



HAL
open science

La perception multisource : une nouvelle approche de la perception de scènes

Emmanuelle Ménétrier, André Didierjean

► **To cite this version:**

Emmanuelle Ménétrier, André Didierjean. La perception multisource : une nouvelle approche de la perception de scènes. *Psychologie Française*, A paraître, 10.1016/j.psfr.2021.10.003 . hal-03683658

HAL Id: hal-03683658

<https://hal.science/hal-03683658v1>

Submitted on 31 May 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

La perception multisource :
Une nouvelle approche de la perception de scènes

Multisource perception: a new way to conceptualize scene perception

Emmanuelle Ménétrier *
Maître de conférences, Université d'Angers

André Didierjean
Professeur des universités, Université de Bourgogne-Franche-Comté
Membre honoraire de l'Institut Universitaire de France

Emmanuelle Ménétrier
Université d'Angers - Maison de la Recherche Germaine Tillion
5 bis Boulevard Lavoisier
49045 Angers Cedex 1, France
Tel. 02.41.22.63.70.
Mail : emmanuelle.menetrier@univ-angers.fr

La perception multisource : Une nouvelle approche de la perception de scènes

Résumé

Les recherches menées ces dernières années dans le domaine de la cognition spatiale font état d'observations empiriques originales dont l'approche "traditionnelle" parvient parfois difficilement à rendre compte. L'une des contributions majeures de ces observations a été d'inviter à repenser la perception de scènes, considérant que celle-ci bénéficierait à ne plus être appréhendée comme étant centrée principalement sur la modalité sensorielle considérée. A ce titre, le *Modèle multisource de la perception de scènes* (Intraub et coll.) offre un cadre théorique alternatif, selon lequel la perception de scènes serait structurée autour d'un cadre spatial égocentrique, complété par différentes sources d'information (externe, interne). Toutes travailleraient de concert de façon à créer une simulation des régions environnantes probables, intégrant la vue perçue dans un contexte spatial plus large. Cette revue de question se propose notamment de présenter comment l'une des avancées du modèle d'Intraub est d'unifier différents champs de la cognition jusqu'alors appréhendés isolément.

Mots-clés : perception de scènes ; mémoire visuelle ; *boundary extension* ; modèle multisource.

Multisource perception: a new way to conceptualize scene perception

Abstract

Research conducted these last years in the field of spatial cognition report empirical findings that are difficult to account for with the traditional visual cognitive model of scene perception. One of the major contributions of these findings has been to invite rethinking scene perception, which would benefit from not being apprehended as centered mainly on the sensory modality considered. On the contrary, the *Multisource model of scene perception* developed by Intraub and coll. offers an alternative theoretical framework considering visual perception as an act of spatial cognition, with spatial information at its core. According to this model, during the initial understanding of a view, the cognitive system would be elaborating a multisource representation, with spatial information constituting an egocentric framework that conveys to the observer a sense of the environment in which he/she is embedded. Scene representation would be organized around an amodal spatial structure combining different sources of information: a bottom-up and external source of information derived from different modalities (e.g., visual, haptic), as well as internal sources of high-level information (i.e., amodal, conceptual and contextual information). These different sources of information would work together to create a simulation of the likely environment, integrating the perceived view into a broader spatial context. Beyond rethinking scene perception, one of the advances of the model is to unify different fields of cognition apprehended until then in isolation. The current paper aims to present this model and some of the results it allows to account for.

Keywords: scene perception; visual memory; boundary extension; multisource model

La perception multisource :

Une nouvelle approche de la perception de scènes

Principale source de connaissance de notre environnement, la perception visuelle est essentiellement appréhendée sur la base de l'information sensorielle. Alors que son caractère automatique et permanent lors des périodes d'éveil pourrait laisser penser à tort à un mécanisme très élémentaire sur le plan des processus mobilisés, se cache une réalité beaucoup plus complexe, liée notamment à des contraintes d'ordre physiologique. Si le champ visuel binoculaire couvre un angle visuel global de 220 par 110 degrés (e.g., Streri, 1993), l'acuité visuelle optimale se limite toutefois à la région de la fovéa, petite dépression se trouvant au centre de la rétine et couvrant seulement deux degrés du champ visuel humain. L'acuité visuelle se dégrade à mesure que l'on s'éloigne de la région fovéale pour se diriger vers les régions parafovéale (2-5 degrés de l'angle visuel), puis périphérique (au-delà de 5 degrés). Les raisons d'une telle dégradation sont liées à la structure rétinienne, dont la densité en cônes décroît soudainement au-delà de la région fovéale pour ne plus être qu'éparse dans les régions périphériques. Cette acuité limitée contraint la prise d'information, conduisant l'œil à changer régulièrement de point de fixation (3 à 4 fois par seconde) par le biais de saccades, mouvements balistiques d'une durée moyenne de 42 ms au cours desquels la vision est interrompue¹ (e.g., Kreckelberg, 2010). Le processus perceptif se caractérise ainsi par sa discontinuité, reposant sur une alternance permanente entre prise d'information et maintien de l'information extraite en mémoire afin de combler les lacunes informationnelles liées aux interruptions régulières des entrées sensorielles (e.g., Intraub, 2014).

Ce constat de décalage entre réalité physiologique du processus perceptif et expérience subjective d'un environnement continu, stable et cohérent, est à l'origine de nombreuses interrogations, liées notamment à la question de l'intégration visuelle (i.e., processus de combinaison de l'information extraite d'une fixation oculaire à l'information extraite de la vue suivante). Si l'approche « traditionnelle » de la perception de scènes, désignée ultérieurement par les termes « cognition visuelle », permet aisément de rendre compte de certains phénomènes mettant l'accent sur le caractère épars et volatile des représentations maintenues entre deux fixations (cf., phénomène de cécité au changement, ou *change blindness*, Grimes, 1996), celle-ci présente davantage de difficultés à apporter une explication satisfaisante à un phénomène d'enrichissement de la trace mnésique survenant suite à des intervalles de rétention aussi brefs que de la durée d'une saccade. Ce phénomène, connu sous le nom d'extension des limites (ou *boundary extension*, BE dans la suite du texte ; pour des revues, e.g., Intraub, 2010, 2012 ; Ménétrier, Didierjean, & Marmèche, 2011), fait référence à une distorsion de la mémoire visuo-

¹ Les données indiquées sont celles traditionnellement relevées dans le champ de la perception de scènes visuelles. Celles-ci ne sont données qu'à titre indicatif, et peuvent varier en fonction du matériel inspecté ou de l'activité cognitive en cours (e.g., Rayner, 2009).

spatiale caractérisée par une tendance à surestimer l'étendue des scènes préalablement perçues. Depuis sa découverte en 1989 par Intraub et Richardson, le phénomène fait l'objet d'un intérêt grandissant, comme en témoigne l'augmentation régulière du nombre de publications sur le sujet (i.e., une douzaine dans les années 1990 contre près d'une cinquantaine au cours de la dernière décennie). Outre le fait que celui-ci se situe à la frontière entre perception et mémoire (Roediger, 1996), l'une des raisons de cet intérêt repose probablement sur le fait que le BE constitue un moyen original d'étudier le processus perceptif et la représentation de scènes : les observateurs se remémorent en effet avoir perçu des régions de l'environnement se trouvant au-delà des limites d'une vue physique en l'absence de corrélat sensoriel, témoignant du caractère constructif de la perception de scènes (e.g., Intraub, 2014).

L'objectif poursuivi dans cet article est avant tout de montrer comment des données récentes issues de la cognition spatiale ont permis de repenser le processus perceptif de manière originale, à travers la présentation d'un modèle développé par Intraub et coll., le *Modèle multisource de la perception de scènes* (e.g., Intraub & Dickinson, 2008 ; Intraub, 2010, 2012, 2014). Si la perspective qui prédomine en psychologie postule que le percept serait principalement structuré à partir de l'information en provenance de la modalité sensorielle considérée², l'originalité du *Modèle multisource* (MMS) est de centrer le processus perceptif sur l'information de nature spatiale, issue de la position de l'observateur dans l'environnement. A l'information spatiale, constituant l'ossature de la perception de scène, sont intégrées d'autres sources d'information : de l'information en provenance des différentes modalités sensorielles, ainsi que de l'information de haut niveau, issue des connaissances.

Avant de présenter le contexte ayant conduit à l'émergence de ce modèle et les avancées théoriques dont il est à l'origine, certains des aspects fondamentaux de la perception de scènes visuelles et des explications apportées jusque-là en vue d'en rendre compte seront exposés dans une première partie. Le phénomène d'extension des limites, sur lequel s'appuie principalement Intraub, ne sera ensuite présenté que succinctement, l'objectif n'étant pas ici d'en proposer un exposé exhaustif (voir pour une revue en ce sens, Ménétrier et al., 2011), mais de montrer en quoi les recherches réalisées dans ce domaine ont permis de mettre en exergue le caractère fondamentalement spatial de la perception de scènes. La dernière partie de cette note théorique sera quant à elle consacrée à une présentation détaillée du *Modèle multisource*, auquel les travaux sur le BE ont donné naissance.

² Cette idée n'exclut en rien l'importance accordée par cette approche à d'autres sources informationnelles dans la structuration du percept, telles que les connaissances, émotions ou buts de l'observateur (e.g., Bruner & Goodman, 1947 ; Gasper & Clore, 2002 ; Henderson, 2003 ; Ménétrier, Didierjean & Vieillard, 2013).

1. La perception de scènes visuelles selon l'approche de la cognition visuelle

Bien que classiquement appréhendée isolément des autres champs de la cognition (e.g., Intraub, 2012), la perception visuelle constitue une activité cognitive à part entière, comme en témoignent les nombreux traitements qu'elle met en jeu afin d'assurer l'expérience d'un percept fluide et continu, malgré une interruption régulière de la prise d'information (e.g., Krekelberg, 2010). Si le processus perceptif soulève de nombreuses questions, cette première partie sera abordée dans le contexte de la perception de scènes visuelles (e.g., Intraub, 1997, 2007). Les mécanismes élémentaires impliqués dans les processus perceptifs seront présentés brièvement, en vue notamment de rendre compte du processus d'intégration visuelle, impliqué dans la perception d'un environnement stable et cohérent.

1.1. *La perception d'un environnement stable et cohérent : l'intégration visuelle*

Historiquement, l'une des premières réponses apportées à la question du processus d'intégration reposait sur l'existence hypothétique d'un tampon visuel intégratif dont la fonction aurait été de combiner l'information visuelle extraite au cours des fixations successives. L'idée sous-jacente était celle d'une structure mnésique de capacité importante qui permettrait de maintenir, entre deux fixations, des représentations précises et détaillées de notre environnement physique. Les différentes informations provenant des saccades visuelles y seraient fusionnées de sorte à former une représentation composite de notre environnement visuel (e.g., Jonides, Irwin, & Yantis, 1982). L'hypothèse du maintien de représentations de nature photographique a toutefois été remise en question dès la fin des années 70, à travers des travaux réalisés dans le domaine de la lecture notamment. Faisant l'hypothèse du maintien de représentations visuelles hautement fidèles et détaillées, McConkie et Zola (1979) s'attendaient en toute logique à ce que le processus d'intégration soit perturbé par des changements des caractéristiques physiques des stimuli intervenant au cours des saccades. En présentant des fragments de texte dont La CaSsE éTaIt iNvErSeE d'une fixation à la suivante (i.e., ce fragment devenait « LA cAsSe EtAiT iNvErSeE »), les auteurs pensaient non seulement observer une perturbation du processus de lecture, mais s'attendaient également à ce que les participants remarquent les changements opérés. Les auteurs ont en réalité échoué à observer les effets escomptés : non seulement le processus d'intégration semblait ne pas avoir été perturbé, mais aucun des participants n'avait par ailleurs perçu les changements opérés. Ces travaux ont constitué une avancée majeure dans la conceptualisation du processus d'intégration visuelle : en invalidant l'hypothèse du maintien de représentations hautement fidèles et détaillées d'une fixation à la suivante, c'est l'existence même du tampon visuel intégratif qui était remise en question (Voir cependant, Paeye, Collins, & Cavanagh, 2017).

Différentes recherches menées par la suite dans le domaine de la représentation de scènes visuelles sont venues corroborer l'idée d'un processus d'intégration reposant sur des représentations éparses et faiblement détaillées, comme en témoignent les travaux réalisés dans le domaine de la détection du changement (e.g., Currie, McConkie, Carlson-Radvansky, & Irwin, 2000 ; Henderson & Hollingworth, 1999a ; Irwin, Brown, & Sun, 1988 ; McConkie & Currie, 1996 ; Rayner, McConkie, & Ehrlich, 1978). L'un des phénomènes ayant majoritairement contribué à en illustrer le caractère schématique est sans conteste le phénomène de cécité au changement (e.g., O'Regan, Rensink, & Clark, 1999 ; Rensink, 2002 ; Simons & Rensink, 2004). Celui-ci se caractérise par une difficulté à détecter un changement, même important, survenant au cours d'une interruption du flux optique, que celle-ci soit d'origine naturelle (e.g., saccade oculaire, clignement de paupière), ou artificielle (e.g., présentation d'un écran vierge dans le cadre de dispositifs expérimentaux). Le phénomène s'observe également lorsque le changement n'est pas masqué, mais généré progressivement (e.g., Auvray & O'Regan, 2003). De même, le phénomène se manifeste sans masque, par le biais de dispositifs simulant des éclaboussures (*mudsplashes*) qui apparaissent de façon arbitraire lors du changement d'image, capturant l'attention en différents régions de la scène (O'Regan, Rensink, & Clark, 1999). Décrit pour la première fois par Grimes en 1996, ce sont néanmoins les travaux de Rensink, O'Regan et Clark (1997) qui ont marqué une avancée considérable dans sa compréhension. Outre le fait que ce phénomène soit d'origine attentionnelle³ et non de nature perceptive, les observations réalisées par ces auteurs sont venues corroborer l'idée du caractère épars et faiblement détaillé des représentations maintenues d'une fixation à la suivante : les représentations visuelles se caractériseraient avant tout par leur nature schématique. Celles-ci renverraient à des versions appauvries des scènes réelles, ne maintenant que de l'information d'intérêt central par rapport à la catégorie sémantique de la vue perçue (e.g., O'Regan et al., 1999 ; voir également, Boloix, 2007a, pour une revue en français). Le bien-fondé du maintien de représentations éparses et peu détaillées est cependant loin de faire l'unanimité, certains auteurs, tels que O'Regan et Noë (1992 ; O'Regan et al., 1999 ; Noë et O'Regan, 2000), considérant que de telles représentations sont inutiles, notre environnement physique constituant une mémoire externe continuellement disponible que l'observateur serait libre d'interroger à loisir. Ceci permettrait notamment d'expliquer l'impression de richesse perceptive dont nous faisons en permanence l'expérience.

La question de la précision des détails remémorés reste sujette à débat. S'il apparaît que les représentations élaborées ne sont jamais d'une précision photographique, ce notamment dans des situations permettant un encodage approfondi de l'information sensorielle (i.e., fixations multiples), différents auteurs soutiennent que les individus seraient capables de former

³ Plus globalement, les travaux réalisés dans le domaine de la détection du changement ont étayé l'idée selon laquelle l'attention est nécessaire à la prise de conscience des éléments de l'environnement, comme en témoignent également d'autres phénomènes d'aveuglement attentionnel, tels que le phénomène de cécité inattentionnelle (e.g., Simons, 2000, 2010 ; Simons & Chabris, 1999).

Ces recherches s'inscrivent par ailleurs dans la lignée de travaux ayant contribué à montrer que le déploiement attentionnel au sein des scènes visuelles n'est pas guidé uniquement par les caractéristiques de bas niveau (e.g., couleur, contraste, texture, orientation ; cf., travaux d'Itti & Koch, 2000, sur les cartes de saillance), mais également par les connaissances dont dispose l'observateur vis-à-vis de la scène perçue (Pour un modèle intégrant cartes de saillance, connaissances et contraintes liées à la tâche en cours, voir par exemple, le *Modèle du guidage contextuel* développé par Torralba Oliva, Castelhana, & Henderson, 2006).

des représentations visuelles détaillées de l'ensemble d'une scène (e.g., Brady, Konkle, Alvarez, & Oliva, 2008 ; Henderson & Hollingworth, 2003 ; Standing, 1973). Si ces observations semblent difficilement conciliables avec l'idée de représentations visuelles éparpillées et peu détaillées, ce paradoxe peut toutefois être levé si l'on considère que les auteurs soutenant ce point de vue appréhendent la question différemment (i.e., mesures de nature différente : détection explicite vs. implicite du changement ; stades différents du traitement informationnel : précoce vs. tardif, opérationnalisés via des durées de présentation allant de 240 ms à 20 s, e.g., Hollingworth, 2003 ; Rensink et al., 1997), rendant les comparaisons difficiles à établir (cf., Boloix, 2007a, pour une discussion détaillée de ces différences). Cette remarque rejoint l'une des questions centrales dans le domaine de la perception de scènes visuelles, celle de la question de la nature des représentations élaborées en fonction des opportunités d'approfondissement du traitement de l'information sensorielle offertes à l'observateur (e.g., Intraub, 2012).

1.2. *Caractérisation des scènes visuelles*

Définir ce que l'on entend par scène visuelle est moins aisé qu'il n'y paraît, en raison de nombreuses questions et problèmes soulevés par leur caractérisation (e.g., problèmes liés à l'absence de métrique permettant d'en décrire le niveau de détails, questions liées à leur cadrage, e.g., Intraub, 2010). Un consensus semble toutefois émerger autour du fait que les scènes doivent représenter le monde « réel » (e.g., Intraub, 2010 ; Henderson & Hollingworth, 1999b), en opposition notamment au caractère très artificiel de certains stimuli parfois employés en laboratoire. Par ailleurs, celles-ci doivent être interprétées par le système cognitif comme des vues partielles d'un monde continu, c'est-à-dire comme des localisations ou surfaces du monde réel se prolongeant au-delà des limites du champ visuel ou de l'image perçue (e.g., Gottesman & Intraub, 2002 ; Intraub, Gottesman, & Bills, 1998).

Parmi les problèmes posés par la caractérisation des scènes visuelles, se pose notamment la question de leur cohérence structurelle. A ce titre, Biederman, Mezzanotte et Rabinowitz (1982) se sont attachés à définir les relations structurales objet/contexte permettant de considérer qu'une scène est correctement formée. Les auteurs ont ainsi identifié cinq règles relationnelles, distinguant règles physiques et règles sémantiques. Les deux premières décrivent des relations physiques entre l'objet et le contexte : (1) selon la règle d'interposition, l'opacité de l'objet interrompt le fond de la scène, et (2) selon la règle de support, les objets reposent sur des surfaces, à l'exception d'une minorité. Les trois autres règles décrivent des relations sémantiques, dont la particularité est de requérir l'accès à la signification de l'objet. Si la règle de probabilité (3) met l'accent sur la vraisemblance de l'objet au sein du contexte présenté, la règle de position (4) précise les emplacements prototypiques d'un objet au sein de ce contexte. La règle de taille (5) met quant à elle l'accent sur la cohérence de la taille de l'objet par rapport aux autres objets habituellement rencontrés dans ce contexte.

1.3. *Vitesse d'identification des scènes visuelles*

Lorsque l'on s'intéresse à la perception de scènes visuelles, les questions posées vont toutefois au-delà de questions relatives aux processus élémentaires impliqués dans le fonctionnement du système perceptif, et soulèvent des questions aussi variées que celles de leur vitesse d'identification, de la priorité de traitement accordée à leurs éléments constitutifs ou, comme évoqué précédemment, de la quantité et de la précision des détails remémorés à différents stades du traitement de l'information.

Le fonctionnement du système visuel se caractérise par la rapidité et l'efficacité des traitements opérés, dont on sait qu'elles revêtent une fonction adaptative importante. Si l'on considère la question d'un point de vue évolutionniste, cette capacité constituerait une composante essentielle des systèmes vivants biologiques, les chances de survie d'un animal dépendant directement de sa capacité à identifier rapidement un prédateur ou une proie potentielle (e.g., VanRullen & Thorpe, 2001). Dès les années 70, les recherches sur la vitesse d'identification de scènes ont montré que le système visuel est capable d'en extraire la catégorie sémantique dans les 100 ms suivant leur apparition, les travaux les plus récents suggérant que la reconnaissance serait possible dans les 13 premières ms (Potter & Haggmann, 2015 ; Potter, Wyble, Haggmann, & McCourt, 2014. Voir également, Greene & Oliva, 2009). En présentant des photographies de scènes de la vie quotidienne en rafale, paradigme connu sous le nom de séquences RSVP (*Rapid Serial Visual Presentation*), Potter (1976) a montré que l'identification de scènes visuelles est possible sous des conditions extrêmes (Intraub, 2012). L'une des particularités de ce paradigme repose en effet sur la mise en œuvre de situations très éloignées de conditions naturelles de perception, à travers la présentation séquentielle de photographies n'entretenant aucun lien sémantique et ce, à des fréquences de présentation excédant le rythme auquel est habituellement réalisée la prise d'information (e.g., une image toutes les 113 ms, voire toutes les 13 ms en l'absence d'intervalle interstimulus ; voir, respectivement, Potter, 1976, et Potter et al., 2014). De par ces caractéristiques, le paradigme élimine la continuité inhérente à la perception de l'environnement et le contrôle que peut exercer l'observateur sous des conditions normales de balayage. Les résultats observés par l'auteure indiquent que la classification opère extrêmement rapidement car ne requérant pour cela qu'une seule fixation et ce, malgré le coût cognitif élevé induit par le paradigme (Intraub, 2012). Du reste, l'identification scénique serait facilitée par le fait qu'une analyse grossière de la structure spatiale globale suffirait à la catégorisation (e.g., Oliva & Torralba, 2006), le système visuel traitant en priorité la structure globale d'une scène avant de procéder à une analyse plus détaillée des objets qui la constituent (e.g., Navon, 1977 ; Oliva & Torralba, 2001 ; Palmer, 1975)⁴.

L'un des intérêts de ces études est que Potter (1976 ; voir également Potter et al., 2014 ; Potter & Haggmann, 2015) s'est par la même occasion attachée à étudier le souvenir relatif aux scènes perçues sous ces conditions de présentation extrêmement brèves et saccadées. Si les résultats indiquent qu'une fixation oculaire suffit à la catégorisation scénique, il en va

⁴ Ceci ne signifie en rien que l'identification d'objet ne joue aucun rôle dans le processus de catégorisation scénique, comme en témoigne l'importance des associations objet/contexte (e.g., Davenport & Potter, 2004).

différemment du souvenir relatif au contenu des scènes présentées. La reconnaissance des scènes présentées initialement au sein des séquences RSVP s'est en effet avérée inférieure au seuil du hasard. Ces faibles taux de performance indiquent que l'on peut identifier momentanément davantage de scènes qu'il n'est possible de s'en remémorer par la suite, suggérant qu'identification et mémoire des scènes reposent sur des mécanismes distincts (Intraub, 2012 ; Potter, 1976). Il apparaît toutefois que ces observations seraient spécifiques aux caractéristiques temporelles mises en jeu dans les séquences présentées. En effet, pour des durées de présentation similaires, le niveau de performance dépasse le seuil du hasard lorsque les scènes sont séparées par un intervalle interstimulus de 1.5 seconde. Le rythme de présentation des scènes constituerait ainsi un facteur critique du faible niveau de performance mnésique relevé, ne permettant pas la consolidation de l'information conceptuelle dans un registre de MLT. Selon Potter, la dissociation observée entre capacités de catégorisation et performances mnésiques résulterait d'un effet de masquage conceptuel (e.g., Intraub, 1981 ; Potter, 1993) : au cours de séquences RSVP, la catégorie sémantique de la vue perçue serait maintenue sous forme verbale dans un registre conceptuel à court terme, l'apparition soudaine de la scène subséquente écrasant le label verbal de la vue précédente. Les performances différentielles observées (Potter, 1976 ; Potter et al., 2014) posent ainsi la question du format des représentations maintenues. Sur ce point, les modèles classiques de la perception de scènes envisagent le stockage de l'information visuelle et de la catégorie sémantique associées à la scène perçue sous forme de représentations distinctes : de nature sensorielle pour la première, et verbale/abstraite pour la seconde (e.g., Hollingworth, 2003, 2004 ; Hollingworth & Henderson, 2002). Cette relation est toutefois envisagée différemment dans le cadre du *Modèle multisource*, point qui sera abordé ultérieurement.

Le tampon conceptuel, activé dans les 100 ms suivant l'apparition d'une scène visuelle (e.g., Potter, 1976), ne renvoie toutefois qu'à l'un des tampons à court terme par lesquels transite une information visuelle extrêmement volatile avant sa consolidation potentielle dans un registre mnésique à long terme. L'approche dominante de la cognition visuelle a en effet coutume de mettre l'accent sur le devenir de l'information lors des différentes étapes du processus perceptif (e.g., Intraub, 2010). Une fois parvenue aux registres d'information sensorielle (e.g., Atkinson & Shiffrin, 1968), l'information visuelle extraite d'une fixation oculaire serait maintenue en mémoire iconique, registre mnésique assurant très brièvement (300-400 ms) la prolongation de l'information visuelle au cours des fixations sous une forme brute et hautement détaillée (Sperling 1960 ; voir, Loftus, Johnson, & Shimamura, 1985, pour une généralisation à des scènes visuelles). Au cours des saccades en revanche, ou en cas de masquage visuel⁵, l'information serait maintenue sous une forme tout autre (i.e., schématique) en mémoire transsaccadique (MT), décrite comme un sous-système de la mémoire visuelle à court terme (ou MVCT, Hollingworth, Richard, & Luck, 2008 ; Phillips, 1974). Toutes deux partagent en effet un certain nombre de caractéristiques, telles qu'une capacité de stockage limitée à quatre objets tout au plus (e.g., Irwin, 1992 ; Luck & Vogel, 1997 ; Pashler, 1988 ; voir cependant Irwin & Zelinsky, 2002), ou le maintien de représentations dont la précision est

⁵ Le phénomène de masquage visuel se révèle différent des effets de masquage conceptuel décrits précédemment, car ne relevant pas de l'étiquette verbale associée à la catégorie sémantique de la scène perçue, mais de ses caractéristiques physiques.

inférieure à celle de la mémoire iconique (Irwin, 1991 ; Phillips, 1974). Si la MVCT assure le maintien de l'information visuelle sur des fenêtres temporelles plus longues que la MT (i.e., quelques secondes), toutes deux, contrairement à la mémoire iconique, résistent aux effets de masquage, les objets étant encodés durablement (e.g., Irwin, 1991, 1992). Le dernier tampon visuel à court terme par lequel l'information est susceptible de transiter n'est autre que la mémoire visuelle de travail (e.g., Hollingworth & Luck, 2009), sollicitée lorsque l'observateur est engagé dans des tâches mobilisant de l'information visuelle, et pour lesquelles il est nécessaire d'en assurer le maintien sur des durées plus longues qu'en MVCT.

Comme le souligne Intraub (2012), ces différents tampons ne renverraient pas à des entités structurelles distinctes mais davantage à des phases différentes d'un processus continu de traitement de l'information sensorielle, assurant son maintien sur des fenêtres temporelles allant de quelques dizaines de millisecondes à quelques secondes après sa disparition. Du fait des capacités limitées de traitement de l'information auxquelles est assujéti le système cognitif, la mémoire visuelle serait sous le joug d'un goulet d'étranglement (Broadbent, 1958 ; Pashler, 1988 ; Phillips, 1974), ne maintenant qu'une quantité limitée d'information au cours des traitements. Si ces capacités limitées de traitement semblent ne pas avoir d'effet sur la compréhension des scènes perçues, c'est leur souvenir qui serait impacté, comme le suggèrent les travaux de Potter (1976 ; 2014) exposés plus haut. Il apparaît ainsi que, lorsqu'il s'agit de former une représentation robuste de scènes visuelles complexes (i.e., qui n'est plus sujette aux effets de masquage), le système visuel serait contraint de s'appuyer sur une structure mnésique de capacité plus importante, la mémoire visuelle à long terme (MVLT, e.g., Hollingworth & Henderson, 2002 ; Melcher, 2006). Celle-ci agirait de concert avec la MVCT⁶ de sorte à maintenir une représentation visuelle précise et détaillée de la scène perçue (e.g. Hollingworth, 2003, 2004).

Si la mémoire semble impactée par les capacités limitées de traitement du système cognitif, la catégorisation des scènes à un stade très précoce du traitement informationnel se révèle essentielle (Intraub, 2012). Outre le déclenchement de stratégies de reconnaissance des objets (cf., effets de contexte, e.g., Palmer, 1975 ; Biederman et al., 1982 ; Davenport & Potter, 2004), celle-ci guiderait le déploiement des fixations oculaires, c'est-à-dire l'exploration visuelle au sein des scènes perçues (e.g., Boloix, 2007b, pour une revue). Différents travaux s'accordent sur le fait que ces effets seraient causés par les connaissances, abstraites et stockées en MLT, dont dispose l'observateur vis-à-vis des scènes perçues. La compréhension initiale d'une scène s'accompagne en effet de l'activation automatique de représentations prototypiques des scènes perçues, les schémas de scènes (e.g., Palmer, 1975), générant des attentes relatives tant aux objets qu'il est plausible de trouver au sein de la scène rencontrée qu'aux relations spatiales que ceux-ci sont susceptibles d'entretenir (e.g., Bar, 2004).

⁶ Les concepts de mémoire visuelle à court terme et mémoire visuelle de travail (MVDT) sont employés de manière interchangeable par certains auteurs (e.g., Hollingworth & Henderson, 2002 ; Luck & Vogel, 1997). Melcher (2006) souligne toutefois la perte informationnelle à confondre ces deux concepts. Alors que la MVCT correspond davantage à un niveau de stockage de l'information, situé entre mémoire sensorielle et MLT, la seconde assurerait notamment la récupération d'informations stockées en MLT.

1.4. Synthèse et problématique

Si l'approche dominante dans le champ de la cognition visuelle a permis de faire d'importantes découvertes et continue de nourrir de nombreuses questions, elle n'en présente pas moins certaines limites car ne permettant pas de rendre compte de certaines observations réalisées au cours de la dernière décennie dans le champ de la cognition spatiale. Alors que cette approche permet aisément d'expliquer la pauvreté des fixations maintenues d'une fixation à la suivante, aucune des théories développées dans ce cadre ne permet véritablement de prédire ni d'expliquer le phénomène d'enrichissement de la trace mnésique reporté dans le BE, dont l'une des particularités est de survenir suite à des intervalles de rétention aussi brefs que de la durée d'une saccade (42 ms ; e.g., Intraub & Dickinson, 2008). Pour rendre compte de ce phénomène selon l'approche dominante de la cognition visuelle, il faudrait, soit intégrer un nouveau tampon visuel à court terme à l'ensemble des tampons décrits précédemment (i.e., un tampon d'extrapolation de scène), soit inclure le processus de prédiction impliqué dans le BE à l'un des tampons existants (e.g., la mémoire trans-saccadique ; Intraub, 2010). Une telle manière de conceptualiser le processus d'extrapolation pose problème d'un point de vue épistémologique, dans la mesure où ces tampons et processus additionnels sont proposés *a posteriori* et n'en présentent de ce fait qu'une puissance explicative très limitée (Intraub, 2010).

Avant de présenter le *Modèle multisource de la perception de scènes*, développé par Intraub et coll. (e.g., Intraub, 2014) en vue de pallier certaines des lacunes inhérentes au modèle de la cognition visuelle, l'objet de la partie suivante sera de présenter succinctement le BE en mettant l'accent sur l'ancrage fondamentalement spatial de ce phénomène. Comme évoqué précédemment, l'objectif ne sera pas d'en faire une présentation détaillée (cf., Ménétrier et al., 2011, pour une revue en français), mais davantage de montrer en quoi les travaux réalisés dans ce domaine ont permis de faire évoluer la conceptualisation de la perception de scènes et des représentations qui en émergent.

2. Le caractère spatial de la représentation de scènes : Apports du phénomène d'extension des limites

L'une des conséquences d'un champ visuel limité spatialement est de ne révéler que des vues partielles d'un monde continu, dont l'individu doit prendre connaissance par le biais de mouvements oculaires ou corporels. Du fait de l'internalisation de la structure physique de l'environnement (e.g., Intraub, 2001), ces vues partielles sont à l'origine d'une attente de continuité spatiale. Le système visuel (SV) interprète en effet les vues perçues comme se prolongeant au-delà de leurs contours, qu'il s'agisse des limites du champ visuel ou, à un niveau opérationnel, des limites de stimuli en 2D comme des photographies (e.g., Intraub, 2010). Lors de la compréhension initiale des vues perçues, le SV est ainsi engagé dans un processus d'extrapolation de la structure spatiale probable des régions de l'environnement échappant à la perception de l'observateur, l'extension des limites⁷ (*Boundary Extension*, BE). Observé pour la première fois par Intraub et Richardson en 1989, le BE est décrit comme une distorsion de la mémoire visuo-spatiale caractérisée par une tendance à surestimer l'étendue des vues préalablement perçues, le souvenir de l'observateur incluant de l'information que ne comportait pas la scène originale mais que le contexte rend plausible d'inférer.

Bien que présenté comme une erreur mnésique relative au contenu des vues perçues, le phénomène se distingue en différents points des autres formes de faux souvenirs classiquement rapportées à travers la littérature, car se produisant notamment dans des conditions sous lesquelles il semble raisonnable d'attendre de bonnes performances mnésiques. Alors qu'erreurs mnésiques et faux souvenirs (e.g., Corson & Verrier, 2013 ; Loftus, 1997) sont traditionnellement décrits comme se produisant dans des situations impliquant par exemple une forte charge mnésique, le BE a été rapporté dans des conditions de faible charge (i.e., après présentation très brève d'une unique photographie testée immédiatement après, e.g., Intraub & Dickinson, 2008 ; Munger, Owens, Conway, 2005), ou des durées d'exposition des stimuli suffisamment longues pour permettre un traitement approfondi de l'information sensorielle (e.g., Gagnier, Intraub, Oliva, & Wolfe, 2011). Ceci peut paraître d'autant plus surprenant que nombre de recherches font état d'une excellente mémoire visuelle chez les individus, s'apparentant à des capacités de reconnaissance exceptionnelles (e.g., Standing, 1973 ; Brady et al., 2008). Ce dernier point met l'accent sur l'une des principales caractéristiques du BE, le phénomène mettant en jeu de l'information de nature sensorielle, plus concrète que l'information verbale classiquement impliquée dans les autres formes de faux souvenirs (cf., paradigmes DRM ou de désinformation, Deese, 1959 ; Roediger & McDermott, 1995 ; Loftus, Miller, & Burns, 1978 ; Ménétrier, Iralde, & Le Bohec, 2019). De même, alors qu'erreurs mnésiques et faux souvenirs sont traditionnellement observés suite à des délais suffisamment longs pour causer un affaiblissement de la trace, le BE a été rapporté par le biais de dispositifs

⁷ Si nous avons opté ici pour une traduction littérale du nom du phénomène, il est également courant de désigner celui-ci par les termes « extension du champ », en référence au champ visuel limité de l'observateur.

simulant et enregistrant l'activité oculomotrice en temps réel. Ce résultat indique que le phénomène se produit tant avec des durées d'exposition coïncidant avec la durée d'une fixation oculaire que suite à des intervalles de rétention de la durée d'une saccade (Intraub & Dickinson, 2008). Outre le fait de suggérer que le phénomène se situerait à la frontière entre perception et mémoire (Roediger, 1996), ces observations sont relativement importantes sur le plan fonctionnel, car laissant supposer que le BE participerait à des processus perceptifs aussi élémentaires que le processus d'intégration visuelle. Celui-ci contribuerait ainsi à l'impression de continuité et de stabilité dont nous faisons en permanence l'expérience (Dickinson & Intraub, 2008 ; Intraub & Dickinson, 2008). L'implication du phénomène à un niveau élémentaire du processus perceptif a été corroborée par des recherches ayant montré que celui-ci faciliterait le traitement des scènes visuelles en amorçant le contenu de la vue suivante (Gottesman, 2011). Loin de renvoyer à un processus mnésique reconstitutif *ad hoc* (i.e., se produisant une fois que la scène a disparu) comme l'ont initialement supposé Intraub et coll., ces observations suggèrent que le BE serait activé à un stade très précoce du traitement de l'information visuelle, l'identification d'une configuration spatiale activant automatiquement un processus d'extrapolation de l'environnement alentour probable (Gottesman & Intraub, 2002). De ce fait, le phénomène trouverait son origine dans le mode de traitement perceptif des scènes (global vs. analytique), le souvenir de l'observateur constituant un reliquat de l'activité d'extrapolation de la structure spatiale environnante déployée lors de la compréhension initiale de la vue perçue. Un traitement basé sur la structure globale de la scène serait en effet associé à un phénomène d'extension des limites, alors qu'un traitement basé sur les détails (i.e., objets) serait davantage associé à une perte informationnelle résultant en un effet de restriction des limites (Ménétrier et al., 2019). Ainsi, si le BE est avant tout décrit comme une erreur mnésique relative au contenu des vues perçues, il n'en présente pas moins une valeur adaptative importante : en permettant à l'observateur de réaliser des prédictions remarquablement justes sur la structure spatiale des régions de l'environnement échappant à sa perception, le BE faciliterait les interactions avec celui-ci à travers des tâches telles que la navigation (e.g., Chadwick, Mullally, & Maguire, 2013).

Nature spatiale des représentations scéniques

Si le processus d'extrapolation mis en jeu dans le BE semble jouer un rôle fondamental tant dans des processus perceptifs élémentaires que dans l'adaptation de l'individu à son environnement, son importance fonctionnelle semble corroborée par le fait que celui-ci s'observe tout au long de l'ontogenèse - aussi bien chez des personnes âgées que chez des nourrissons de trois mois (Ménétrier et al., 2019 ; Multhaup, Munger, & Smith, 2016 ; Quinn & Intraub, 2007 ; Seamon, Schlegel, Hiester, Landau, & Blumenthal, 2002) ; dans des troubles impliquant un traitement de l'information privilégiant le détail au détriment de la structure globale des stimuli, telles que le syndrome d'Asperger (Chapman, Ropar, Mitchell, & Ackroyd, 2005) ; ou chez des individus présentant une altération du développement de l'hippocampe, comme dans le syndrome de Down (Spanò, Intraub, & Edgin, 2017). De même, la compensation des dégradations de l'acuité visuelle par le système cognitif constitue l'une des

hypothèses avancées pour rendre compte de la prise de conscience tardive des symptômes de pathologies ophtalmiques affectant la vision périphérique, telles que les rétinites pigmentaires (Intraub, 2010).

La généralisation du phénomène à d'autres modalités sensorielles semble par ailleurs témoigner de l'importance du processus d'extension sur le plan fonctionnel. Si les tentatives de généralisation à la modalité auditive se sont à ce jour révélées peu fructueuses (Hutchison et al., 2012), le BE a toutefois été mis en exergue dans le domaine du toucher (Intraub, 2004), représentations visuelle et haptique reposant sur une représentation spatiale commune (Intraub, Morelli, & Gagnier, 2015). Cette dernière observation se révèle fondamentale sur le plan théorique, car suggérant que la perception de scènes serait avant tout un acte de cognition spatiale. Si le processus d'extrapolation mis en jeu dans le BE se produit également en l'absence de perception, le simple fait de penser à une scène suffisant à activer des attentes relatives au contexte spatial environnant (Intraub et al., 1998), celui-ci ne serait toutefois pas nécessairement lié à des processus d'imagerie visuelle, comme en témoignent les recherches menées auprès d'une jeune femme sourde et aveugle depuis son plus jeune âge en raison du syndrome de Leber, et dont le toucher constitue le principal mode d'exploration de son environnement (Intraub, 2004, 2014).

Le BE apparaît ainsi comme un processus hautement adaptatif, se produisant face à des configurations interprétées comme des scènes par le système cognitif, et dans des conditions sous lesquelles le besoin de prédiction est particulièrement important. Il a en effet été montré que l'ampleur de l'extrapolation dépend notamment des caractéristiques physiques des scènes perçues et, plus particulièrement, de la quantité d'espace représentée. Qu'il s'agisse du plan de la scène ou de son angle de prise de vue, plus la quantité d'espace visible est importante, moins l'extrapolation sera forte. Les scènes en gros plan sont en effet à l'origine d'un effet BE plus important que les mêmes scènes présentées en plan large (e.g., Dickinson & Intraub, 2008), de même qu'une scène présentée de face entraînera davantage de BE que la même scène photographiée à un angle de 45° (Gagnier et al., 2011). Ces observations sont en accord avec des travaux récents montrant que, lorsque l'observateur dispose de connaissances spécifiques sur le contexte spatial plus large dans lequel la scène perçue prend place, l'extrapolation est moins importante que face à une scène inconnue (Ménétrier, Didierjean, & Barbe, 2018). L'ensemble de ces observations semble cohérent d'un point de vue fonctionnel, le besoin de prédictions se révélant plus important dans les situations où l'observateur n'a pas connaissance de l'environnement dans lequel il évolue. Cette idée est d'ailleurs renforcée par le fait que le BE se produit lors de la compréhension initiale des vues perçues (i.e., lors de la première fixation réalisée sur une scène), c'est-à-dire, lorsque l'observateur est confronté à une scène pour la première fois. Comme le souligne Bar (2007), il semble en effet que le système cognitif minimise le traitement de l'information lorsque celle-ci est prédictible, le but étant de diriger les ressources attentionnelles en direction d'objets ou régions de l'environnement pertinents face à la situation rencontrée.

Parallèlement aux études comportementales exposées précédemment, différentes recherches réalisées dans le domaine de la neuroimagerie et de la pathologie ont contribué à

étayer l'hypothèse du caractère fondamentalement spatial de la représentation de scènes visuelles. Ainsi, Hassabis, Kumaran, Vann, et Maguire (2007) ont rapporté un déficit dans le processus de construction de scènes chez des patients amnésiques présentant une atteinte bilatérale de l'hippocampe. Outre l'amnésie antérograde, trait caractéristique majeur de ce type de syndrome, les auteurs ont montré que ces patients ont des difficultés à imaginer ce à quoi pourrait ressembler l'environnement au-delà des limites de photographies qui leur étaient présentées. Si leurs descriptions se sont révélées relativement correctes concernant les objets susceptibles d'être contenus par la scène, l'analyse de leurs verbalisations a en revanche révélé un manque de cohérence spatiale, les environnements imaginés s'avérant fragmentés et manquant d'indications relatives au positionnement des objets. S'interrogeant sur l'effet d'un déficit dans le processus de construction de scène lors de la compréhension initiale des scènes perçues, Mullally, Intraub, et Maguire (2012) ont montré que ces patients se révèlent peu enclins au BE, leur souvenir relatif à l'étendue des scènes perçues s'avérant plus précis que celui de participants contrôles. Ces observations suggèrent ainsi que, outre le rappel d'expériences passées ou l'imagination de scènes ou d'événements futurs, l'hippocampe serait impliqué dans la construction de scènes et les représentations spatiales et, par extension, des tâches cognitives telles que la navigation (e.g., Chadwick et al., 2013).

Le rôle central de l'hippocampe dans le processus d'extrapolation de la structure spatiale environnante a par ailleurs été corroboré par les recherches menées dans le domaine de l'imagerie cérébrale. Chadwick et al. (2013) montrent ainsi qu'au cours de la compréhension initiale des vues perçues, le BE serait associé à l'activation de l'hippocampe et du cortex parahippocampique. Plus précisément, le processus de construction de scènes, initié lors de la compréhension initiale de la vue perçue, serait amorcé par le cortex préfrontal ventromédian (CPFvm). Cette structure agirait de manière complémentaire à l'hippocampe (DeLucia et al., 2018), le CPFvm étant à l'origine de l'activation de schémas de scènes permettant à l'hippocampe d'élaborer une première représentation spatiale scénique. Cette représentation, très élémentaire, serait par la suite enrichie de détails appropriés au contexte par le CPFvm qui, engagé dans un processus itératif avec l'hippocampe et le néocortex, permettrait la formation d'une représentation scénique riche et complexe. Au cours de la remémoration en revanche, ce seraient l'aire parahippocampique des lieux et le cortex rétrosplénial qui seraient sollicités (Park, Intraub, Yi, Widders, & Chun, 2007). Ces structures sous-tendant toutes deux l'exploration visuelle et haptique seraient respectivement impliquées dans la perception de la structure spatiale et dans l'intégration des vues perçues dans un contexte spatial plus large (Epstein, 2008). Ces observations étayaient l'hypothèse d'un phénomène en deux temps, exposée plus amplement dans la partie suivante.

3. Le modèle multisource de la perception de scènes : une alternative aux approches « classiques » de la cognition visuelle

3.1. *Idée centrale du Modèle multisource*

Alors que la plupart des modèles de la cognition visuelle considèrent que l'origine des représentations scéniques n'est autre que la/les modalités sensorielles considérées, le MMS défend le point de vue selon lequel la perception de scènes renverrait avant tout à un acte de cognition spatiale. L'information spatiale constituerait en effet un cadre de référence égocentrique donnant à l'observateur une idée de l'environnement dans lequel il/elle se trouve (e.g., ce qu'il est plausible de trouver sur sa gauche et sa droite, mais aussi derrière, au-dessus ou au-dessous de lui). La structure spatiale sous-tendant la représentation de scènes se révélerait ainsi très proche de la notion d'image mentale proposée par Loomis, Klatzky et Giudice (2013), toutes deux renvoyant à une représentation spatiale environnante (i.e., non limitée au plan frontal) de nature amodale. A la différence de la notion d'image mentale toutefois, la structure spatiale mobilisée dans le MMS n'est pas générée en réponse à une stimulation, mais sert de cadre organisateur permanent aux représentations (Intraub et al., 2015).

Si les recherches réalisées dans le domaine de la perception de scènes reposent majoritairement sur des simulations du champ visuel à travers la présentation de stimuli en 2D, il convient de garder à l'esprit que, sous des conditions naturelles de perception, l'individu prend place au sein des scènes perçues, c'est-à-dire au sein d'un environnement se prolongeant au-delà des limites de son champ visuel (e.g., Intraub, 2014). Bien que la prise d'information soit réduite, lors de la perception initiale d'une scène, aux régions de l'environnement se trouvant face à lui, les connaissances dont dispose l'observateur lui permettent d'extrapoler la structure probable des régions échappant à sa perception. L'une des idées centrales du modèle repose ainsi sur le fait qu'au cours de la première fixation réalisée sur une scène, le SV serait engagé dans un processus de simulation⁸ de la structure probable des régions environnantes (cf., modèles de la cognition située, e.g., Barsalou, 2008), dont le contenu témoignerait de l'internalisation de la structure physique de l'environnement (e.g., Intraub, 2001 ; Shepard, 1984). Si l'information visuelle transmet de l'information relative à la localisation de l'individu au sein du cadre spatial égocentrique, les lacunes informationnelles liées au fait que la prise d'information ne révèle que des vues partielles de l'environnement seraient comblées par d'autres sources d'information de haut niveau. Activées en vue de placer mentalement la scène perçue dans un contexte spatial plus large, ces différentes sources d'information sont de nature amodale, conceptuelle et contextuelle (e.g., Intraub, 2012).

⁸ Le terme « simulation » reprend le terme employé par Intraub dans ses écrits (Intraub, 2014).

Comme son nom l'indique, l'information amodale résulte d'une forme abstraite de perception, les processus de continuation et de complétion amodale permettant respectivement de prolonger mentalement les surfaces et textures et de "compléter" les objets tronqués au-delà des limites de la vue perçue (Gottesman & Intraub, 2002 ; Kanizsa, 1979 ; Kellman, Yin, & Shipley, 1998 ; Michotte, Thinès, & Crabbe, 1964). Bien qu'aucune tentative de quantification de l'étendue couverte par ces processus n'ait été entreprise à ce jour dans le cadre du BE, il semble que ces mécanismes ne concernent qu'une étendue très restreinte de la simulation opérée, à savoir les régions se trouvant juste au-delà des limites de la scène (Intraub, 2014). D'un point de vue fonctionnel, les processus amodaux sont loin de jouer un rôle négligeable dans le processus perceptif. Au-delà de prolonger la scène juste au-delà de ses limites, ceux-ci sont en effet impliqués dans la compréhension de l'environnement, permettant à l'observateur d'interpréter correctement les scènes perçues. Ce sont ces processus qui permettent par exemple de comprendre que les poubelles représentées sur la Figure 1 ne sont pas cassées, ou qu'une personne sur un portrait n'a pas été décapitée (!).

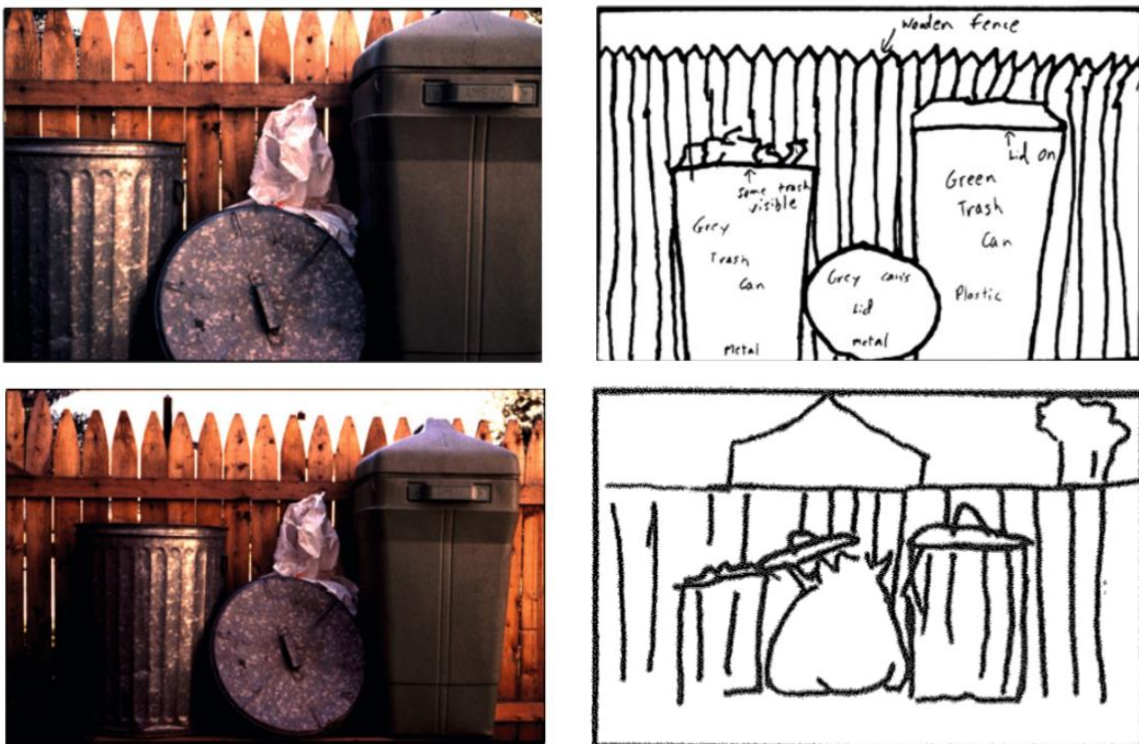


Figure 1. Exemples de scènes en gros plan (haut) et plan large (bas) et dessins présentant un biais d'extension des limites pour chacune de ces scènes. Extrait de Intraub & Richardson (1989). Adapté avec l'aimable autorisation d'Helene Intraub.

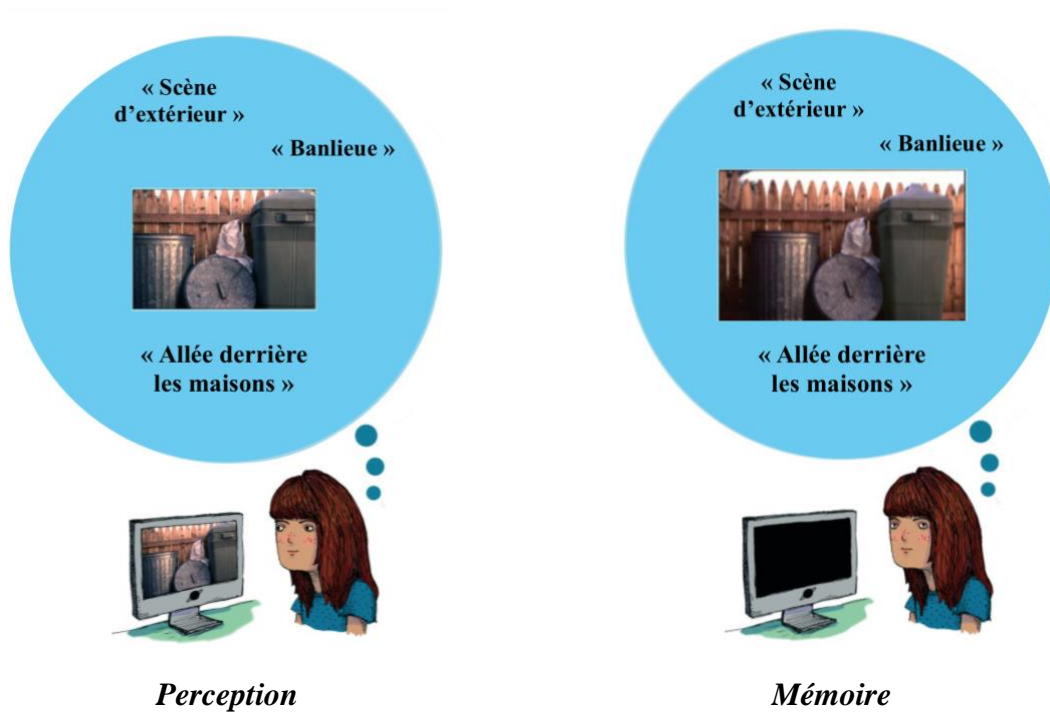
La simulation générée lors de la perception initiale d'une scène résulterait également de l'activation précoce de connaissances conceptuelles (i.e., quelques dizaines de millisecondes suivant l'apparition de la scène, e.g., Potter, 1976, Potter et al., 2014) conduisant à l'identification de la catégorie sémantique de la scène perçue, mais aussi de connaissances contextuelles, permettant d'établir une cartographie mentale de la structure probable des régions environnantes. La notion de connaissances contextuelles est sans doute très proche de celle de schémas de scènes (e.g., Palmer, 1975), car se référant à une représentation schématique des objets qu'il est plausible de trouver au sein du type de scène rencontré et de leur localisation prototypique dans ce contexte. La simulation opère ici à une échelle plus étendue que dans le cadre des processus amodaux décrits précédemment, permettant notamment de rendre compte de l'inclusion, dans le cadre de tâches de dessin par exemple, de détails que ne comportait pas la scène originale (e.g., Figure 1, où l'on observe l'ajout d'une habitation en arrière-plan, voir par exemple, Intraub & Richardson, 1989).

Toutefois, les connaissances de l'observateur ne sauraient à elles seules à rendre compte de la simulation opérée. Celle-ci serait également guidée par l'expérience personnelle (i.e., expérience liée au milieu dans lequel évolue l'individu), une même scène pouvant donner lieu à des interprétations différentes d'un individu à l'autre. Comme le fait remarquer Intraub (2014) à titre illustratif, certains de ses collègues ont proposé des interprétations très différentes de la scène représentée Figure 1 selon la région des États-Unis dans laquelle ils vivaient, décrivant tour à tour les poubelles adossées à la clôture comme se trouvant dans un jardin, dans la rue, ou dans l'allée menant d'une maison à l'autre.

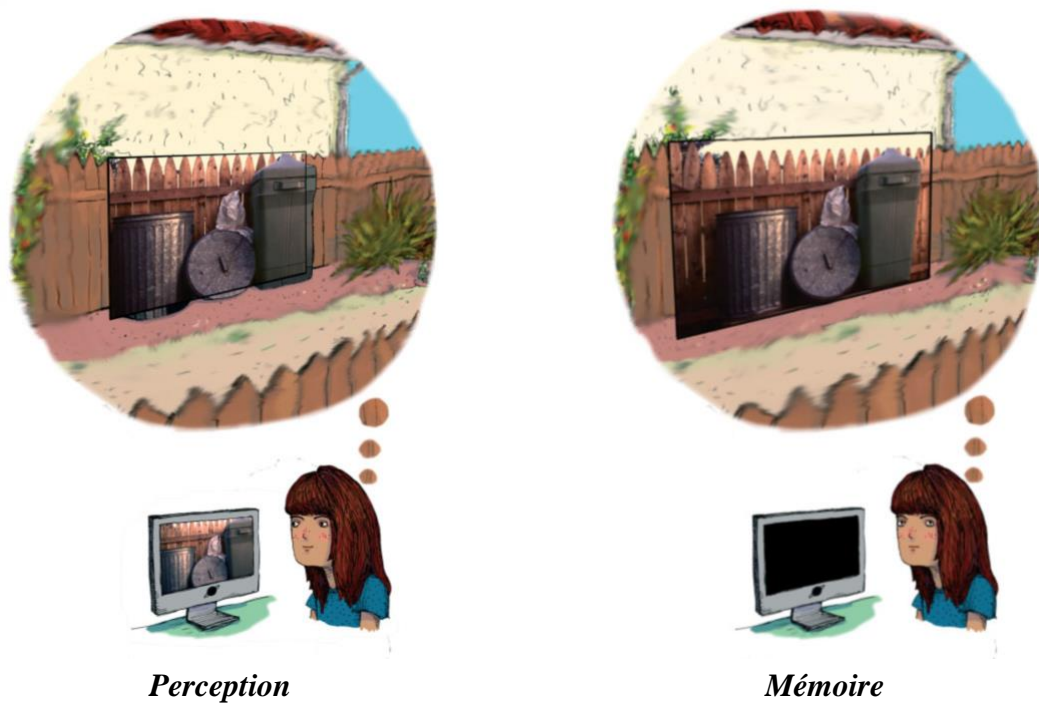
Ainsi, outre le fait de reconsidérer la structure organisatrice des représentations scéniques, l'une des idées centrales du MMS est que la représentation de scènes ne doit pas être appréhendée comme principalement structurée à partir de l'information sensorielle mais à partir d'une représentation multisource, c'est-à-dire d'une représentation organisée autour d'une structure spatiale amodale intégrant différentes sources d'information. Ces sources informationnelles se révèlent être à la fois de nature externe (i.e., relative à la/aux modalités sensorielles considérées) et interne (i.e., information amodale, conceptuelle et contextuelle), et sont activées très précocement, c'est-à-dire lors de la compréhension initiale de la vue perçue (e.g., Intraub, 2010).

A ce titre, l'une des questions centrales posées par le modèle est celle du format des représentations engagées dans le processus de simulation et, plus spécifiquement, de la relation entre information visuelle et concepts relatifs à la scène perçue. De nombreux modèles considèrent qu'information visuelle et connaissances conceptuelles associées à la scène perçue font référence à deux types de représentations distincts, la perception de scènes reflétant l'information visuelle et son association à un concept, représenté par exemple sous forme d'un label verbal (Intraub, 2012). Le MMS, en revanche, envisage différemment la relation entre ces deux types de représentations. La simulation opérée serait en effet associée à une représentation mentale imagée du contexte spatial plus large, intégrant la scène perçue au sein d'une vue prototypique. L'information visuelle serait ainsi "complétée" par une représentation concrète résultant de l'activation des différentes sources d'information de haut niveau évoquées précédemment, à l'origine d'une représentation multisource capturant la continuité de l'espace

environnant (Intraub, 2010, 2012, 2014 ; cf., Figure 2 pour une représentation schématique de ces deux types de relations).



Approche de la cognition visuelle



Modèle multisource

Figure 2. Alors que l'approche de la cognition visuelle envisage la représentation de scènes visuelles à partir de la coexistence de deux types de représentations distincts, visuelle et conceptuelle (e.g., verbale), le Modèle multisource considère qu'au cours de la compréhension initiale de la scène perçue, le système visuel serait engagé dans une activité de simulation de la structure spatiale environnante sous une forme imagée, complétant le cadre spatial égocentrique et l'information visuelle disponible. La figure ci-dessus offre une comparaison du phénomène BE selon ces deux modèles. *Adapté de Intraub (2012) avec l'aimable autorisation d'Helene Intraub. Dessins par Stevie French.*

3.2. *Explication d'erreurs mnésiques se produisant après une saccade*

Si le MMS propose une approche alternative et originale de la perception de scènes visuelles, il n'en rend pas moins compte d'observations que ne permettent pas d'expliquer la plupart des autres approches de la cognition visuelle : l'enrichissement des scènes perçues suite à des intervalles de rétention de la durée d'une saccade. Selon ce modèle, le BE se déroulerait en deux phases successives. La première se produirait au cours de la compréhension initiale des scènes perçues, lors du processus de simulation. Cette première phase, décrite comme le processus d'extension à proprement parler (e.g., Chadwick et al., 2013), est à distinguer de l'erreur d'extension qui, elle, s'observerait lors de la remémoration. Contrairement à la phase d'extrapolation, pendant laquelle interviennent des processus constructifs (i.e., de nature inférentielle), la seconde phase mobiliserait des processus décisionnels puisqu'il serait question pour l'observateur de discriminer les régions extrapolées mentalement des régions réellement perçues (e.g., Intraub, 2012).

Pour illustrer comment les erreurs de jugement sont possibles à ce stade, il convient de se référer à l'analogie proposée par Intraub et coll. (e.g., Intraub & Dickinson, 2008) entre structure rétinienne et représentation de la simulation en mémoire. De manière analogue au percept, dont l'acuité est maximale en vision centrale et faible en périphérie, les représentations scéniques seraient elles aussi graduelles : relativement précises dans les régions de la scène associées à une acuité visuelle optimale et plus faiblement détaillées dans les régions de basse acuité. De ce fait, le souvenir serait relativement flou au niveau des limites de la scène, l'information réellement perçue se confondant avec les régions prolongées par les mécanismes de continuation amodale. Au-delà, la simulation ne renverrait plus qu'à une représentation d'autant plus générale et schématique que l'on s'éloigne des limites de la scène préalablement perçue (Intraub, 2014). En écho au modèle développé par Johnson, Hashtroudi et Lindsay (1993), l'observateur se trouverait confronté, lors de la remémoration, à une tâche de *source monitoring*. Selon ce modèle, les individus seraient face à une tâche de ce type à chaque fois qu'ils doivent juger de l'origine d'un souvenir, notamment lorsqu'il est question de déterminer si celui-ci fait référence à un événement réellement advenu ou s'il s'agit au contraire

d'informations inférées. Le *source monitoring* renvoie ainsi à un processus décisionnel basé sur l'évaluation d'indices associés au souvenir, les expériences vécues présentant des caractéristiques plus rarement retrouvées dans l'information inférée, telles que des informations liées au contexte dans lequel s'est déroulé l'événement, certains détails perceptifs, ou les émotions associées à l'événement. Comme il s'agit d'un processus décisionnel, les individus sont susceptibles de commettre des erreurs d'attribution de la source, à l'origine d'erreurs mnésiques. Dans le cadre du BE, l'observateur se doit ainsi de discriminer les régions de l'environnement réellement perçues de celles qui ont été extrapolées mentalement, la discrimination portant sur les sources d'information suivantes : externe vs. interne ; visuelle vs. amodale. Comme le postule le modèle du *source monitoring*, plus la similarité entre deux sources d'information est importante, plus le risque d'erreurs d'attribution est accru. Selon cette logique, les erreurs de *source monitoring* associées au BE résulteraient de la très forte ressemblance entre information réellement perçue et régions prolongées par les mécanismes de perception amodale. Ainsi, ce qui est attribué à tort à la vision n'inclut pas la simulation dans sa globalité mais, pour des raisons adaptatives (Intraub & Gagnier, 2018), la mince portion de l'espace générée de façon amodale. Ainsi, ce qui serait inclus en mémoire renverrait à une partie seulement des régions de la simulation actives au cours de la première fixation, renforçant l'idée que le BE ne renvoie pas à un processus reconstructif *ad hoc* mais à un processus se produisant au cours de la compréhension initiale de la vue perçue. Ce dernier point revêt une importance cruciale sur le plan théorique car il souligne que le BE ne serait autre qu'un processus perceptif en temps réel, les régions extrapolées étant incluses à la représentation de la scène sous une forme amodale alors que celle-ci est toujours visible (Intraub et al., 2015 ; Mullaly et al., 2012). Une telle conceptualisation a des répercussions sur la manière d'appréhender certains processus élémentaires à l'œuvre dans la perception visuelle, en particulier le processus d'intégration. Si la question de l'intégration visuelle est au cœur des interrogations soulevées par les approches dominantes dans le champ de la cognition visuelle, l'une des particularités du MMS est d'en remettre en cause l'utilité fonctionnelle. Ces approches avancent en effet que les représentations de l'environnement seraient construites progressivement, par accumulation de l'information glanée au cours des fixations successives (e.g., Melcher, 2006). Le MMS considère en revanche que le SV placerait la scène perçue dans un contexte spatial plus large au cours de la première fixation réalisée sur une scène, la simulation opérée reposant sur une représentation multisource pouvant être affinée et corrigée lors des fixations suivantes. De cette manière, les représentations de l'environnement seraient continuellement mises à jour, la perception de scènes renvoyant avant tout à une activité constructive en temps réel. Le MMS propose ainsi une conception alternative expliquant mieux la cohérence de la perception de scènes (Intraub, 2012), tout en constituant un modèle plus parcimonieux que la plupart des autres approches développées dans le champ de la cognition visuelle, qui ont recours à des opérations cognitives supplémentaires pour rendre compte de notre sentiment d'un percept fluide et continu.

Conclusion

Développé afin de rendre compte d'observations que la majorité des autres approches dans le champ de la cognition visuelle parviennent difficilement à expliquer, le modèle d'Intraub et coll. propose une alternative originale, différant en bien des points de l'approche dominante. Alors que selon la plupart des modèles la perception de scènes aurait l'information sensorielle pour principale source d'information, le *Modèle multisource* permettrait de mieux capturer la complexité des scènes et des processus cognitifs engagés dans leur traitement (Intraub, 2010). Celui-ci appréhende en effet la perception de scènes visuelles comme un acte de cognition spatiale, considérant que le cœur de la représentation n'est pas l'information sensorielle, mais le cadre spatial égocentrique issu du point de vue de l'observateur. Le cadre spatial établi constituerait en effet une structure amodale servant de cadre organisateur permanent aux représentations. Lors de la compréhension initiale d'une scène, ce cadre serait complété par différentes sources d'information engageant le système cognitif dans une simulation de l'espace environnant. Ainsi, la représentation mentale d'une scène refléterait tant sa réalité physique probable que l'interprétation qu'en fait l'observateur (Intraub, 2012, 2014). Au-delà de proposer une approche différente de la relation entre information sensorielle et représentation des concepts associés à la scène perçue, le modèle se révèle proche des modèles de la cognition située (e.g., Barsalou, 2008), établissant un lien entre ces derniers et ceux relevant d'une approche computo-symbolique⁹. Le MMS propose par ailleurs de reconsidérer le processus perceptif, estimant que le processus d'intégration renverrait à un *construct* théorique obsolète, dans la mesure où les fixations réalisées successivement sur une scène opéreraient de façon à permettre une mise à jour en temps réel de l'information disponible, et affiner la représentation initiale de la scène perçue (e.g., Chadwick et al., 2013).

Bien que présenté dans le cadre de la perception de scènes visuelles, l'une des contributions majeures du modèle multisource est d'offrir une approche généralisable à l'ensemble des modalités sensorielles, celui-ci n'étant pas centré sur une modalité spécifique. Le MMS présente de la sorte un cadre conceptuel motivant des questions plus générales sur la représentation de scènes, telles que la perception intermodale ou la relation entre perception de scènes et navigation (Intraub, 2014). De même, l'une des contributions majeures du modèle est d'intégrer des champs de la cognition jusqu'alors appréhendés isolément : la cognition spatiale, la perception, et la mémoire, faisant le lien pour cette dernière entre certains tampons mnésiques à très court terme (i.e., mémoire trans-saccadique) et la MLT à travers le modèle du *source monitoring* (Johnson et al., 1993). Les travaux les plus récents suggèrent par ailleurs que le processus de construction de scène opéré lors de la compréhension initiale de la vue perçue serait similaire au processus de simulation impliqué dans la planification d'événements futurs (e.g., Intraub, 2014), le *Modèle multisource* constituant potentiellement une piste prometteuse de théorie unificatrice de la cognition.

⁹ Dans la mesure où l'approche incarnée et située de la cognition ne sera pas développée davantage, car ne relevant pas de l'objet de cet article, le lecteur intéressé pourra consulter, entre autres, la revue de question proposée par Dutriaux et Gyselinck (2016), ainsi que le récent ouvrage de Versace, Brouillet, et Vallet (2018). Voir Goldinger, Papesh, Barnhart, Hansen, & Hout (2016), pour une critique de cette approche.

Références

- Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K.W. Spence, & J.T. Spence (Ed.), *The psychology of learning and motivation*, 2 (pp. 89–195). New York: Academic Press.
- Auvray, M., & O'Regan, J. K. (2003). L'influence des facteurs sémantiques sur la cécité aux changements progressifs dans les scènes visuelles. *L'Année Psychologique*, 103, 9-32.
- Bar, M. (2004). Visual objects in context. *Nature Reviews: Neuroscience*, 5, 617-629.
- Bar, M. (2007). The proactive brain: Using analogies and associations to generate predictions. *Trends in Cognitive Sciences*, 11, 280–289.
- Barsalou, L.W. (2008). Grounded cognition. *Annual Review of Psychology*, 59, 617–645.
- Biederman, I, Mezzanotte, R.J., & Rabinowitz, J.C. (1982). Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive Psychology*, 14, 143-177.
- Boloix, E. (2007a). La représentation des scènes visuelles en mémoire : les apports de la cécité au changement. *L'Année Psychologique*, 107, 459-487.
- Boloix, E. (2007b). Attention et perception de scènes visuelles. *L'Année Psychologique*, 107, 113-151.
- Brady, T.F., Konkle, T., Alvarez, G.A., & Oliva A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Science of the United States of America*, 105, 14325–14329.
- Broadbent, D. (1958). *Perception and communication*. Oxford: Pergamon Press.
- Bruner, J. S., & Goodman, C.C. (1947). Value and need as organizing factors in perception. *Journal of Abnormal Social Psychology*, 42, 33-44.
- Chadwick, M.J., Mullally, S.L., & Maguire, E.A. (2013). The hippocampus extrapolates beyond the view in scenes: An fMRI study of boundary extension. *Cortex*, 49, 2067-2079.
- Chapman, P., Ropar, D., Mitchell, P., & Ackroyd, K. (2005). Understanding boundary extension: Normalisation and extension errors in picture memory among adults and boys with and without Asperger's syndrome. *Visual Cognition*, 12, 1265-1290.
- Corson, Y., & Verrier, N., (2013). *Les faux souvenirs*. Bruxelles : De Boeck.
- Currie, C.B., McConkie, G.W., Carlson-Radvansky, L.A., & Irwin, D.E. (2000). The role of the saccade target object in the perception of a visually stable world. *Perception and Psychophysics*, 62, 673-683.
- Davenport, J.L., & Potter, M.C. (2004). Scene consistency in object and background perception. *Psychological Science*, 15, 559-564.

- Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology*, *58*, 17–22.
- De Lucia, F., McCormick, C., Mullally, S.L., Intraub, H., Maguire, E.A., Ciaramelli, E. (2018). Boundary extension is attenuated in patients with ventromedial prefrontal cortex damage. *Cortex*, *108*, 1-12.
- Dickinson, C.A., & Intraub, H. (2008). Transsaccadic representation of layout: What is the time course of boundary extension? *Journal of Experimental Psychology: Human Perception and Performance*, *34*, 543-555.
- Dutriaux, L. & Gyselinck, V. (2016). Cognition incarnée : un point de vue sur les représentations spatiales. *L'Année Psychologique*, *116* (3), 419-465. doi:10.4074/S0003503316000373.
- Epstein, R.A. (2008). Parahippocampal and retrosplenial contributions to human spatial navigation. *Trends in Cognitive Sciences*, *12*, 388-396.
- Gagnier, K.M., Intraub, H., Oliva, A., & Wolfe, J.M. (2011). Why does vantage point affect Boundary Extension? *Visual Cognition*, *19*, 234-257.
- Gaspar, K., & Clore, G.L. (2002). Attending to the big picture: Mood and global versus local processing of visual information. *Psychological Science*, *13*, 34-40.
- Goldinger, S.D., Papesh, M.H., Barnhart, A.S., Hansen, W.A., & Hout, M.C. (2016). The poverty of embodied cognition. *Psychonomic Bulletin & Review*, *23*, 959–978.
- Gottesman, C.V. (2011). Mental layout extrapolations prime spatial processing of scenes. *Journal of Experimental Psychology: Human Perception Performance*, *37*, 382–395.
- Gottesman, C.V., & Intraub, H. (2002). Surface construal and the mental representation of scenes. *Journal of Experimental Psychology: Human Perception and Performance*, *28*, 589-599.
- Greene, M.R., & Oliva, A. (2009). The briefest of glances: the time course of natural scene understanding. *Psychological Science*, *20*, 464–472.
- Grimes, J. (1996). On the failure to detect changes in scenes across saccades. In K. A. Akins (Ed.), *Vancouver studies in cognitive science, Vol. 5. Perception* (pp. 89-110). New York, NY, US: Oxford University Press.
- Hassabis, D., Kumaran, D., Vann, S.D., Maguire, E.A. (2007). Patients with hippocampal amnesia cannot imagine new experiences. *Proceedings of the National Academy of Science of the United States of America*, *104*, 1726-1731.
- Henderson, J.M. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences*, *7*, 498-504.
- Henderson, J.M., & Hollingworth, A. (1999a). The role of fixation position in detecting scene changes across saccades. *Psychological Science*, *10*, 438-443.

- Henderson, J.M., & Hollingworth, A. (1999b). High-level scene perception. *Annual Review of Psychology*, *50*, 243-271.
- Henderson, J.M., & Hollingworth, A. (2003). Eye movements and visual memory: Detecting changes to saccade targets in scenes. *Perception and Psychophysics*, *65*, 58-71.
- Hollingworth, A. (2003). Failures of retrieval and comparison constrain change detection on natural scenes. *Journal of Experimental Psychology: Human Perception and Performance*, *29*, 388-403.
- Hollingworth, A. (2004). Constructing visual representations of natural scenes: The roles of short- and long-term visual memory. *Journal of Experimental Psychology: Human Perception and Performance*, *30*, 519-537.
- Hollingworth, A., & Henderson, J.M. (2002). Accurate visual memory for previously attended objects in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance*, *28*, 113-136.
- Hollingworth, A., & Luck, S.J. (2009). The role of visual working memory (VWM) in the control of gaze during visual search. *Attention, Perception & Psychophysics*, *71*, 936–949.
- Hollingworth, A., Richard, A.M., & Luck, S.J. (2008). Understanding the function of visual short-term memory: Transsaccadic memory, object correspondence, and gaze correction. *Journal of Experimental Psychology: General*, *137*, 163-181.
- Hutchison, J. L., Hubbard, T. L., Farrandino, B., Brigante, R., Wright, J. M., & Rypma, B. (2012). Auditory memory distortion for spoken prose. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 1469-1489.
- Intraub, H. (1981). Rapid conceptual identification of sequentially presented pictures. *Journal of Experimental Psychology: Human Perception and Performance*, *7*, 603-610.
- Intraub, H. (1997). The representation of visual scenes. *Trends in Cognitive Sciences*, *1*, 217-221.
- Intraub, H. (2001). Internalized constraints in the representation of spatial layout. *Behavioral and Brain Sciences*, *24*, 677-678.
- Intraub, H. (2004). Anticipatory spatial representation of 3D regions explored by sighted observers and a deaf-and-blind observer. *Cognition*, *94*, 19-37.
- Intraub, H. (2007). Scene perception: The world through a window. In M.A. Peterson, B. Gillam, & H.A. Sedgwick (Eds.), *In the mind's eye: Julian Hochberg on the perception of pictures, films and the world* (pp.454-466). NY: Oxford University Press.
- Intraub, H. (2010). Rethinking Scene Perception: A Multisource Model. In B. Ross (Ed.), *Psychology of Learning and Motivation*, *52*, (pp.231-264). Burlington: Academic Press.
- Intraub, H. (2012). Rethinking visual scene perception. *Wiley Interdisciplinary Reviews: Cognitive Science*, *3*, 117–127.

- Intraub, H. (2014). Visual scene representation: A spatial-cognitive perspective. In K. Kveraga & M. Bar (Eds.), *Scene vision: Making sense of what we see* (pp. 5-26). Cambridge, MA: The MIT Press.
- Intraub, H., & Dickinson, C.A. (2008). False memory 1/20th of a second later: What the early onset of boundary extension reveals about perception. *Psychological Science, 19*, 1007-1014.
- Intraub, H., & Gagnier, K.M. (2018). Expanding Space: Does Imagination affect Boundary Extension for Visual Scenes? In T.L. Hubbard (Ed.), *Spatial Biases in Perception and Cognition* (193-208). Cambridge, UK: Cambridge University Press.
- Intraub, H., Gottesman, C.V., & Bills, A.J. (1998). Effects of perceiving and imagining scenes on memory for pictures. *Journal of Experimental Psychology: Learning, Memory and Cognition, 24*, 186-201.
- Intraub, H., Morelli, F., & Gagnier, K.M. (2015). Visual, haptic and bimodal scene perception: Evidence for a unitary representation, *Cognition, 138*, 132-147.
- Intraub, H., & Richardson, M. (1989). Wide-angle memories of close-up scenes. *Journal of Experimental Psychology: Learning, Memory and Cognition, 15*, 179-196.
- Irwin, D.E. (1991). Information integration across saccadic eye movements. *Cognitive Psychology, 23*, 420-456.
- Irwin, D.E. (1992). Memory for position and identity across eye movements. *Journal of Experimental Psychology: Learning, Memory and Cognition, 18*, 307-317.
- Irwin, D.E., Brown, J.S., & Sun, J.S. (1988). Visual Masking and Visual Integration Across Saccadic Eye Movements. *Journal of Experimental Psychology: General, 117*, 276-287.
- Irwin, D.E., & Zelinsky, G.J. (2002). Eye movements and scene perception: Memory for things observed. *Perception and Psychophysics, 64*, 882-895.
- Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research, 40*, 1489-1506.
- Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114*, 3-28.
- Jonides, J., Irwin, D. E., & Yantis, S. (1982). Integrating visual information from successive fixations. *Science, 215*, 192-194.
- Kanizsa, G. (1979). *Organization in vision*. New York, NY: Praeger.
- Kellman, P. J., Yin, C., & Shipley, T. F. (1998). A common mechanism for illusory and occluded object completion. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 859-869. doi:10.1037/0096-1523.24.3.859
- Krekelberg, B. (2010). Saccadic suppression. *Current Biology, 20*, R228-R229.

- Loftus, E.F. (1997). Creating false memories. *Scientific American*, 277, 70-75.
- Loftus, G.R., Johnson, C.A., Shimamura, A.P. (1985). How much is an icon worth? *Journal of Experimental Psychology: Human Perception and Performance*, 28, 113–136.
- Loftus, E.F., Miller, D.G., & Burns, H.J. (1978). Semantic integration of verbal information into a visual memory. *Journal of Experimental Psychology: Human Learning and Memory*, 4, 19-31.
- Loomis, J.M., Klatzky, R.L., & Giudice, N.A. (2013). Representing 3D space in working memory: Spatial images from vision, hearing, touch, and language. In S. Lacey, R. Lawson (eds), *Multisensory imagery* (pp. 131–155). New York: Springer.
- Luck, S.J., & Vogel, E.K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, 390, 279-281.
- McConkie, G.W., & Currie, C.B. (1996). Visual stability across saccades while viewing complex pictures. *Journal of Experimental Psychology: Human Perception and Performance*, 22, 563-581.
- McConkie, G.W., & Zola, D. (1979). Is visual information integrated across successive fixations in reading? *Perception and Psychophysics*, 25, 221-224.
- Melcher, D. (2006). Accumulation and persistence of memory for natural scenes. *Journal of Vision*, 6, 8-17.
- Ménétrier, E., Didierjean, A., & Barbe, V. (2018). Effect of contextual knowledge on spatial layout extrapolation. *Attention, Perception, & Psychophysics*, 80(8), 1932–1945. DOI: 10.3758/s13414-018-1569-9
- Ménétrier, E., Didierjean, A., & Marmèche, E. (2011). Le système visuel traite-t-il les photographies comme des fenêtres ouvertes sur le monde ? *L'Année psychologique*, 111, 749-769.
- Ménétrier, E., Didierjean, A., & Vieillard, S. (2013). Is Boundary Extension emotionally selective? *The Quarterly Journal of Experimental Psychology*, 66, 635-641.
- Ménétrier, E., Iralde, L., & Le Bohec[#], L. (2019). Spatial layout extrapolation and aging: Underlying cognitive and executive mechanisms. *Visual Cognition*, 27 (9-10), 668-686. DOI: 10.1080/13506285.2019.1634663
- Michotte, A., Thinès, G., & Crabbe, G. (1964). Les compléments amodaux des structures perceptives. Louvain : Institut de psychologie de l'Université de Louvain.
- Mullally, S.L., Intraub, H., & Maguire, E.A. (2012). Attenuated boundary extension produces a paradoxical memory advantage in amnesic patients. *Current Biology*, 22, 261-268.
- Multhaup, K.S., Munger, M.P., & Smith, K.C. (2016). Boundary Extension Is Sensitive to Hand Position in Young and Older Adults, *The Journals of Gerontology: Series B*, gbw011, <https://doi.org/10.1093/geronb/gbw011>
- Munger, M. P., Owens, R. T., & Conway, J. E. (2005). Are boundary extension and representational momentum related? *Visual Cognition*, 12, 1041–1056.

- Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, 9, 353-383.
- Noë, A., & O'Regan, J.K. (2000). Perception, attention, and the grand illusion. *Psyche: an interdisciplinary journal of research on consciousness*, 6.
- Oliva, A., & Torralba, A. (2001). Modeling the Shape of the scene: a holistic representation of the spatial envelope. *International Journal in Computer Vision*, 42, 145-175.
- Oliva, A., & Torralba, A. (2006). Building the gist of a scene: the role of global image features in recognition. *Progress in Brain Research: Visual Perception*, 155, 23-36.
- O'Regan, J.K. (1992). Solving the “real” mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology*, 46, 461-488.
- O'Regan, J.K., Rensink, R.A., & Clark, J.J. (1999). Change blindness as a result of “mudsplashes”. *Nature*, 398, 34.
- Paeye, C., Collins, T., & Cavanagh, P. (2017). Transsaccadic perceptual fusion. *Journal of Vision*, 17, 1–11.
- Palmer, S.E. (1975). The effects of contextual scenes on the identification of objects. *Memory and Cognition*, 3, 519-526.
- Park, S., Intraub, H., Yi, D.-J., Widders, D., & Chun, M.M. (2007). Beyond the edges of a view: Boundary extension in human scene-selective visual cortex. *Neuron*, 54, 335-342.
- Pashler, H. (1988). Familiarity and visual change detection. *Perception and Psychophysics*, 44, 369-378.
- Phillips, W.A. (1974). On the distinction between sensory storage and short-term visual memory. *Perception and Psychophysics*, 16, 283-290.
- Potter, M.C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and memory*, 2, 509-522.
- Potter, M.C. (1993). Very short-term conceptual memory. *Memory & Cognition*, 21, 156-161.
- Potter, M.C., & Haggmann, C.E. (2015). Banana or fruit? Detection and recognition across categorical levels in RSVP. *Psychonomic Bulletin and Review*, 22, 578–585.
- Potter, M.C., Wyble, B., Haggmann, C.E., & McCourt, E. S. (2014). Detecting meaning in RSVP at 13 ms per picture. *Attention, Perception, & Psychophysics*, 76, 270-279.
- Quinn, P.C., & Intraub, H. (2007). Perceiving “outside the box” occurs early in development: Evidence for boundary extension in 3- to 7-month old infants. *Child Development*, 78, 324-334.
- Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, 62, 1457-1506.

- Rayner, K., McConkie, G.W., & Ehrlich, S. (1978). Eye movements and integrating information across fixations. *Journal of Experimental Psychology: Human Perception and Performance*, 4, 529-544.
- Rensink, R.A. (2002). Change detection. *Annual Review of Psychology*, 53, 245-277.
- 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.
- Roediger, H.L. (1996). Memory illusions. *Journal of Memory and Language*, 35, 76-100.
- Roediger, H.L., & McDermott, K.B. (1995). Creating false memories: Remembering words not presented in list. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 803-814.
- Seamon, J.G., Schleger, S.E., Heister, P.M., Landau, S.M., & Blumenthal, B.F. (2002). Misremembering pictured objects: People of all ages demonstrate the boundary extension illusion. *American journal of psychology*, 115, 161-167.
- Shepard, R.N. (1984). Ecological constraints on internal representation: Resonant kinematics of perceiving, imagining, thinking, and dreaming. *Psychological Review*, 91, 417-447.
- Simons, D.J. (2000). Attentional capture and inattention blindness. *Trends in Cognitive Sciences*, 4, 147-155.
- Simons, D. J. (2010). Monkeying around with the gorillas in our midst: Familiarity with an inattention-blindness task does not improve the detection of unexpected events. *i-Perception*, 1, 3-6.
- Simons, D.J., & Chabris, C.F. (1999). Gorillas in your midst: Sustained inattention blindness for dynamic events. *Perception*, 28, 1059-1074.
- Simons, D.J., & Rensink, R.A. (2004). Change blindness: Past, present and future. *Trends in Cognitive Sciences*, 9, 16-20.
- Spanò, G., Intraub, H., & Edgin, J. O. (2017). Testing the “boundaries” of boundary extension: Anticipatory scene representation across development and disorder. *Hippocampus*, 27, 726-739.
- Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied*, 74, 1-29.
- Standing, L. (1973). Learning 10000 pictures. *Quarterly Journal of Experimental Psychology*, 25, 207-222.
- Streri, A. (1993). Percevoir. In A. Weil-Barras (Eds.), *L'homme cognitif* (pp.99-118). Paris : PUF.
- Torralba, A., Oliva, A., Castelano, M.S., & Henderson, J.M. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychological Review*, 113, 766-786.

- Van Rullen, R., & Thorpe, S.J. (2001). The time course of visual processing: From early perception to decision-making. *Journal of Cognitive Neuroscience*, *13*, 454-461.
- Versace, R. Brouillet, D., & Vallet, G. (2018). *Cognition Incarnée : une cognition située et projetée*. Bruxelles : Mardaga.

Remerciements

Les auteurs remercient chaleureusement Helene Intraub pour ses remarques pertinentes, le partage de ses clichés et pour leur avoir permis d'adapter la Figure 2 de ses travaux.