BIEDERMAN I., MEZZANOTTE R.J., RABINOWITZ J.C., FRANCOLINI C.M., PLUDE D. - (1981) Detecting the unexpected in photointerpretation, Human Factor, 23, 153-154.
- CHABRIER R. - (1999) La cécité aux changements progressifs dans les scènes complexes (rapport polycopié), Diplôme d'études approfondies de sciences cognitives, Ecole des Hautes Etudes en Sciences Sociales, Paris.
- CHRISTIE J., KLEIN R. - (1995) Familiarity and attention : Does what we know affect what we notice ? Memory and Cognition, 23, 547-550.
- COLTHEART M. - (1980) Iconic memory and visible persistence, Perception and Psychophysics, 27, 183-228.
- COLTHEART M. - (1983) Iconic memory, Philosophical Transactions of the Royal Society, London, B 302, 283-294.
- FOLK C.L., REMINGTON R.W., JOHNSTON C.J. - (1992) Involuntary covert orienting is contingent on attentional control setting, Journal of Experimental Psychology : Human Perception and Performance, 18 (4), 1030-1040.
- FRIEDMAN A. - (1979) Framing pictures : The role of knowledge in automatized encoding and memory for gist, Journal of Experimental Psychology : General, 108, 316-355.
- HABER R.N. - (1983) The impending demise of the icon: A critique of the concept of iconic storage in visual information processing, Behavioural and Brain Sciences, 6, 1-54.
- HENDERSON J.M. - (1992) Object identification in context : the visual processing of natural scenes, Canadian Journal of Psychology, 4 (special issue), 319-341.
- HENDERSON J.M. - (1997) Transsaccadic memory and integration during real-world object perception, Psychological Science, 8 (1), 51-55.
- HENDERSON J.M., HOLLINGWORTH A. - (1998) Eye movements during scene viewing: An overview, in G. UNDERWOOD (Edit.), Eye guidance in reading and scene perception, Oxford, UK, Elsevier, 269-283.
- HOCHBERG J. - (1986) Representation of motion and space in video and cinematic displays, in K. R. BOFF, L. KAUFMAN et J. P. THOMAS (Edit.) Handbook of perception and human performances : Vol.1 : Sensory processes and perception, London, John Wiley & Sons, 22.1-22.64.
- HOLLINGWORTH A., HENDERSON J.M. - (2000) Semantic informativeness mediates the detection of changes in natural scenes, Visual Cognition, 7 (1/2/3/), 213-235.
- IRWIN D.E. - (1991) Information integration across saccadic eye movements, Cognitive Psychology, 23, 420-456.

- IRWIN D.E., ANDREWS R.V. - (1996) Integration and accumulation of information across saccadic eye movements, in T. INUI et J.L. MCCLELLAND (Edit.), Attention and performance XVI : Information integration in perception and communication, Cambridge, MA, USA, MIT Press, 125-155.
- KAHNEMAN D., TREISMAN A., GIBBS B. - (1992) The reviewing of object files : Object-specific integration of information, Cognitive Psychology, 24, 175-219.
- LEVIN D.T., SIMONS D.J. - (1997) Failure to detect changes to attended objects in motion pictures, Psychonomic Bulletin and Review, 4 (4), 501-506.
- MARR D. - (1982) Vision : A computational investigation into the human representation and processing of visual information, San Francisco, W.H. Freeman.
- O'REGAN J.K. - (1992) Solving the « real » mysteries of visual perception : The world as an outside memory, Canadian Journal of Psychology, 46 : 3, 461-488.
- O'REGAN J.K., DEUBEL H., CLARK J.J., RENSINK R.A. - (2000) Picture changes during blinks : Looking without seeing and seeing without looking, Visual Cognition, 7 (1/2/3/), 191-211.
- O'REGAN J.K., NOË A. - (2001) A sensorimotor account of vision and visual consciousness, Behavioral and Brain Sciences, 24 (5).
- O'REGAN J.K., RENSINK R.A., CLARK J.J. - (1999) Change blindness as a result of « mudsplashes », Nature, 398, p. 34.
- PASHLER H. - (1995) Attention and visual perception : Analysing divided attention, in S.M. KOSSLYN et D.N. OSHERSON (edit.), Visual cognition, an invitation to cognitive science, Vol. 2, Cambridge, Mass, MIT press, 71-100.
- PASHLER H. - (1988) Familiarity and visual change detection, Perception and Psychophysics, 44 (4), 369-378.
- PHILLIPS W.A. - (1974) On the distinction between sensory storage and short-term visual memory, Perception and Psychophysics, 16, 283-290.
- PYLYSHYN Z.W., STORM R.W. - (1988) tracking multiple independent targets : Evidence for a parallel tracking mechanism, Spatial Vision, 3, 179-197.
- RAYNER K., POLLATSEK A. - (1983) Is visual information integrated across saccades ? Perception and Psychophysics, 34 (1), 39-48.
- RENSINK R.A. - (2000) The dynamic representation of scenes, Visual Cognition, 7 (1/2/3/), 191-211.
- RENSINK R.A., O'REGAN J.K., CLARK J.J. - (1997) To see or not to see : The need for attention to perceive changes in scenes, Psychological Science, 8, 368-373.
- SCHACTER D.L. - (1987) Implicit memory : History and current status, Journal of Experimental Psychology : Learning, Memory, and Cognition, 118, 501-518.
- SHAPIRO K.L., ARNELL K.A., RAYMOND J.E. - (1997) The attentional blink : A view on attention and a glimpse on consciousness, Trends in Cognitive Sciences, 1, 291-296.
- SHEPARD R.N. - (1967) Recognition memory for words, sentences and pictures, Journal of Verbal Learning and Verbal Behaviour, 6, 156-163.
- SIMONS D.J. - (1996) In sight, out of mind : When object representation fail, Psychological Science, 7 (5), 301-305.
- SIMONS D.J. - (2000) Current approaches to change blindness, Visual Cognition, 7 (1/2/3/), 1-15.
- SIMONS D.J., FRANCONERI S.L., REIMER R.L. - (2000) Change blindness in the absence of a visual disruption, Perception, 29 (10), 1143-1154.
- SIMONS D.J., LEVIN D.T. - (1997) Change Blindness, Trends in Cognitive Sciences, 1 (7), 261-267.
- SIMONS D.J., LEVIN D.T. - (1998) Failure to detect changes to people in a real-world interaction, Psychonomic Bulletin and Review, 5 (4), 644-649.
- SPERLING G. - (1960) The information available in brief visual presentations, Psychological Monographs: General and Applied, 74 (11), 1-29.
- STANDING L. - (1973) Learning 10,000 pictures, Quarterly Journal of Experimental Psychology, 25, 207-222.
- STANDING L., CONEZIO J., HABER R.N. - (1970) Perception and memory for pictures : Single-trial learning of 2500 visual stimuli, Psychonomic Science, 19 (2), 73-74.
- YANTIS S. - (1993) Stimulus-driven attentional capture and attentional control setting, Journal of Experimental Psychology : Human perception and Performance, 19 (3), 676-681.
- YANTIS S., EGETH H.E. - (1999) On the distinction between visual salience and stimulus-driven attentional capture, Journal of Experimental Psychology : Human perception and Performance, 25 (3), 661-676.
- YANTIS S., JONIDES J. - (1990) Abrupt visual onsets and selective attention : Voluntary versus automatic allocation, Journal of Experimental Psychology: Human Perception and Performance, 16 (1), 121-134.

Annexe 1. Expérience 1 : détail des images.

		Scène	Changement
Image 01	IC	Une banque, surmontée d'un panneau « change », en lettres bleues.	Les lettres deviennent rouges.
	IM	Même chose, avec beaucoup de monde et un vendeur de marrons devant la banque.	
Image 02	IC	Une pharmacie.	Les lettres et la croix de la pharmacie deviennent rouges
	IM	Même chose avec une jongleuse et des amoureux se chamaillant en premier plan.	
Image 03	IC	Une voiture bleue	Devient rouge
	IM	Une femme nous fait un signe de la main à côté de la voiture, et un bus passe derrière la voiture.	
Image 04	IC	Une boîte à lettre de la poste	Prend une couleur rouge.
	IM	Deux amoureux marchent en se tenant la main en premier plan.	
Image 05	IC	Dans un appartement, deux pommes vertes sont posées sur une table.	Les pommes deviennent rouges.
	IM	Une femme est assise sur le canapé derrière la table et téléphone.	
Image 06	IC	Deux bananes sont posées sur la même table.	Les bananes deviennent rouges.
	IM	Sur le canapé, un homme est en conversation téléphonique.	

(IC : intérêt central, IM : Intérêt marginal)

Annexe 2. Expérience 2 : détail des images.

		Scène		Changement
Image 01		Un grand arbre devant la Seine, le ciel est chargé de grands nuages blancs.	C	Les nuages deviennent bleus.
			NC	Les nuages deviennent rouges
Image 02		Un garçon aux yeux noirs en gros plan.	C	Ses yeux deviennent bleus.
			NC	Ses yeux deviennent rouges.
Image 03		Sous une arcade, un guitariste joue devant une plante dans un pot marron.	C	Le pot de la plante devient gris foncé.
			NC	Il devient violet.
Image 04		Dans un appartement, un garçon lit un journal, sur la table est posée une rose blanche dans un vase.	C	La rose devient rouge
			NC	Elle devient bleue.
Image 05		Vue de haut, une route séparée par une ligne blanche.	C	La ligne devient jaune.
			NC	Elle devient bleue.
Image 06		Image de bande dessinée, un lapin traverse un ciel bien bleu.	C	Le ciel devient jaune.
			NC	Le ciel devient jaune.

(C : Cohérent, NC : non cohérent)

RESUME

Les expériences proposées étudient la capacité à identifier un changement survenant progressivement dans un élément d'une scène visuelle. Nous montrons que la capacité d'identification des changements s'accroît en fonction de l'« intérêt » de l'élément et en fonction de son degré d'« incohérence » au sein de la scène. Ces travaux suggèrent qu'une partie seulement des informations présentes dans une scène est préservée en mémoire.

La discussion porte sur les explications possibles de ces résultats et leurs implications quant à l'étude de la nature de nos représentations.

Mots clés : vision, cécité aux changements, facteurs sémantiques.