



HAL
open science

Anticipation et accumulation active d'information sensorielle dans la prise de décision en situations de vision normale et dégradée

Boris Quetard

► **To cite this version:**

Boris Quetard. Anticipation et accumulation active d'information sensorielle dans la prise de décision en situations de vision normale et dégradée. Psychologie. Université Clermont Auvergne [2017-2020], 2018. Français. NNT : 2018CLFAL006 . tel-02058194

HAL Id: tel-02058194

<https://theses.hal.science/tel-02058194>

Submitted on 5 Mar 2019

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.



Université Clermont-Auvergne
École doctorale Lettres, Sciences Humaines et Sociales

THÈSE
Pour l'obtention du grade de Docteur en Psychologie

Anticipation et accumulation active d'information sensorielle dans la prise de décision en situations de vision normale et dégradée

Présentée par Boris QUÉTARD

Dirigée par
Martial MERMILLOD, Marie IZAUTE et Jean-Charles QUINTON

Soutenue le 10 Avril 2018

Membres du jury :

Catherine GABAUDE, Rapporteur, Chargée de recherche HDR, IFSTTAR Bron

Thérèse COLLINS, Rapporteur, Professeure, LPP, Université Paris Descartes

Michel DHOME, Examineur, Directeur de recherche CNRS, Institut Pascal, Université Clermont Auvergne

Martial MERMILLOD, Directeur de thèse, Professeur, LPNC, Université Grenoble-Alpes

Marie IZAUTE, Co-directrice de thèse, Professeure, LAPSCO, Université Clermont Auvergne

Jean-Charles QUINTON, Co-directeur de thèse, Maître de conférence, LJK, Université Grenoble-Alpes

Invités :

Michèle COLOMB, Invitée, Directrice de recherche, CEREMA, Clermont Ferrand

Giovanni PEZZULO, Invité, Professeur, ICST, Rome

Ce travail a bénéficié d'une aide de l'État gérée par l'Agence Nationale de la Recherche au titre du programme Investissements d'avenir dans le cadre du projet LabEx IMobS3 (ANR-10-LABX-16-01), d'une aide de l'Union Européenne au titre du Programme Compétitivité Régionale et Emploi 2007-2013 (FEDER – Région Auvergne) et d'une aide de la Région Auvergne.

Remerciements

Tout d'abord, merci à Thérèse Collins, Catherine Gabaude et Michel Dhome, membres du jury de cette thèse, pour avoir accepté de prendre de leur temps pour lire, évaluer et discuter de mon travail.

Je remercie chaleureusement Martial Mermillod, Marie Izaute et Jean-Charles Quinton pour m'avoir accompagné sur la durée de mon doctorat. Merci de m'avoir fait confiance pour mener à bien ce projet de recherche, pour vos conseils précieux lors de la conduite des travaux menés durant tout ce temps, pour votre aide à chaque étape de ce parcours, autant sur des aspects techniques que théoriques, et pour votre patience. Merci pour m'avoir permis de travailler sur ces thématiques, j'ai énormément appris avec vous (mais ai encore tant à apprendre). Merci pour m'avoir accompagné au bout de ce parcours qui, j'espère, n'est qu'un prélude.

Un grand merci à Michèle Colomb, Giovanni Pezzulo et Laura Barca, collaborateurs associés au projet de recherche dans lequel s'inscrivait mon travail. Merci à Michèle pour son accompagnement sur les aspects applicatifs de ce travail, notamment lors de la mise en place de nos expériences en situation de brouillard. Merci à Giovanni et Laura pour leurs conseils précieux et leurs apports aux articles présentés dans cette thèse, ainsi que pour d'enrichissantes discussions sur les aspects relatifs à la modélisation et aux méthodes de mouse-tracking. Merci à eux pour m'avoir accueilli dans leur laboratoire à Rome à deux reprises.

Je tiens aussi à remercier les personnes avec qui j'ai pu collaborer, et qui m'ont aidé à divers niveaux dans mes expériences. En particulier : Jean-Luc Bicard et Philippe Morange (avec mes pensées à sa famille) pour leur aide technique lors de la création de scènes visuelles avec brouillard produit sur la plateforme du CEREMA, M. et Mme. Mialon pour le prêt d'animaux empaillés employés dans ces mêmes scènes, Owen Kevin Appadoo pour son aide sur la construction de scènes 3D, Mélanie Marcotte et Jessica Guilhe pour leur aide dans la préparation et les passations de l'Étude 2.

Merci à tous les membres du LAPSCO, qui ont toujours été disponibles pour répondre, ou tenter de répondre à mes questions, ou simplement en discuter. Notamment, merci à Alain Méot, Laetitia Silvert et Ludovic Ferrand. Merci à Sophie Monceau, Pierre Chausse et Julien Guégan pour leur aide technique précieuse.

Merci à Kenny et Marlène pour leur amitié sur toute la durée de cette thèse. Et un grand merci collectif à tous les collègues et amis, doctorants et assimilés passés et présents, trop nombreux pour ne pas risquer d'omettre quelqu'un, mais m'ayant tous, individuellement, aidé personnellement et professionnellement. Merci à tous pour avoir égayé cette période de ma vie, au labo comme à l'extérieur, merci pour le soutien que vous m'avez apporté et pour les souvenirs passés et, je le souhaite, futurs.

Merci à Jason Webley dont la musique m'a accompagné durant tout mon doctorat.

Merci à Elmo et Crocro pour avoir épargné les canards et ne pas avoir effacé mon travail (avant de partir embêter d'autres gens), et merci à Chat! et Hedgy pour les encouragements durant toute la fin de la thèse, sur la rédaction du présent manuscrit.

Enfin, merci à ma famille, Robin et Jordy, Marie-Claude et Patrice, sans qui rien de tout cela n'aurait été possible. Merci pour les moments passés m'ayant permis de me ressourcer, merci votre soutien durant cette période, durant mes études et durant ma vie.

Table des matières

Résumé	4
Abstract	5
Glossaire des abréviations	6
Introduction	7
Chapitre 1	11
Traitement de l'information visuelle et perception de scènes visuelles	
1. Intégration de l'information visuelle et modèles neurophysiologiques.....	11
1.1. Vision centrale et vision périphérique	11
1.2. Intégration progressive de l'information visuelle et traitement <i>coarse-to-fine</i>	12
1.3. Catégorisation rapide de scènes et d'objets	13
1.4. Influences <i>top-down</i> et le modèle de Bar	14
1.5. Résumé et conclusions	17
2. Sélection de l'information et attention.....	18
2.1. Attention exogène et attention endogène	19
2.2. Attention couverte et attention ouverte	20
2.3. Attention et recherche visuelle	21
2.3.1. <i>L'attention comme un processus de sélection sériel : Modèle d'intégration des caractéristiques et</i>	
<i>Modèle de recherche guidée</i>	22
2.3.2. <i>L'attention comme un processus compétitif : Hypothèse de compétition biaisée</i>	23
2.4. Résumé et conclusions	25
3. Exploration de scènes visuelles et mouvements des yeux dans la recherche visuelle.....	26
3.1. Modèles <i>bottom-up</i> pour l'allocation du regard	26
3.1.1. <i>Cartes de saillances visuelle bottom-up</i>	27
3.1.2. <i>Limites et reformulation de la saillance bottom-up</i>	28
3.2. Influences <i>top-down</i> dans l'exploration visuelle : les connaissances <i>a priori</i>	29
3.2.1. <i>Connaissances relatives au contexte visuel</i>	30
3.2.2. <i>Connaissances sur l'apparence de la cible</i>	32
3.3. Résumé et conclusion	34

Prise de décision et collecte d'information pour la perception visuelle : Méthode expérimentale intégrant des méthodes de *mouse-tracking* et d'oculométrie

1. La prise de décision perceptuelle.....	36
1.1. Modèles de prise de décision perceptuelle	37
1.1.1. <i>Théorie de la détection du signal</i>	37
1.1.2. <i>Modèles d'échantillonnage séquentiel</i>	39
1.1.3. <i>Accumulation d'évidence sensorielle et attentes</i>	41
1.2. Mécanismes actifs d'accumulation d'information	42
1.2.1. <i>Compromis vitesse-exactitude</i>	43
1.2.2. <i>Accumulation d'évidence via les mouvements des yeux</i>	46
1.3. Résumé et conclusion	48
2. Problématique et méthodologie générale.....	50
2.1. Problématique générale	50
2.2. Mesurer la prise de décision : le paradigme de <i>mouse-tracking</i>	53
2.3. Présentation des études expérimentales	56
2.3.1. <i>Évaluer la collecte active d'évidence via le compromis vitesse-exactitude</i>	57
2.3.2. <i>Évaluer la collecte active d'évidence via les mouvements des yeux</i>	59

Études expérimentales

1. Étude 1 : Effets conjugués de l'incertitude perceptive et des attentes générées par le contexte dans une tâche de choix forcé binaire.....	63
Article 1	65
<i>Tracking the time course of perceptual decisions reveals the effects of contextual and noisy sensory information on the speed and accuracy of the responses</i>	
1.1. Résumé des résultats de l'Étude 1	111
2. Étude 2 : Influence de l'incertitude visuelle et de l'indication contextuel sur l'accumulation d'évidence via les mouvements des yeux et la prise de décision lors de la recherche visuelle.....	112
Article 2	114
<i>Differential effects of visual uncertainty and contextual guidance on perceptual decisions: Evidence from eye and mouse tracking in visual search</i>	
2.1. Résumé des résultats de l'Étude 2	136

3. Étude 3 : Influence du brouillard et des attentes sur l'identité de la cible sur la prise de décision et l'accumulation d'information via les mouvements des yeux.....	137
Article 3	140
<i>Combined effects of expectations and visual uncertainty upon detection and identification of a target in the fog</i>	
Article 4	147
<i>A shape in the fog: Expectations, visual uncertainty and the accumulation of evidence in visual search using a combination of mouse and eye tracking methods</i>	
3.1. Résumé des résultats de l'Étude 3	184
Chapitre 4	185
<u>Discussion générale</u>	
1. Prise de décision perceptuelle pour la catégorisation : influence du contexte et <i>mouse-tracking</i>.....	187
1.1. Dégradation de l'information visuelle et accumulation d'évidence	187
1.2. Prise de décision en présence d'information contextuelle	188
1.3. Compromis vitesse-exactitude dans le paradigme de <i>mouse-tracking</i>	189
2. Détection et vérification de la cible dans la recherche visuelle.....	191
2.1. Effet des attentes quant à la localisation de la cible et à son identité	191
2.2. Effet de l'incertitude sur l'information visuelle	193
2.3. Distinction entre processus de détection et de vérification	195
3. Prise de décision perceptuelle dans la recherche visuelle.....	196
3.1. Accumulation d'évidence pour la réponse <i>cible présente</i>	197
3.2. Accumulation d'évidence pour la réponse <i>cible absente</i>	199
3.3. Modéliser l'accumulation d'évidence pour l'absence de la cible	201
4. Limites et perspectives pour la prise de décision lors de la recherche visuelle.....	203
4.1. Limites méthodologiques et théoriques	203
4.2. Perspectives pour la modélisation de la recherche visuelle comme un processus de prise de décision	205
Conclusion	210
Bibliographie	214
Annexe 1	231
<i>Saliency in the fog : Application to real & 2.5D pictures</i>	
Annexe 2	233
<i>Exemples de scènes visuelles plongées dans le brouillard, employées dans l'Étude 3-article 2</i>	

Résumé

Anticipation et accumulation active d'information sensorielle dans la prise de décision en situations de vision normale et dégradée

Conduire un véhicule dans le brouillard requiert d'intégrer de l'information visuelle bruitée avec des attentes sur la scène routière pour rechercher des indices visuels importants pour la navigation. Les tâches d'identification et de recherche visuelle peuvent être vues comme des processus de prise de décision où l'information est accumulée et où des attentes sur l'objet et son contexte sont intégrées. L'accumulation d'information est souvent modélisée comme un processus passif. Cette thèse vise à mettre en avant des mécanismes actifs, intégrant les attentes sur la cible (sur sa position, sur son identité) et de l'information sensorielle dégradée (e.g., brouillard).

Nous avons employé le paradigme de *mouse-tracking*, permettant d'inférer des aspects dynamiques du processus de prise de décision via les mouvements de la souris d'ordinateur. L'Étude 1 évalue l'effet du contexte dans la catégorisation de cible et suggère un compromis entre rapidité et exactitude de l'accumulation d'évidence pouvant être vu comme influençant activement la décision. Mais elle n'évalue pas directement la collecte active d'évidence.

Les Études 2 et 3 incluent la mesure de la détection et vérification de la cible via les mouvement des yeux lors de la recherche visuelle dans des scènes dégradée. Les attentes sur la localisation (Étude 2) et sur l'identité de la cible (Étude 3) sont manipulées. Ces études éclairent les contributions de la détection et de la vérification dans l'accumulation d'évidence pour la réponse *cible absente* et *cible présente*. Pour conclure, nous proposons une ébauche de modèle de prise de décision intégrant une dynamique entre accumulation d'évidence et système oculomoteur.

Mots-clefs :

Prise de décision, Recherche visuelle, Scènes naturelles, Anticipation, Bruit visuel, Brouillard, Mouvement des yeux, Mouse-tracking

Abstract

Anticipation and active accumulation of sensory evidence in perceptual decision-making with normal and degraded visual information

Driving a vehicle in the fog requires the integration of noisy visual information with expectations about the visual road scene, in order to search for visual clues important for navigating. The visual search and identification of relevant objects can be seen as decision-making processes where sensory information is accumulated and where the expectations about the target object and its context are integrated. The accumulation of information is often modelled as a passive process. This thesis focuses on the contribution of active mechanisms integrating expectations about the target (its identity, its location) with degraded sensory information (with fog or artificial noise).

We used the mouse-tracking paradigm, allowing to infer dynamic aspects of the decision-making process through a computer mouse movements. Study 1 evaluates the effect of the context for categorizing a target and suggests a trade-off between the speed and accuracy of the evidence accumulation process which can be seen as actively influencing the decision. But this study cannot directly evaluate the active collection of evidence.

In Studies 2 and 3, target detection and verification are directly measured through eye movements during visual search tasks in visually degraded scenes. We manipulated the expectations about the location (Study 2) and the target's identity (Étude 3). These studies emphasize the contributions of the detection and verification processes in the accumulation of evidence toward the *target present* and *target absent* responses. In conclusion, we propose the draft of a decision-making model which integrates the dynamics between the accumulation of evidence, and the oculomotor system.

Keywords :

Decision-making, Visual search, Real-world scenes, Anticipation, Visual noise, Fog, Eye movements, Mouse-tracking

Glossaire des abréviations

Français (tous chapitres, excluant les articles)

ASC : Aire sous la courbe
BFS : Basses fréquences spatiales
CGL : Corps genouillé latéral
CVE : Compromis vitesse-exactitude
Circuit K : Circuit koniocellulaire
Circuit M : Circuit magnocellulaire
Circuit P : Circuit parvocellulaire
EEG : Electroencephalogramme
HFS : Hautes fréquences spatiales
IRMf : Imagerie par résonance magnétique fonctionnelle
MD : Maximum de déviation
MEG : Magnetoencephalographie
TDS : Théorie de la détection du signal
V1 : Cortex visuel primaire
V2 : Cortex visuel secondaire

Anglais (articles du chapitre 3, Études expérimentales)

AOI : Area of interest
AUC : Area under the curve
CI : Confidence interval
CV : Causal variable
ddf : Denominator degree of freedom
IT : Initiation time
IV : Independent variable
LMM : Linear mixed(-effects) models
MD : Maximal deviation
MV : Maximal velocity
MV-time : Time of maximal velocity
RT : Response times
SAT : Speed-accuracy trade-off
SE : Standard error

Introduction

Rechercher des yeux un objet dans l'espace est une activité cognitive contribuant à un grand nombre de situations de la vie de tous les jours. Elle implique l'intégration d'information visuelle avec des attentes et des connaissances sur la situation courante dans le but de déterminer la présence d'un objet dans l'environnement ainsi que sa position, ou déterminer qu'il est absent. Cette décision (objet présent ou absent) permet par la suite de choisir en conséquence entre diverses actions pour effectuer la tâche en cours (Land, 2009).

Par exemple, la recherche visuelle est capitale lors de la conduite automobile, notamment pour collecter de l'information importante pour naviguer dans la circulation, repérer d'éventuels obstacles ou diagnostiquer la dangerosité de la situation. De ce fait, des conditions de vision dégradée, provoquées par exemple par des conditions météorologiques défavorables, rendent l'environnement visuel incertain et la recherche visuelle d'information plus difficile, affectant en conséquence la conduite. Par exemple, le brouillard impose aux conducteurs de limiter leur vitesse, la détection d'éventuels obstacles étant plus difficile et plus tardive (Mueller & Trick, 2012). Ces conditions de conduite présentent des risques directs pour les usagers des réseaux routiers. Ainsi, des conditions météorologiques défavorables, telles que la pluie ou le brouillard, ont tendance à augmenter la fréquence et l'intensité des accidents routiers (Billot, El Faouzi, & De Vuyst, 2009; Saffarian, Happee, & de Winter, 2012). Cela sera dû en partie à des déficiences de la recherche visuelle dans ces situations où l'information visuelle est dégradée (Konstantopoulos, Chapman, & Crundall, 2010).

Les difficultés à intégrer de l'information visuelle pour détecter les risques dans l'environnement routier ne sont pas les seuls facteurs pouvant mener à des accidents. Il a été montré que si une partie de l'information capitale pour la conduite se situe dans des régions saillantes de l'environnement visuel (donc raisonnablement facile à trouver), une autre partie se situe de façon consistante dans des régions non-saillantes (McCarley, Steelman, & Horrey, 2014). Cela rend les connaissances de l'environnement routier capitales pour repérer au bon moment l'information peu visible mais pertinente à la conduite, ces connaissances permettant de générer des attentes sur la position des objets ainsi que leur comportement et mouvement. Ainsi, le taux plus élevé d'accidents chez les conducteurs inexpérimentés a été attribué au fait que ces conducteurs aient un modèle interne des risques routiers plus pauvre et ainsi, une idée moins précise de leurs causes et des indices visuels pouvant permettre de les remarquer ou anticiper (Pradhan, Pollatsek, Knodler, & Fisher, 2009; Underwood, Chapman, Bowden, & Crundall, 2002). De ce fait ils explorent du regard plus rarement les régions où ces indices se trouvent habituellement. Par exemple il a été rapporté que les conducteurs inexpérimentés regardaient moins souvent les rétroviseurs latéraux

Introduction

(Konstantopoulos et al., 2010).

La recherche visuelle permet donc d'accumuler de l'information utile pour un certain nombre de décisions devant être prises en situation de conduite. Celle-ci semble donc revêtir une importance capitale dans la conduite. Sa compréhension est un enjeu majeur à la fois pour la compréhension des mécanismes cognitifs sous-jacents à la perception visuelle et à la prise de décision en situation de conduite (la recherche visuelle en étant une activité prototypique), et pour son application à des questions de sécurité routière impliquant des usagers humains comme pour le développement de systèmes artificiels autonomes ou assistant la conduite. Ces enjeux concernent des questions applicatives aussi diverses que l'anticipation de risques humains dus à l'échec à détecter un danger, la perception visuelle en situation routière où la vision est dégradée (*e.g.*, brouillard, pluie), la recherche d'information tant pour se diriger (quel chemin prendre) que pour contrôler le véhicule (*e.g.*, comment suis-je positionné par rapport aux autres usagers).

C'est dans ce contexte théorique applicatif vaste et dans le cadre d'un financement du Labex IMobS³ (Innovative Mobility: Smart and Sustainable Solutions), que ce travail de thèse a été effectué. Il est issu d'un projet au sein de ce Labex, porté par une collaboration du LAPSCO (Marie Izaute, Martial Mermillod), de l'Institut Pascal (Jean-Charles Quinton) et du laboratoire de Clermont-Ferrand du CEREMA (Michèle Colomb et Frédéric Bernardin). Additionnellement, l'ICST de Rome (Giovanni Pezzulo et Laura Barca) a été associé au projet et enfin, nous pouvons citer le LPNC et le LJK (tous deux à Grenoble) que, respectivement, Martial Mermillod et Jean-Charles Quinton ont intégré au cours de la thèse.

Ce travail de thèse porte plus particulièrement le rôle de mécanismes actifs tels que les mouvements des yeux dans le processus de prise de décision permettant de déterminer si un objet est présent ou absent de l'environnement. Nous avons choisi de nous focaliser sur la recherche visuelle via une approche fondamentale, permettant de traiter globalement la question de la vision dans des scènes naturelles (*e.g.*, photographies d'intérieur/extérieur, de villes/campagnes/environnements sauvages), mais sans perdre de vue des aspects applicatifs autour de la vision en situation de conduite. D'une part, adopter un point de vue fondamental permet d'apporter des éléments de réponse génériques à diverses problématiques et ainsi, à différents enjeux propres à la perception visuelle en milieu routier tels que ceux cités en exemple. D'autre part, nous nous sommes focalisés sur des aspects particulièrement pertinents dans le domaine routier, à savoir éclairer les mécanismes décisionnels à l'œuvre lors de la recherche visuelle, et rendre compte de l'intégration d'information visuelle dégradée (notamment par le brouillard) avec des attentes sur la composition de l'environnement pour déterminer l'identité et la présence d'une cible. Le rapprochement entre les modèles de prise de décision et la recherche visuelle apporte des éléments de discussion sur des tâches spécifiques à la conduite, reposant sur des décisions prises sur la base d'information sensorielle captée dans l'environnement, telles que décider s'il faut freiner ou passer

Introduction

dans un carrefour complexe, ou bien déterminer la direction à prendre (Lemonnier, Brémond, & Baccino, 2014, 2015). Enfin, nous proposons une étude basée thématiquement sur la conduite automobile, impliquant la détection d'objets (piétons et cerf/chevreuil) dans un brouillard produit dans une salle spécialement équipée (Plateforme R&D Brouillard et Pluie du CEREMA), et où les attentes quant aux objets présentés sont générées par des panneaux de signalisation. Ainsi, si par la suite, le présent manuscrit adopte une approche plus fondamentale, il conserve une pertinence certaine pour des thématiques de recherche applicatives sur la perception visuelle et la prise de décision en environnement routier.

Dans une perspective généraliste, le Chapitre 1 vise donc à présenter des notions souvent classiques dans littérature théorique, sur le traitement de l'information visuelle, l'attention visuelle et l'exploration de scènes visuelles via les mouvements des yeux. Ce chapitre cherche avant tout à mettre en avant l'idée que la totalité de l'information visuelle d'une scène naturelle ou d'un objet présenté n'est pas accessible dès le début de l'observation (*e.g.*, lors d'une observation libre, identification d'un objet, recherche visuelle). Elle est intégrée progressivement, au fur et à mesure de l'observation, autant durant une fixation des yeux qu'en bougeant les yeux à différents endroits de la scène visuelle. De plus, des connaissances sur la situation d'observation et des attentes permettent de guider ou faciliter l'intégration de l'information visuelle. Les trois sections composant ce chapitre apportent des éléments théoriques illustrant l'aspect progressif de l'intégration d'information sur la scène visuelle avec des attentes, au niveau perceptif (*i.e.*, traitement de l'information visuelle, Section 1.), attentionnel (*i.e.*, sélection de l'information pertinente à traiter en priorité, Section 2.) et oculomoteur (*i.e.*, allocation du regard vers l'information pertinente, Section 3.). Cette revue permet de présenter les bases théoriques et aspects neurocognitifs justifiant l'emploi de modèles d'identification de cible et de recherche visuelle centrés sur l'accumulation d'information pertinente.

Cette accumulation d'information est effectuée dans le cadre d'un processus de prise de décision sous-jacent à la tâche de recherche visuelle ou d'identification en cours d'exécution. La prise de décision et ses relations avec la recherche visuelle constitue ainsi le cadre théorique direct du présent travail. Le Chapitre 2 vise à définir ce cadre, ainsi qu'à présenter la problématique qui en découle et les méthodes expérimentales employées pour y répondre. Ainsi, il couvre la notion de prise de décision perceptuelle (*i.e.*, le choix entre des réponses alternatives basé sur l'information sensorielle disponible, Section 1.), cherchant à mettre en évidence des mécanismes actifs visant à guider la collecte d'information pertinente tels que l'exploration visuelle de la scène via les mouvements des yeux, ou bien le compromis entre la vitesse de la réponse et son exactitude. La Section 2. donne une vision d'ensemble du travail expérimental de cette thèse, incluant ses objectifs théoriques, le paradigme expérimental employé (*i.e.*, la tâche de *mouse-tracking*, permettant d'inférer différents aspects continus de la prise de décision) et les principales

Introduction

méthodes de mesure (notamment, les mouvements des yeux).

Le Chapitre 3 présente les trois études que comprend cette thèse. L'Étude 1, composée d'un article, inclut trois expériences et présente le paradigme de *mouse-tracking* au travers d'une tâche de catégorisation d'animaux et d'objets plus ou moins bruités. Elle explore en particulier l'effet du contexte visuel sur la reconnaissance et sur le compromis vitesse-exactitude, vu comme une stratégie permettant d'adapter la durée de l'accumulation d'évidence à la quantité d'information disponible et au temps imparti. La seconde étude propose de mesurer directement l'accumulation d'évidence via les mouvements des yeux dans une tâche de recherche visuelle. Elle comprend un article incluant une expérience focalisée sur les effets du bruit visuel et des attentes sur la localisation de la cible, à la fois sur les mouvements des yeux et sur le processus de prise de décision, et cherchant à combiner *mouse-tracking* et oculométrie pour examiner leurs relations. La troisième étude, composée de deux articles incluant chacun une expérience, emploie la même approche dans une tâche combinant recherche visuelle et identification de cible et ayant pour thématique la conduite en conditions de vision dégradée. Ainsi, elle manipule la quantité de brouillard présent dans la scène et induit des attentes sur l'identité de la cible (piéton ou cerf/chevreuil) via des panneaux de signalisation (proportion d'essais congruents vs. incongruents).

Enfin, le Chapitre 4 discute des résultats de ces trois études au regard du cadre théorique de la prise de décision perceptuelle, puis met en avant les limites et les perspectives de ces études via, notamment, la présentation d'un modèle général synthétisant nos résultats ainsi que des modèles computationnels récents combinant prise de décision et mouvements des yeux. En conclusion sont exposées les principales perspectives dans le cadre applicatif de la perception visuelle et la prise de décision en milieu routier, notamment sous des conditions climatologiques dégradant la vision.

Chapitre 1

Traitement de l'information visuelle et perception de scènes visuelles

1. Intégration de l'information visuelle et modèles neurophysiologiques

La perception de l'environnement visuel résulte de la captation de la lumière et du traitement du signal généré au travers du système nerveux. Les photons, après avoir traversé l'œil puis les différentes couches de cellules composant la rétine, sont captés par les photorécepteurs et transformés en signaux électriques. Dès ce moment, le signal lumineux est structuré tout au long de circuits pouvant être vu comme hiérarchiques et *bottom-up* –i.e., de la rétine jusqu'aux aires corticales associatives. De plus, ces circuits intègrent des signaux *top-down* –i.e., des aires associatives vers les aires de plus bas niveau telles que les cortex visuels primaire (V1) et secondaire (V2)– permettant de guider ce traitement hiérarchique et *in fine*, de donner un sens à ce qui est perçu.

Nous emploierons le terme *bottom-up* pour désigner des traitements ascendants, partant de caractéristiques visuelles de bas niveau (*e.g.*, orientations, couleurs, fréquences spatiales) pour structurer une information de plus en plus complexe en des représentations de plus haut niveau (*e.g.*, représentation d'objets ou de scènes visuelles, connaissances ou attentes sur la composition de l'environnement). À l'inverse, nous emploierons le terme *top-down* pour désigner les traitements descendants, où des représentations de haut niveau et des influences cognitives affectent des étapes antérieures du processus de traitement de l'information (Gilbert & Li, 2013). La décomposition et recombinaison progressive du signal s'illustre à différents niveaux de ce traitement hiérarchique. Nous présentons dans cette section certains aspects de ces traitements ascendants et descendants mettant en évidence que différentes parties du signal visuel sont intégrées selon des décours temporels différents découlant de l'organisation des structures neurophysiologiques du système visuel. Cela permet l'émergence d'une intégration progressive d'information, pouvant être guidée à différents niveaux, par des mécanismes *top-down*.

1.1. Vision centrale et vision périphérique

Deux types de photorécepteurs captent différentes parties du signal ; les bâtonnets sont sensibles à un large intervalle de luminances mais pas à la couleur alors que trois types de cônes permettent de distinguer les couleurs mais sont très peu sensibles aux signaux de faible luminance. La répartition spatiale

asymétrique des photorécepteurs dans la rétine –*i.e.* bâtonnets absents au niveau fovéal au profit des cônes (vision centrale) et cônes quasi-absent dans la périphérie au profit des bâtonnets (vision périphérique)– induit une distinction fonctionnelle entre vision centrale et périphérique (Purves et al., 2001). On peut notamment citer deux phénomènes accentuant cette distinction. Ainsi, chaque cellule ganglionnaire (transmettant l'information de la rétine aux structures sous-corticales) de la périphérie rétinienne (majorité de cellules parasol) convoie l'information provenant d'un plus grand nombre de photorécepteurs que pour les cellules provenant de la fovéa (majorité de cellules midget) (Dacey & Petersen, 1992). Par ailleurs, le phénomène de magnification corticale désigne le fait qu'à différents niveaux du cortex visuel, une plus grande surface corticale est allouée à l'information captée au niveau fovéal. Ainsi, dans ces aires visuelles, les neurones traitant de l'information captée en vision périphérique ont des champs récepteurs d'autant plus larges (*i.e.*, un neurone convoie l'information originellement captée par un grand nombre de bâtonnets et cônes) et inversement pour les neurones traitant de l'information captée en vision centrale. La vision centrale permet ainsi une vision plus détaillée alors que la vision périphérique ne permet qu'une vision de faible résolution mais effective dans des environnements sombres ou sur des stimuli ayant une faible luminance (voir Séré, Marendaz, et Hérault, 2000, ou Rosenholtz, Huang, Raj, Balas, et Ilie, 2012, pour des modélisations de cette dégradation de la résolution spatiale en fonction de l'excentricité rétinienne).

1.2. Intégration progressive de l'information visuelle et traitement *coarse-to-fine*

En parallèle, au moins trois circuits neuronaux principaux convoient l'information visuelle de la rétine aux structures sous-corticales (notamment corps genouillé latéral, CGL) et cortex visuels, puis aux aires associatives (Bullier, 2001; Kaplan, 2004). Le circuit magnocellulaire (circuit M, provenant des cellules parasol, ~10 % des cellules du CGL) est composé de cellules ayant une faible sensibilité à la couleur, très sensibles au contraste de luminance, transmettant rapidement l'information et ayant un champ récepteur plus large. À l'inverse, le circuit parvocellulaire (circuit P, provenant des cellules midget, ~80 % des cellules du CGL) est composé de cellules plus sensibles à la couleur, peu sensibles aux contrastes de luminance, transmettant plus lentement l'information et ayant des champs récepteurs plus réduits. Nous excluons de cette revue le circuit koniocellulaire (circuit K) ; en effet, les cellules composant ce circuit sont très hétérogènes et ses fonctions sont moins bien comprises (Kaplan, 2004, 2012).

Cela induit une seconde distinction. Le circuit M convoie une information spatiale de faible résolution spatiale, dite globale, mais dont la vitesse de traitement et le caractère phasique permet d'extraire l'information de mouvement (Kaplan, 2004). Le circuit P convoie une information de haute

résolution spatiale, dite locale, mais de faible résolution temporelle (cellules à activité toniques), le rendant insensible au mouvement. De plus, traiter l'information locale, plus détaillée, implique plus lourdement des connexions horizontales et ainsi des cellules avec une vitesse de conduction plus faible que les celle des cellules ascendantes et descendantes. Les caractéristiques respectives de ces deux circuits génèrent ainsi une asymétrie dans les dynamiques temporelles de traitement de l'information globale et locale (Bullier, 2001). L'information globale d'une scène visuelle peut être vue comme extraite des basses fréquences spatiales (BFS) de l'image et l'information locale extraite des hautes fréquences spatiales (HFS) (Kauffmann, Ramanoël, & Peyrin, 2014; Oliva & Torralba, 2006) bien que d'autres conceptions existent (e.g., Rosenholtz et al., 2012). Enfin, de plus récentes études mettent en avant que l'hypothèse des trois circuits (M, P et K), décrite ci-dessus, est sans doute simpliste (Kaplan, 2012), mais elle nous semble suffisante dans le cadre de cette revue. *A minima*, pointer une asymétrie des dynamiques temporelles entre traitement de l'information globale et locale nous semble raisonnable et étayé expérimentalement (Hegdé, 2008; Kauffmann et al., 2014).

1.3. Catégorisation rapide de scènes et d'objets

Il a été proposé que l'asymétrie des dynamiques temporelles entre traitement global et local permette un décours *coarse-to-fine* du traitement de l'information visuelle (Bullier, 2001; Fabre-Thorpe, 2011; Kauffmann, Chauvin, Guyader, & Peyrin, 2015; Kauffmann et al., 2014; Kveraga, Ghuman, & Bar, 2007). Suivant ce décours, l'information à l'échelle globale est traitée plus tôt et/ou plus rapidement que l'information à l'échelle locale, permettant d'intégrer rapidement une représentation générale de la scène visuelle ou de l'objet perçu (structure globale), puis d'intégrer progressivement les détails (pour une revue critique, voir Hegdé, 2008). La perception peut ainsi être vue comme une chaîne d'inférences basées sur une information graduellement plus complexe à mesure que le temps s'écoule.

L'un des piliers sur lesquels repose le traitement visuel *coarse-to-fine* est la catégorisation rapide de scènes ou d'objets sur la base de leur structure globale, rapidement accessible. Lorsque présentée pendant un très court intervalle (26 ms sans masque), il est ainsi possible de déterminer très rapidement et efficacement la catégorie de base, ou *gist*, d'une scène visuelle –e.g., ville, autoroute, plage, forêt– et sa catégorie supra-ordonnée –e.g., naturelle, artificielle– ou bien la catégorie de base d'une cible –e.g., chat, poisson, chien– et sa catégorie supra-ordonnée –e.g., animal vs non-animal (Fabre-Thorpe, 2011; Rousselet, Joubert, & Fabre-Thorpe, 2005). Le maintien de cet effet lorsque les stimuli sont présentés en vision périphérique, en noir et blanc ou avec un très bas niveau de contraste pourrait indiquer que cette catégorisation rapide se base sur l'information convoyée par les circuits M (Fabre-Thorpe, 2011).

Alternativement, il a été proposé que l'extraction rapide de statistiques résumant les régularités structurales de l'information visuelle centrale comme périphérique permette une description précoce suffisante pour effectuer ce type de catégorisation (Banno & Saiki, 2015; Gerhard, Wichmann, & Bethge, 2013).

Un temps plus long est nécessaire pour distinguer des scènes ou cibles issues des catégories de base comparé aux catégories supra-ordonnées, indiquant la nécessité d'intégrer de l'information plus élaborée (Fabre-Thorpe, 2011). Différents jugements peuvent être effectués à propos d'une scène visuelle, sur la base d'une information plus ou moins complexe. Ainsi, juger des propriétés globales –*e.g.*, dans quelle mesure une scène est ouverte, navigable, d'un climat chaud ou froid, sa profondeur de champ– nécessite une présentation (masquée) plus courte (~34 ms) que juger de sa catégorie de base (~50 ms) (Greene & Oliva, 2009). Juger des propriétés globales nécessite donc une information visuelle moins élaborée (*e.g.*, surfaces horizontales grossières où se tenir, présence d'obstacles pour juger de la navigabilité) que juger de la catégorie de base. Cela suggère que l'information visuelle accumulée à différents instants permet d'effectuer des jugements sur le contenu des scènes nécessitant l'intégration d'information de plus en plus complexe et riche.

De manière schématique et probablement réductrice, cela sous-tend une conception générale du système visuel où l'information est intégrée hiérarchiquement, depuis la rétine, les aires visuelles bas niveau sous-corticales (CGL, colliculus supérieur) et corticales (cortex visuel primaire V1, secondaire V2) jusqu'aux aires associatives temporales et pariéto-frontales. De plus, certaines parties du signal visuel sont intégrées plus rapidement que d'autres. A chaque instant du traitement, des inférences peuvent être effectuées, permettant d'intégrer progressivement des caractéristiques ou propriétés de la scène perçue sur la base de l'information disponible.

1.4. Influences *top-down* et le modèle de Bar

Le traitement de l'information n'est pas seulement intégré de manière *bottom-up* ou ascendante. D'une part, l'asymétrie de vitesse de conduction entre les circuits transmettant l'information globale et locale permet de faciliter l'intégration d'information locale grâce à l'information globale précédemment traitée. Ainsi, il a été proposé que l'information rapidement convoyée par les circuits M depuis les cortex visuels primaire V1 et secondaire V2 vers les aires de plus haut niveau, notamment pariétales et préfrontales, soit projetée vers V1 et V2 de manière à guider le traitement de l'information convoyée plus lentement (Bullier, 2001; Peyrin et al., 2010). D'autre part, si les circuits M comme P convoient majoritairement l'information visuelle vers le cortex visuel primaire V1, une part des cellules ganglionnaires

ont des projections vers des structures sous-corticales telles que le colliculus supérieur (~ 10 % des cellules issues de la rétine) ou le pulvinar, permettant à de l'information visuelle d'atteindre les aires associatives plus rapidement (Baluch & Itti, 2011; Kveraga et al., 2007; Peyrin et al., 2010).

Le modèle de Bar (Bar, 2004, 2007; Kveraga et al., 2007) est un bon exemple de modèle neurophysiologique reposant sur ces mécanismes, portant plus particulièrement sur la reconnaissance d'objets. La Figure 1 reproduit le schéma de ce modèle, issu de (Bar, 2004). Dans ce modèle, les aires corticales bas niveau et les voies de traitement sous-corticales intègrent l'information directement disponible, le *gist* de la scène et de l'objet cible (à partir de leur BFS), et projettent rapidement cette information vers deux voies. La première voie transmet l'information relative à la scène visuelle, catégorisée très rapidement à partir d'une représentation globale. Ainsi, le contexte de la scène est extrait de l'information globale dès les premiers instants du traitement, permettant d'activer les représentations en mémoire des objets associés à ce contexte. L'information imprécise relative à l'objet est intégrée via la seconde voie, où elle est suffisante pour générer des analogies entre l'information *bottom-up* et les représentations d'objets similaires. Dans le cadre de ce modèle, le terme d'analogies est utilisé pour désigner le lien entre l'information visuelle rudimentaire rapidement intégrée sur l'objet observé et les représentations similaires présentes en mémoire. À partir de ces analogies, les représentations les plus probables compte tenu du *gist* de l'objet, sont activées. Le croisement entre représentations candidates quant à l'identité de la cible et contexte permet de prédire l'identité de la cible. Ces prédictions et la composition de la scène guident par la suite l'intégration plus lente et progressive d'information locale complexe (HFS).

Dans le modèle de Bar, l'information globale est ainsi vue comme base sur laquelle des prédictions sur ce qui est observé sont générées à un stade précoce du processus d'intégration visuelle (à un moment où peu d'information locale a pu être intégrée). Ces prédictions sont mises en compétition et au fur et à mesure que l'information visuelle est intégrée et accumulée, sont graduellement mises à jour et exclues si elles en viennent à être infirmées (Trapp & Bar, 2015). Ce modèle rend compte de l'avantage comportemental pour la catégorisation de scènes présentées suivant un schéma *coarse-to-fine* plutôt que *fine-to-coarse* (Kauffmann et al., 2015; Schyns & Oliva, 1994) mais aussi de la facilitation de la reconnaissance des objets lorsque le contexte dans lequel ils sont présentés est congruent (Auckland, Cave, & Donnelly, 2007; Biederman, Rabinowitz, Glass, & Stacy, 1974; Davenport & Potter, 2004; S. E. Palmer, 1975) –*e.g.*, un tigre dans la jungle par opposition à un tigre dans une salle de bain.

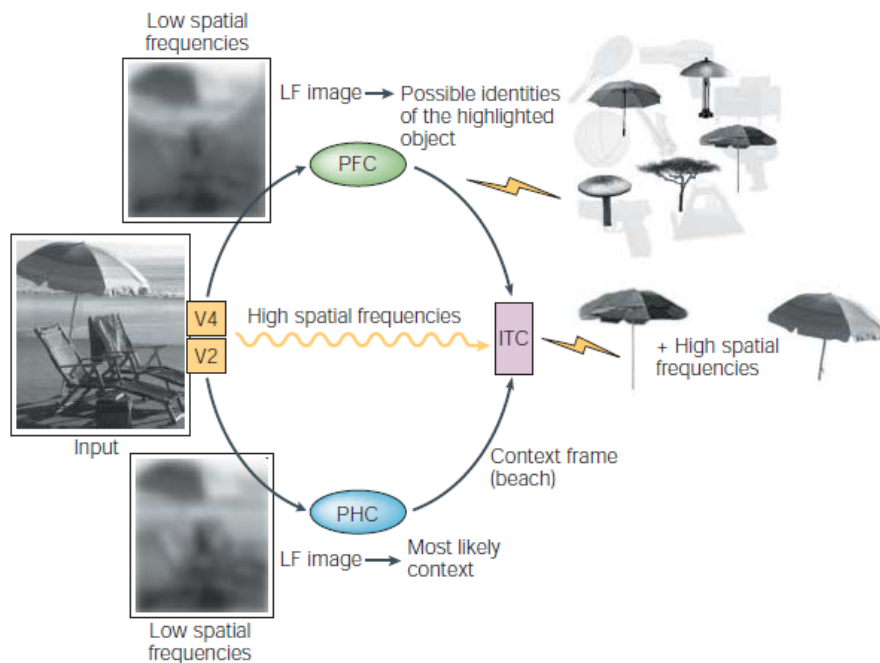


Figure 1 : Schéma du modèle de Bar (provenant de Bar, 2004). L'information visuelle basse fréquence spatiale (LF, *low frequency*) de l'objet cible est extraite vers le cortex préfrontal (PFC). Sur la base de cette information des analogies sont générées (identités possibles de l'objet). L'information visuelle basse fréquence spatiale du contexte est extraite vers le cortex para-hippocampique (PHC) pour déterminer sa catégorie la plus probable. Le croisement de ces deux sources d'information globale permet de générer des prédictions sur l'identité probable de la cible. Ces prédictions sont intégrées au niveau du cortex inféro-temporal (ITC), et guident l'intégration lente et progressive de l'information visuelle haute fréquence spatiale.

Ainsi, des connections descendantes permettent à la fois de faciliter le traitement de l'information visuelle à un bas niveau de traitement (aires visuelles V1, V2), mais aussi les fonctions de haut niveau telles que la reconnaissance d'objet par l'intégration de l'information visuelle entrante avec des représentations stockées en mémoire (objets et catégories de scènes visuelles). Jusqu'à présent, nous nous sommes focalisés sur l'idée que l'information visuelle est traitée selon un décours *coarse-to-fine*. Néanmoins, Oliva et Schyns (1997) ont montré que les demandes spécifiques à la tâche de reconnaissance pouvaient induire un décours *fine-to-coarse* où l'information à l'échelle locale (HFS dans leur expérience) est traitée prioritairement lorsque plus diagnostique pour reconnaître la scène et inversement dans le cas où l'information à l'échelle globale est plus diagnostique. Selon ces auteurs, dans chacun de ces cas, l'information non prioritaire est décodée rapidement, à un bas niveau de traitement (intégration visuelle) et sans reconnaissance de la scène. Lorsque la tâche n'impose pas de contrainte particulière, la scène serait traitée par défaut selon un décours *coarse-to-fine*, cela pouvant s'expliquer par de l'information étant, par

défaut, plus diagnostique à l'échelle globale pour reconnaître certaines catégories de scènes visuelles (Mermillod, Guyader, & Chauvin, 2005) ou expression faciales émotionnelles (Mermillod, Bonin, Mondillon, Alleysson, & Vermeulen, 2010). Ainsi, le traitement visuel pourrait être biaisé vers un traitement global, cette information étant généralement plus diagnostique. Oliva et Schyns (1997) mettent ainsi en avant un mécanisme *top-down* permettant de sélectionner l'échelle à laquelle l'information pertinente pour la tâche de reconnaissance se situe probablement. Autrement dit, ils proposent l'implication de mécanismes attentionnels.

1.5. Résumé et conclusions

Les travaux exposés jusqu'à présent décrivent (de façon simplifiée) un processus d'intégration de l'information visuelle s'effectuant hiérarchiquement de façon *bottom-up*, et intégrant des projections en provenance des aires associatives de manière *top-down*. L'aspect nous semblant le plus important dans cet exposé, est la mise en avant de l'aspect continu et progressif de l'intégration d'information visuelle, la totalité de l'information captée n'étant pas forcément disponible à tout moment. D'une part, l'asymétrie de vitesse de traitement entre différentes caractéristiques de l'information visuelle (*e.g.*, échelle, couleur) permet à l'information intégrée plus rapidement de servir de base à la génération de prédictions sur ce qui est vu, via des projections *top-down*. Cela permet d'anticiper le contenu d'une scène sans que toute l'information disponible ne soit intégrée. D'autre part, comme le soulignent Oliva et Schyns (1997), le décours de l'intégration d'information est guidé par la tâche de reconnaissance, impliquant probablement un mécanisme de sélection d'information ; toute l'information captée n'est pas nécessairement intégrée jusqu'aux aires associatives et une sélection de l'information pertinente est sans doute nécessaire pour effectuer les différentes tâches pour lesquelles la vision est nécessaire. Traditionnellement, c'est le rôle attribué à l'attention visuelle.

2. Sélection de l'information et attention

Pour mener à bien une tâche quelconque, l'information pertinente doit être accessible et intégrée correctement alors que l'information sans rapport avec la tâche, voire pouvant interférer avec celle-ci, doit pouvoir être atténuée ou ignorée. Dans ce but, « l'attention nous permet de traiter sélectivement la vaste quantité d'information à laquelle nous sommes confrontés, en priorisant certains aspects de l'information tout en ignorant d'autres, par focalisation sur un certain endroit ou un certain aspect de la scène visuelle » (Carrasco, 2011, p. 1484). Ainsi, cette notion repose sur l'idée que les ressources de traitement sont limitées (Beck & Kastner, 2009; Broadbent, 1958). En conséquence, la totalité de l'information visuelle entrante ne peut être intégrée instantanément. Ainsi, l'attention visuelle permet de sélectionner l'information à traiter prioritairement. Elle est souvent décrite comme un projecteur (*attentional spotlight*) éclairant une partie du champ visuel (Posner, 1980; Theeuwes, 1989) ou une lentille (*attentional lense*) zoomant sur une partie du champ visuel (Eriksen & St. James, 1986). En plus de permettre une focalisation sur des endroits particuliers (attention spatiale), l'attention permet aussi de sélectionner des caractéristiques visuelles particulières présentes dans la scène visuelle (attention pour les caractéristiques, *feature-based attention*) –e.g., couleur rouge, orientation verticale, forme particulière, mouvement droite-gauche. L'attention pour les caractéristiques permet ainsi d'augmenter les ressources allouées à l'intégration de certaines de ces caractéristiques au détriment d'autres (Carrasco, 2011; Theeuwes, 2013). Cela permet notamment de faciliter la recherche d'objets particuliers, sans *a priori* sur leur localisation dans l'espace. Enfin, les objets –définis comme des entités visuelles structurées dont la surface est limitée par des contours définis, ayant à la fois une valeur sémantique et fonctionnelle (Cavanagh, 2011)– sont en mesure de capturer l'attention de par leur potentielle pertinence pour la compréhension de la scène visuelle et/ou pour la tâche en cours d'exécution (Chen, 2012; Malcolm & Shomstein, 2015; Nuthmann & Henderson, 2010).

Deux distinctions importantes peuvent être faites concernant l'attention. La distinction entre attention exogène vs. endogène renvoie à la manière dont est traitée l'information –*bottom-up* ou *top-down*– guidant la sélection attentionnelle alors que la distinction attention couverte vs ouverte (*covert vs overt*) renvoie aux mécanismes par lesquels les traitements perceptifs sont améliorés aux endroits sélectionnés attentionnellement (e.g., les mouvements des yeux).

2.1. Attention exogène et attention endogène

L'attention exogène renvoie à un système *bottom-up* où l'attention est capturée par les caractéristiques de l'information visuelle entrante, par le mouvement, etc. L'attention est ainsi automatiquement capturée par la saillance perceptive d'un stimulus par rapport à son environnement et aux autres stimuli présentés (*i.e.*, dans quelle mesure il s'en distingue). Au contraire, l'attention endogène renvoie à un système *top-down* où l'attention est guidée par les objectifs, intentions et attentes de l'individu. Ainsi, Posner (1980) a montré qu'indiquer la position d'une cible préalablement à sa présentation attirait l'attention, que l'indice indique la position de la cible par des stimuli présentés en vision périphérique (exogène) ou par des flèches présentées en vision centrale (endogène).

L'attention exogène est souvent comprise comme automatique alors que l'attention endogène est souvent assimilée à une attention vue comme intentionnelle (Carrasco, 2011). Ces notions ne sont pourtant pas si faciles à définir. S'il existe un certain consensus définissant l'attention exogène comme automatique, l'attention endogène n'est pas forcément intentionnelle ou volontaire ; Baluch et Itti (2011) distinguent l'attention endogène volontaire (*volitional*) et l'attention endogène systématique (*mandatory*). Cette dernière désigne des « processus attentionnels influençant les processus sensoriels de manière persistante et automatique » (p. 210). Par exemple, avec l'expertise, l'attention aura tendance à être dirigée vers les endroits d'une scène connus pour être pertinents dans la tâche, sans forcément impliquer de volonté à porter l'attention à cet endroit (*e.g.*, situation de conduite, Underwood, Chapman, Bowden, & Crundall, 2002).

D'autre part, le fait que la capture attentionnelle par un objet (et/ou ses caractéristiques visuelles) relève plutôt d'un système endogène ou exogène est discuté dans la littérature. Traditionnellement considérée comme exogène, il a été montré que la capture attentionnelle par des stimuli saillants pouvait être modulée par les objectifs de la tâche (Folk, Remington, & Johnston, 1992), dans le sens où une caractéristique saillante de la cible ou d'un distracteur capture moins l'attention lorsque cette caractéristique n'a pas de valeur par rapport aux buts de la tâche, que lorsqu'elle est contingente à ces buts (*e.g.*, faible capture par un distracteur rouge lorsque la couleur rouge la rend distincte des autres distracteurs mais que la couleur n'est pas une caractéristique distinguant la cible des distracteurs). Les auteurs concluent à l'implication de mécanismes *top-down* dans la capture attentionnelle par l'information visuelle, bien que cette conclusion soit discutée (Theeuwes, 2013; Theeuwes & Van der Burg, 2013). De même, il a été rapporté que lors de la perception de scènes visuelles naturelles, la capture attentionnelle par des objets apparaissant ou changeant de couleur brutalement était réduite en situation de double tâche, suggérant l'implication de mécanismes *top-down* (Matsukura, Brockmole, Boot, & Henderson, 2011).

2.2. Attention couverte et attention ouverte

La distinction entre attention exogène et endogène met en évidence l'origine de l'information sur laquelle se base la sélection attentionnelle (information issue de traitements *bottom-up* vs. *top-down*). La distinction entre attention couverte et ouverte repose sur les mécanismes par lesquels l'attention facilite le traitement de l'information visuelle. Plus précisément, l'attention ouverte vise à sélectionner une zone ou un objet en effectuant un mouvement du regard à l'endroit correspondant pour capter l'information visuelle avec la partie fovéale de la rétine et profiter de sa meilleure résolution, améliorant la perception à l'endroit fixé (Kietzmann & König, 2015). Au contraire, l'attention couverte vise à sélectionner une zone ou un objet du champ visuel pour en faciliter le traitement sans porter le regard à cet endroit (Carrasco, 2011). Ainsi, à l'endroit sélectionné, l'attention couverte permettrait d'augmenter la sensibilité au contraste ou la résolution spatiale par modulation du flux d'information dans le système visuel, permettant d'améliorer la sensibilité aux HFS en vision périphérique ou bien l'acuité visuelle. Pour permettre cela, différents mécanismes ont été proposés, parmi lesquels la réduction du bruit externe (bruit sur l'information visuelle entrante) par filtrage du bruit concomitant au signal d'intérêt et/ou de l'information distractive autour du focus de l'attention, l'amélioration de la captation du signal (réduisant le bruit interne) ou la modulation (réduction) des champs récepteurs de groupes de neurones à différents niveaux du traitement visuel. Pour une revue sur l'attention visuelle couverte, voir Carrasco (2011).

Cette distinction entre attention ouverte et couverte n'implique pas nécessairement deux systèmes indépendants ou deux systèmes distincts. Ainsi, bien que le découplage entre focus attentionnel et destination d'une saccade oculaire soit possible (*e.g.*, dans les tâches expérimentales imposant une croix de fixation), il est communément admis qu'un déplacement de l'attention couverte précède un mouvement des yeux à l'endroit correspondant (Rayner, McConkie, & Ehrlich, 1978; Zhao, Gersch, Schnitzer, Doshier, & Kowler, 2012). De même, la planification d'une séquence de saccades induit un déplacement attentionnel pré-saccadique aux destinations prévues, facilitant la discrimination à ces endroits (Gersch, Kowler, Schnitzer, & Doshier, 2009). Rizzolatti et collaborateurs (Rizzolatti, Riggio, Dascola, & Umiltà, 1987) ont émis l'hypothèse que l'allocation de l'attention spatiale couverte découle de la planification des mouvements des yeux (théorie prémotrice de l'attention). En ce sens, lorsque l'attention couverte est engagée dans une zone cible, cette localisation doit être inhibée pour effectuer un mouvement des yeux vers une autre zone, déviant la trajectoire de la saccade à l'opposé de la zone cible (Van der Stigchel & Theeuwes, 2007). De plus, il a été montré que la microstimulation de cellules impliquées dans la planification et le contrôle des saccades oculaires (aire V4, champ oculaire frontal, colliculus supérieur) améliorait le traitement de l'information visuelle dans leur champ récepteur lorsque la stimulation était trop faible pour générer une saccade (Moore, Armstrong, & Fallah, 2003; Muller, Philiastides, & Newsome, 2005). *A minima*, attention

ouverte et couverte partagent donc des structures neuronales et les conceptions les plus radicales postulent que leurs réseaux neuronaux sont confondus (Rizzolatti et al., 1987).

Il est cependant possible de dissocier la destination d'une saccade et le focus de l'attention couverte. C'est le cas dans une situation de double tâche où les participants doivent effectuer une saccade et discriminer une cible située dans la zone de destination de la saccade ou en dehors de celle-ci. Lorsque position de la cible et destination de la saccade sont dissociées, on observe un compromis entre latence et précision de la saccade d'une part, et performance perceptive d'autre part (Zhao et al., 2012). Cela souligne que l'attention est nécessaire à la planification des mouvements des yeux, et qu'attention ouverte et couverte entrent en compétition pour des ressources attentionnelles communes. La relation entre ces deux formes d'attention suggère que les fixations du regard puissent permettre de séquencer le traitement des stimuli sélectionnés par l'attention, ou au moins les plus importants (Zelinsky, 2008).

2.3. Attention et recherche visuelle

L'attention visuelle permet ainsi de sélectionner et traiter préférentiellement l'information pertinente tout en réduisant l'influence de l'information distractive intégrée parallèlement, de manière à faire face à des capacités de traitement limitées. La tâche de recherche visuelle compte parmi les paradigmes les plus utilisés pour étudier l'attention. Cette tâche vise à rechercher une cible parmi un nombre variable de distracteurs lui ressemblant plus ou moins (Treisman & Gelade, 1980). La ressemblance est manipulée en définissant la cible et les distracteurs sur un certain nombre de dimensions telles que la forme, couleur, vitesse ou direction du mouvement, ou l'orientation des éléments composant les stimuli. Les caractéristiques (*features*) désignent les valeurs prises sur ces dimensions (*e.g.*, courbe, rouge, horizontal). Ainsi, les lettres L et T comportent les mêmes caractéristiques –barres horizontale et verticale– sur la dimension « orientation » et les caractéristiques de la lettre X sur cette dimension –barres diagonales– la distingue de ces deux lettres.

Dans cette tâche on distingue la recherche basée sur les caractéristiques (*feature search*, recherche de caractéristiques) où la cible est distincte de tous distracteurs sur les caractéristiques manipulées dans la tâche (*e.g.*, barre horizontale parmi des barres verticales, L rouge parmi des O verts), et la recherche basée sur les conjonctions de caractéristiques (*conjunction search*, recherche de conjonctions) où la cible partage chacune de ses caractéristiques avec les distracteurs et où seule la conjonction de ses caractéristiques permettent de la distinguer (*e.g.*, L rouge parmi des L verts et des O rouges). Dans le premier cas la cible est détectée très rapidement et le nombre de distracteurs n'influence pas la performance (effet *pop-out*). Par contraste, dans le second cas, la cible est plus longue à être détectée et la performance se dégrade avec

l'augmentation du nombre de distracteurs (plus d'erreurs, temps de réponse plus longs). Différents modèles ont été proposées pour expliquer cette distinction entre recherche basée sur les caractéristiques et basée sur les conjonctions.

2.3.1. *L'attention comme un processus de sélection sériel : Modèle d'intégration des caractéristiques et Modèle de recherche guidée*

Une des théories des plus influentes sur le rôle de l'attention dans la recherche visuelle est le Modèle d'intégration des caractéristiques (*Feature Integration Theory*, Treisman & Gelade, 1980). Il postule la mise en œuvre de deux types de traitements : un processus parallèle extrait les caractéristiques des stimuli rapidement au niveau pré-attentif et un processus sériel permettant la combinaison des caractéristiques de chaque stimulus par focalisation de l'attention sur la localisation du stimulus (ancrage des caractéristiques, *feature binding*). Lors d'une recherche de caractéristiques, combiner les caractéristiques n'est pas nécessaire pour détecter la cible et le processus parallèle est suffisant alors que lors d'une recherche de conjonctions, le processus parallèle ne permet pas de détecter la cible, et focaliser l'attention sériellement sur chaque stimulus est nécessaire pour combiner ses caractéristiques et déterminer s'il est un distracteur ou la cible.

Néanmoins, la distinction entre processus parallèle vs. sériel n'est pas aussi claire. Cave et Wolfe (1990) ont effectué une revue des limites de cette distinction. Les auteurs pointent que certaines recherches de conjonctions, impliquant par exemple l'échelle des stimuli ou le plan sur lequel ils sont placés, ne sont pas influencées par le nombre de distracteurs. De même, le modèle ne rend pas compte de différences interindividuelles : certains participants voient leur performance plus ou moins dégradée par le nombre de distracteur lors de recherches de conjonctions alors que d'autres ne sont pas affectés par cet effet. Ils pointent ainsi que l'absence ou la présence de l'effet du nombre de distracteurs n'est pas suffisante pour démontrer l'implication de deux processus distincts, l'un parallèle et l'autre sériel. Il semble exister une gradation de cet effet où, selon les dimensions manipulées dans la recherche de conjonctions et les variations de saillance des stimuli, l'effet du nombre de distracteurs est plus ou moins fort (pouvant être parfois nul avec des variations d'échelle ou de profondeur). Cette gradation n'est pas forcément compatible avec l'implication de ces deux processus distincts.

En lieu et place, les auteurs, dans leur Modèle de recherche guidée (*Guided Search Theory*, Cave & Wolfe, 1990; Wolfe, 1994) proposent une interaction entre processus parallèle et sériel. Pour chaque dimension, les connaissances sur l'apparence de la cible recherchée activent pré-attentivement les caractéristiques pertinentes, produisant des cartes d'activation dont la combinaison permet d'orienter sériellement l'attention vers les stimuli ressemblant le plus à la cible. Ainsi, les processus parallèles et sériels

ne sont plus séparés ; le processus parallèle pré-attentif guide le processus sériel attentif. Si l'on cherche une barre verticale rouge parmi des barres horizontales vertes (recherche de caractéristiques), seule la cible produira une activation sur les cartes des caractéristiques diagnostiques pour cette tâche (couleur et orientation), orientant l'attention vers sa localisation. Si par contre, on cherche une barre verticale rouge parmi des barres horizontales rouges, des barres verticales vertes et des barres obliques bleues (recherche de conjonctions), la cible produira de l'activation sur les deux cartes, les distracteurs partageant une caractéristique avec la cible –horizontal rouge et vertical vert– produiront de l'activation sur les cartes couleur et orientation respectivement, alors que les distracteurs ne partageant aucune caractéristique – obliques bleues– ne produiront d'activation sur aucune carte. Une fois les cartes combinées, la cible reçoit le plus d'activation mais un certain nombre de distracteurs reçoivent une activation moindre. Le modèle postule que l'on ajoute un certain niveau de bruit à ces activations, ce qui peut provoquer la sélection de ces distracteurs avant la cible, ralentissant d'autant plus sa détection que le nombre de distracteurs activés est élevé. Ici, ce sont principalement les objectifs de la recherche visuelle (apparence de la cible) qui guident l'allocation de l'attention en spécifiant les caractéristiques pertinentes pour discriminer la cible. Dans une version révisée du modèle (Wolfe, 1994), Wolfe distingue des cartes de caractéristiques *bottom-up* et *top-down*, pour chaque dimension. Comme dans la première version du modèle, les cartes de caractéristiques *top-down* activent les stimuli composés des caractéristiques de la cible. Les cartes *bottom-up* activent, dans chaque dimension, les stimuli se distinguant visuellement des stimuli environnants et des autres stimuli en général : une barre verticale rouge ayant pour voisins des barres verticales vertes aura une forte activation sur la carte *bottom-up* de la couleur mais pas sur celle de l'orientation. Les cartes *top-down* et *bottom-up* de chaque dimension sont ensuite combinées en une carte d'activation guidant l'attention.

2.3.2. L'attention comme un processus compétitif : Hypothèse de compétition biaisée

L'attention peut aussi être vue, non pas comme un système dédié à sélectionner sériellement les stimuli à discriminer (recherche visuelle) mais comme une « propriété émergente de lentes interactions compétitives travaillant en parallèle au travers du champ visuel » (Desimone & Duncan, 1995, p. 217). En d'autres termes, les auteurs postulent que l'attention visuelle découle du traitement de l'information visuelle au lieu d'être une structure spécifique orientant ce traitement et qu'ainsi, elle est distribuée à tous les niveaux du traitement visuel. Ainsi, l'Hypothèse de compétition biaisée (*Biased Competition Hypothesis*) propose un modèle où les stimuli (ou leurs représentations) sont en compétition pour accumuler de l'activation neuronale, dans un contexte où les ressources de traitement sont limitées (Beck & Kastner, 2005, 2009; Desimone & Duncan, 1995; Fernandez-Duque & Johnson, 2002). Cette compétition est biaisée à la fois par des mécanismes *bottom-up*, visant notamment à séparer les stimuli/formes/objets de leur

arrière-plan (ou contexte), et par des mécanismes *top-down* sélectionnant les stimuli ou objets pertinents à partir de leurs caractéristiques visuelles connues.

Dans le cadre particulier de la tâche de recherche visuelle, une représentation de la cible (*target template*) est stockée en mémoire de travail et elle est utilisée pour biaiser la compétition, par exemple, en faveur des stimuli dont les caractéristiques lui correspondent ou en faveur des stimuli localisés à des endroits pertinents pour la tâche (Desimone & Duncan, 1995). Dans le cas d'une recherche de caractéristiques, il n'y a que peu de compétition entre la cible et les distracteurs car ceux-ci ressemblent très peu à la représentation de la cible et la sélection est fortement biaisée à l'encontre de la cible. Pour une recherche de conjonctions, cible et distracteurs sont similaires et l'avantage de la cible est réduit puisque les distracteurs bénéficient aussi du biais généré par la représentation de la cible.

Globalement, l'hypothèse de compétition biaisée partage des similitudes avec le modèle de recherche guidée. Ainsi, les deux modèles conceptualisent les influences *bottom-up* par une mesure de la dissemblance entre un stimulus et son environnement plus ou moins direct, et les processus *bottom-up* comme les processus *top-down* influencent la sélection en modulant l'activation neuronale. Cependant, la différence majeure est que le modèle de recherche guidée s'appuie sur des cartes rétinotopiques globales pour moduler l'activation, et suppose donc l'existence de structures attentionnelles dédiées. Dans les modèles reposant sur l'hypothèse de compétition biaisée, la compétition s'opère directement au niveau des structures visuelles, un de ses marqueurs étant la réduction de l'activité des neurones dont le champ récepteur est partagé par deux (ou plus) objets concurrents (Beck & Kastner, 2005, 2009). Autrement dit, les stimuli présents dans le même champ récepteur s'inhibent mutuellement.

Beck et Kastner (2009) ont passé en revue les bases neuro-physiologiques des influences endogènes et exogènes sur ces mécanismes de compétition. Ainsi, il a ainsi été montré que l'influence *top-down* modulait la compétition dans le champ récepteur des neurones en faveur du stimulus attendu (notamment en augmentant son contraste, Reynolds & Chelazzi, 2004), et pouvait augmenter le niveau d'activité de référence (*baseline*) des neurones codant une localisation sélectionnée. Beck et Kastner relèvent deux mécanismes *bottom-up*. Premièrement, la compétition est fortement réduite lorsqu'un stimulus est très distinct des autres stimuli dans un champ récepteur (effet *pop-out*) par comparaison avec une situation où tous les stimuli se ressemblent ou sont tous très distincts les uns des autres. Deuxièmement, la compétition est réduite entre les stimuli d'un groupe de stimuli similaires comparé aux stimuli d'un groupe hétérogène, entraînant des phénomènes de regroupement (*grouping*). De plus, plusieurs groupes de stimuli sont en compétition dans un même champ récepteur, permettant une sélection *bottom-up* des objets via ces effets de regroupement et segmentation.

2.4. Résumé et conclusions

L'attention, notamment visuelle, est une fonction cognitive fondamentale permettant de traiter prioritairement les objets pertinents pour la tâche en cours. Cette fonction est rendue nécessaire par le fait que toute l'information visuelle composant l'environnement extérieur n'est pas disponible à tout moment et que les ressources disponibles pour la traiter sont limitées. Ainsi, elle est orientée par un ensemble d'influences *top-down* dites endogènes incluant l'apparence des objets à rechercher et leur localisation probable, et d'influences *bottom-up* dites exogènes incluant des indices visuels tels que les caractéristiques visuelles des objets de l'environnement, leur mouvement ou bien dans quelle mesure un objet ou une partie du champ visuel se distingue de ce qui l'entoure. Traditionnellement, l'attention est vue comme une fonction à part entière orientant le traitement visuel et dont les processus mettent en jeu des structures neurophysiologiques spécifiques. Dans une catégorie de modèles, les influences *bottom-up* et *top-down* sont ainsi intégrées en cartes rétinotopiques modulant les traitements et sélectionnant les zones où porter sérieusement l'attention pour un traitement plus détaillé (Modèle d'intégration des caractéristiques, Treisman et Gelade, 1980 ; Modèle de recherche guidée, Cave et Wolfe, 1990 et Wolfe, 1994). Mais elle peut aussi être vue comme émergente de la compétition pour des ressources de traitement neuronal entre les objets (activation), cette compétition étant biaisée par des influences endogènes et exogènes en faveur des objets et localisations spatiales sélectionnées (Hypothèse de compétition biaisée, Beck et Kastner, 2009 et Desimone et Duncan, 1995). Les mouvements des yeux sont centraux dans la recherche visuelle et sont généralement vus comme guidés par l'attention, permettant un traitement plus détaillé de l'endroit ou de l'objet ciblé.

3. Exploration de scènes visuelles et mouvements des yeux dans la recherche visuelle

La tâche de recherche visuelle fait partie des paradigmes majeurs pour l'étude de l'attention, et est au cœur d'un grand nombre d'activités de la vie de tous les jours. Classiquement, ce type de tâches a été abordé en utilisant des stimuli artificiels simples (singletons, patches de Gabor) de manière à contrôler leurs caractéristiques visuelles. Si ces paradigmes ont permis une meilleure compréhension de l'attention visuelle, un intérêt moindre a été porté à l'allocation du regard. Depuis, de nombreuses implémentations de modèles proches du modèle d'intégration des caractéristiques et du modèle de recherche guidée ont été proposées et adaptées pour des scènes visuelles dont les caractéristiques visuelles sont plus écologiques (Borji & Itti, 2013; Tatler, 2009), et cherchant donc à se rapprocher de situations d'observation naturelle. Ces implémentations ont été considérées à la fois comme modélisant l'allocation de l'attention (au sens large) et la localisation des fixations des yeux (attention ouverte) lorsqu'un observateur regarde une scène visuelle pour effectuer différentes tâches (rechercher un objet, regarder librement, mémoriser la scène). Malgré ce déplacement de l'intérêt vers l'exploration visuelle, les mêmes questions demeurent quant à l'influence relative de sources d'information exogènes extraites de façon *bottom-up*, principalement redéfinie sous le terme « saillance *bottom-up* » (Koch & Ullman, 1985), et de sources d'information endogènes désignant les connaissances de l'observateur sur la situation, les objets environnants et la tâche en cours, influençant de manière *top-down* les processus visuomoteurs.

3.1. Modèles *bottom-up* pour l'allocation du regard

Suivant l'approche initiée avec le Modèle d'intégration des caractéristiques (Treisman & Gelade, 1980) et parallèlement au développement de ce modèle et du Modèle de recherche guidée (Cave & Wolfe, 1990; Wolfe, 1994), les premières modélisations de l'allocation du regard et de l'attention dans une scène visuelle se sont principalement basées sur une mesure de différences entre les caractéristiques visuelles locales par rapport à leur environnement. Ainsi, de manière analogue à la recherche de caractéristiques et à l'effet *pop-out* décrits dans la Section 2.3. du présent chapitre, une zone de l'image dont les caractéristiques visuelles se distinguent particulièrement des zones avoisinantes et du reste de l'image devrait attirer le regard. C'est ainsi que l'on peut définir la saillance *bottom-up* (Itti & Koch, 2000, 2001; Itti, Koch, & Niebur, 1998; Koch & Ullman, 1985).

3.1.1. Cartes de saillances visuelle *bottom-up*

À titre d'exemple de modèles computationnels pour la saillance *bottom-up*, dans le modèle original (Itti et al., 1998), l'image est préalablement filtrée sur neuf échelles spatiales (afin de considérer des objets d'intérêt de différentes tailles apparentes), et pour chaque échelle, les caractéristiques sont ensuite extraites en parallèle sur trois dimensions : couleur, intensité lumineuse et orientation. Ces cartes de caractéristiques sont ainsi calculées en parallèle. Enfin, les contrastes locaux sont extraits (par différence entre les activités sur différentes échelles) et toutes les cartes sont normalisées (compétition spatiale) puis combinées jusqu'à atteindre une carte de saillance unique via une somme pondérée. Pour déterminer la localisation de la prochaine fixation sur la scène visuelle, le modèle implémente un critère de décision de type « *winner takes all* » (la localisation la plus saillante est sélectionnée) doublé d'un mécanisme d'inhibition de retour réduisant la saillance d'une zone déjà sélectionnée pour les n fixations suivantes, permettant d'éviter que la localisation la plus saillante soit fixée systématiquement.

À partir de ce modèle, un grand nombre de modèles ont été proposés. Voir Borji et Itti (2013) pour une revue et Borji, Tavakoli, Sihite et Itti (2013) pour une comparaison standard entre les performances de différents modèles (basée principalement sur la saillance *bottom-up* mais contenant quelques modèles de saillance *top-down*). Ces modèles sont étayés au travers d'un certain nombre d'études expérimentales. Par exemple, pour les tâches d'exploration libre où les participants n'ont pas particulièrement de consigne autre que regarder la scène, les modèles basés sur la saillance *bottom-up* prédisent mieux la répartition des fixations que le hasard ou qu'un modèle gaussien (Borji et al., 2013; Itti & Koch, 2000). Ce dernier modèle simule le biais central de fixation, désignant le fait que lors de l'exploration visuelle d'une image, la distribution des fixations est plus élevée au centre (Tatler, 2007). Par ailleurs, ces modèles prédisent l'ordre des fixations dans une séquence d'exploration (*i.e. scanpath*), avec des performances meilleures que pour un modèle aléatoire uniforme mais globalement plus faible que pour un modèle gaussien (Borji et al., 2013). De plus, dans la tâche d'exploration libre, la saillance aux endroits fixés a tendance à être plus élevée qu'aux endroits non fixés d'une image (Parkhurst, Law, & Niebur, 2002). Dans des tâches d'encodage de scènes visuelles, la saillance *bottom-up* prédit mieux les fixations que différents modèles aléatoires répartissant les fixations suivant une fonction uniforme, reproduisant le biais central de fixation ou bien prenant en compte la localisation des fixations précédentes (Foulsham & Underwood, 2008; Underwood & Foulsham, 2006). Ainsi, il existe une certaine quantité d'études qui pointent le rôle de la saillance *bottom-up* dans l'exploration visuelle, en particulier lors de tâches telles que l'exploration libre ou la mémorisation de scènes.

3.1.2. Limites et reformulation de la saillance *bottom-up*

Cependant, ces résultats sont à contraster avec d'autres études pointant les limites des modèles de saillance *bottom-up*. Il a été pointé que les preuves expérimentales récoltées (exposées précédemment) sont de nature corrélationnelle et par conséquent, ne peuvent pas mener à la conclusion que la saillance *bottom-up* influence directement l'exploration visuelle et est donc un facteur causal dans l'allocation du regard (Henderson, Brockmole, Castelhana, & Mack, 2007; Tatler, 2007). De plus, la saillance a un pouvoir explicatif faible quant à l'allocation du regard dans des scènes visuelles, relativement à d'autres types d'information (Ehinger, Hidalgo-Sotelo, Torralba, & Oliva, 2009; Kanan, Tong, Zhang, & Cottrell, 2009; Rothkopf, Ballard, & Hayhoe, 2016; Tatler & Vincent, 2009; Torralba, Oliva, Castelhana, & Henderson, 2006). Par exemple, Tatler et Vincent (2009) ont montré que la saillance *bottom-up* explique moins bien la répartition des fixations sur des scènes visuelles naturelles en situation d'exploration libre, qu'un modèle implémentant uniquement des biais comportementaux dans le contrôle oculomoteur (distribution d'orientations et d'amplitudes des saccades). De fait, ce modèle est aveugle à l'information visuelle contenue dans l'image et détermine où localiser la prochaine fixation sur la base de tendances systématiques dans l'exécution des saccades en termes d'orientations et amplitudes.

Enfin, un certain nombre d'études ont échoué à mettre en évidence un effet de la saillance sur certaines tâches. Foulsham et Underwood n'ont pas réussi à montrer que la saillance prédit l'ordre des fixations dans des tâches d'encodage de scènes (Foulsham & Underwood, 2008; Underwood, Foulsham, & Humphrey, 2009). Elle ne semble pas non plus jouer de rôle dans l'allocation des fixations lors de tâches de recherche visuelle (Henderson et al., 2007; Henderson, Malcolm, & Schandl, 2009; Underwood & Foulsham, 2006; Underwood, Foulsham, van Loon, Humphreys, & Bloyce, 2006) ni lors de tâches où le regard permet de se diriger dans l'espace et/ou d'accomplir un comportement dirigé vers un but (Ballard & Hayhoe, 2009; Rothkopf et al., 2016).

Récemment, il a été proposé une redéfinition de la notion de saillance pour laquelle les objets ou événements saillants sont vus comme des événements inattendus ou surprenants (Itti & Baldi, 2009; Tatler, Hayhoe, Land, & Ballard, 2011). Ici, la surprise est générée par une déviation statistique dans l'espace et dans le temps. Les statistiques de l'image sont apprises sur une échelle temporelle très courte et permettent de détecter les objets et événements se distinguant de leur environnement (espace) et/ou apparaissant, changeant ou bougeant soudainement (temps). Cette définition permet de rendre compte du fait que le changement dans une scène visuelle, tel que l'apparition, la disparition, le mouvement d'un objet ou un changement soudain de couleur, attire le regard (Brockmole & Henderson, 2005a, 2005b; Matsukura, Brockmole, & Henderson, 2009). Sous cette définition, la saillance ne repose plus exclusivement sur les caractéristiques de l'information extraite de l'environnement, mais implique aussi de l'informations intégrée

a priori, sur laquelle peuvent se fonder des attentes sur la composition de l'environnement. Par exemple, l'information intégrée de manière *top-down* a aussi été modélisée sous forme de cartes de saillance (Borji & Itti, 2013; Wolfe, 1994).

Comme précédemment exposé, la saillance *bottom-up* n'explique que partiellement l'allocation du regard et pour certaines tâches telle que la recherche visuelle, n'a que peu, si ce n'est aucune influence. Comparativement, les connaissances sur la situation dans laquelle se déroule l'observation et sur la tâche à effectuer prédisent plus efficacement l'exploration visuelle.

3.2. Influences *top-down* dans l'exploration visuelle : les connaissances *a priori*

Au cours de sa vie, un individu accumule des connaissances et apprend toutes sortes d'informations sur son environnement. Ces connaissances peuvent ensuite être utilisées pour repérer plus efficacement l'information utile au moment présent, pour accomplir la tâche qu'il ou elle doit effectuer. Ainsi, l'intégration de ces connaissances dites *a priori* dans le processus de sélection de la localisation des fixations du regard constitue une étape importante à la compréhension de la vision lors de tâches de la vie quotidienne. La tâche elle-même, dont l'influence a été montrée précocement (Buswell, 1935; Yarbus, 1967), est aujourd'hui considérée parmi les facteurs les plus déterminants pour l'allocation du regard, notamment par le fait qu'elle définit les objectifs de l'exploration visuelle (Ballard & Hayhoe, 2009).

La variété des tâches dans lesquelles il est possible de s'engager rend compliquée l'élaboration d'un modèle général rendant compte de ces influences. Néanmoins, un cadre général commence à se dessiner. Ainsi, durant les tâches naturelles, différents rôles ont été proposés pour les fixations du regard, parmi lesquels la localisation et l'identification des objets d'intérêt, la prise d'information spécifique permettant de contrôler les mouvements durant la manipulation des objets, et la vérification du résultat de la manipulation (Land & Hayhoe, 2001). L'exploration visuelle permet aussi de repérer et encoder en mémoire de travail l'information pertinente pour la tâche, ainsi que pour naviguer dans l'espace lors de l'accomplissement de cette tâche (Ballard & Hayhoe, 2009; Droll & Eckstein, 2009; Droll, Hayhoe, Triesch, & Sullivan, 2005). Enfin, Land (2009) propose un modèle théorique où le système visuel supporte à la fois le système oculomoteur (*gaze system*, permettant de diriger le regard pour repérer les objets) et le système moteur (permettant de se déplacer dans l'espace et manipuler les objets). Le système des schèmes guide les trois autres systèmes en définissant les séquences d'actions à effectuer (voir par ailleurs Ballard et Hayhoe, 2009, et Rothkopf et Ballard, 2010 pour, respectivement, des discussions théoriques et computationnelles concernant la modélisation de l'influence de la tâche).

Dans ce contexte, la tâche de recherche visuelle a un intérêt particulier puisqu'elle induit un fort

contrôle *top-down* (e.g., comparé à une tâche de mémorisation de scène), contraignant spatialement l'exploration visuelle (Mills, Hollingworth, Van der Stigchel, Hoffman, & Dodd, 2011). De plus, elle constitue à la fois une tâche à part entière mais aussi une étape dans la réalisation de tâches de la vie quotidienne (e.g., rechercher le thé pour le préparer) (Land & Hayhoe, 2001). La présente section se focalise sur la tâche de recherche visuelle et sur l'intégration de connaissances *a priori* pour orienter le regard dans cette tâche. La littérature a principalement traité de l'intégration de deux types de connaissances pour orienter le regard dans le cadre de tâches de recherche visuelle dans des scènes naturelles : le contexte présenté dans la scène visuelle et l'apparence de l'objet à rechercher.

3.2.1. Connaissances relatives au contexte visuel

Le contexte exerce une influence sur l'allocation du regard à différents niveaux. De manière générale, lors de la recherche d'un objet particulier dans l'environnement, les zones de la scène où il y a de fortes probabilités de le trouver sont fixées du regard prioritairement (Castelhano & Heaven, 2010; Henderson et al., 2009; Malcolm & Henderson, 2010; Spotorno, Malcolm, & Tatler, 2014, 2015). Plusieurs formes d'information sont intégrées pour déterminer ces influences contextuelles. Lorsque la scène est connue, l'exploration visuelle peut être guidée par un souvenir de la position de l'objet (Brockmole & Henderson, 2006). Présenter la scène préalablement à la recherche de cible (l'identité de la cible est révélée après cette première présentation de la scène) permet à la fois de définir les endroits où la cible pourrait se trouver (traces du contexte en mémoire) mais aussi de retrouver où celle-ci se situait (ancrage de la cible dans l'espace) (Castelhano & Henderson, 2007; Hollingworth, 2009). Ces deux formes de souvenirs de la scène sont discernables selon que la cible soit présente ou pas lors de la première présentation de la scène. La connaissance préalable de la scène n'est pas forcément nécessaire ; la récupération en mémoire d'une connaissance explicite sur sa localisation spatiale, par exemple via une consigne peut indiquer une partie de la scène contenant la cible (e.g., « au-dessus de l'immeuble ») (Zelinsky & Schmidt, 2009).

Lorsque la scène est inconnue, des connaissances générales sur les objets la composant peuvent permettre de localiser la cible plus facilement. Ces connaissances sont construites par apprentissage lors de situations d'observation vécues antérieurement, et des régularités entre ces différentes instances d'observation passées sont extraites et transférées aux situations d'observation présentes (Brockmole & Vö, 2010; Brooks, Rasmussen, & Hollingworth, 2010; Vö & Wolfe, 2015). Parmi les connaissances générales pouvant être utilisées, celles des lois de la physique permettent d'anticiper la position des objets pour les rechercher dans des zones adéquates de la scène. Une vache aura peu de chances d'être localisée dans le ciel (du moins, sans parachute) rendant l'exploration visuelle de cette partie de la scène sous-optimale (Spotorno et al., 2014). Cependant de légères infractions locales aux lois de la physique (objets flottant à quelques dizaines de centimètres au-dessus d'une surface) n'attirent pas particulièrement le regard ni ne perturbent la recherche visuelle (Vö & Henderson, 2011).

Ces derniers auteurs ont aussi étudié si la congruence sémantique de la cible avec le contexte facilite la recherche et l'exploration visuelle. Les résultats obtenus indiquent que c'est le cas, un objet incongruent avec la catégorie de la scène visuelle (*e.g.*, imprimante dans une cuisine, placée sur les plaques de cuisson) étant plus difficile à repérer et donc, fixé du regard plus tardivement. Néanmoins, Castelhana & Heaven (2011) ont contrôlé la cohérence de la localisation de la cible dans des scènes visuelles sémantiquement congruentes ou incongruentes (*e.g.*, un mug dans une chambre à coucher –scène sémantiquement incongruente– placé sur une table de chevet *vs.* par terre). Elles ne retrouvent pas d'effet de congruence sémantique scène-cible dans ce cas. Par contraste, l'exploration visuelle est facilitée pour la recherche d'une cible placée à un endroit cohérent comparé à un endroit incohérent, que la scène soit sémantiquement congruente avec la cible ou non. Cela peut souligner que des connaissances sur la fonction des objets permet de les localiser plus facilement dans l'espace.

La relation entre les objets dans la scène permet aussi une telle facilitation ; une cible liée spatialement à d'autres objets (*e.g.*, une cheminée relativement à une maison) est plus facilement repérable lorsque placée de manière cohérente par rapport à ces objets (cheminée sur le toit), les objets pouvant indiquer sa position (Brockmole & Vö, 2010; Eckstein, Drescher, & Shimozaki, 2006; Malcolm & Shomstein, 2015; Wu, Wick, & Pomplun, 2014). Il a été argumenté que si le contexte global représenté par la scène visuelle permet de déterminer de larges zones où la cible peut se trouver, les objets contenus dans la scène permettent une détermination plus précise de la position de la cible (Brooks et al., 2010; Pereira & Castelhana, 2014). Voir par ailleurs Vö et Wolfe (2015) ainsi que Wu, Wick et Pomplun (2014) pour des revues d'articles.

Enfin, peu de modèles intégrant une forme d'influence contextuelle ont été proposés. Les principales tentatives (Ehinger et al., 2009; Kanan et al., 2009; Torralba et al., 2006) implémentent dans un cadre Bayésien, une carte de saillance contextuelle *top-down* sur la base de la relation apprise entre cibles et *gist* de l'image, ne prenant donc pas en compte le contenu précis de la scène (objets). La rigidité d'une telle implémentation de l'influence contextuelle a cependant été critiquée sur le fait qu'elle ne représente qu'un des types de contraintes exercés sur la position des objets (Zelinsky & Schmidt, 2009), et que la présente section a tenté de passer en revue. Notons cependant, un autre modèle (Eckstein et al., 2006) qui implémente l'influence contextuelle sur la base de la concomitance des localisations de deux objets (*e.g.*, cheminée sur le toit d'une maison), toujours dans un cadre Bayésien, mais dont le but est de prédire la destination de la première saccade plutôt que la séquence de fixation complète effectuée lors de la recherche visuelle.

3.2.2. Connaissances sur l'apparence de la cible

Comme exposé dans la partie précédente sur l'attention visuelle (Section 2.), la connaissance *a priori* des caractéristiques de la cible, et plus généralement de l'apparence de la cible, a été extensivement prise en compte dans la construction de modèles pour l'allocation de l'attention visuelle (Beck & Kastner, 2009; Desimone & Duncan, 1995; Wolfe, 1994). C'est effectivement une forme d'information qui ne facilite pas seulement la recherche visuelle, elle en définit l'objectif (Rao, Zelinsky, Hayhoe, & Ballard, 2002). Dans les tâches de recherche visuelle dans des scènes naturelles, ces connaissances sur l'apparence de l'objet cible constituent une représentation visuelle de la cible (*target template*) en mémoire de travail (Hout & Goldinger, 2015). Habituellement, la représentation de la cible est manipulée en indiquant la cible à rechercher via la présentation d'une image plus ou moins précise de celle-ci (*vs.* présenter son nom écrit) préalablement à la scène visuelle, ou en construisant cette représentation via l'apprentissage de l'association nom-image des objets présentés (*e.g.*, espèces de poissons tropicaux dans Bravo & Farid, 2009).

Ainsi, dans des scènes visuelles naturelles, la présentation préalable d'une image de la cible (l'objet à trouver) facilite la recherche visuelle subséquente par comparaison avec la présentation préalable de son nom, accélérant autant sa détection dans la scène que son identification (Castelhano & Heaven, 2010; Malcolm & Henderson, 2009, 2010, Spotorno et al., 2014, 2015). De plus, les résultats de Spotorno et al. (2014) suggèrent qu'une représentation visuelle précise de la cible favorise sa détection extra-fovéale dès l'initiation de la recherche, augmentant notamment la probabilité que la première saccade soit faite dans sa direction. Même lors d'une recherche de patterns visuels dans du bruit visuel non structuré, le regard est attiré par les zones ressemblant au pattern cible (Rajashekar, Bovik, & Cormack, 2006).

Ces représentations doivent être suffisamment spécifiques pour permettre la discrimination de la cible avec les distracteurs, tout en tolérant la variabilité des conditions de présentation de la cible dans l'environnement (échelle, masquage, orientation). Castelhana, Pollatsek et Cave (2008) ont observé que lorsqu'on active une représentation très vague de la cible préalablement à l'apparition de la scène (via son nom écrit), une cible typique de sa catégorie est reconnue plus facilement qu'une cible atypique, la cible typique étant plus ressemblante à la représentation vague activée par le nom de l'objet-cible. Cet effet disparaît lorsqu'une représentation visuelle plus précise est activée préalablement à la recherche visuelle via la présentation d'une image de la cible. De même, il a été montré qu'une représentation imprécise de la cible (*e.g.*, même objet, détail non diagnostique modifié) dégrade les performances par rapport à une représentation exacte (image exacte de la cible), bien qu'elle les facilite par rapport à une représentation inexacte (*e.g.*, autre exemplaire de la même catégorie que la cible) (Hout & Goldinger, 2015). Cette dégradation affecte autant l'exploration visuelle lors de la recherche visuelle que le temps de décision pour l'identification de la cible. Par ailleurs, Reeder et Peelen (2013), ont comparé d'une part, la présentation préalable de silhouettes noires des cibles à rechercher (piéton ou voitures) dans différentes orientations et d'autre part, de stimuli contenant seulement leur texture et couleur. Ils ont ainsi montré que ce type d'effet facilitateur, observé uniquement avec les silhouettes, était bien dû à la construction d'une représentation de la cible, et non au simple amorçage de ses caractéristiques visuelles. De plus, ils soulignent l'indépendance de ces représentations par rapport au point de vue.

La représentation de la cible peut aussi être construite par apprentissage de la relation image-nom de l'objet préalablement à la tâche de recherche visuelle (*e.g.*, poissons tropicaux de différentes espèces). Elle peut être activée en mémoire de travail en nommant simplement l'objet-cible à rechercher. Cette méthode permet d'éviter un éventuel amorçage visuel *bottom-up* de la cible. C'est la méthode employée par Bravo et Farid (2009, 2012). Ils ont montré que la représentation ainsi construite est spécifique à l'objet appris, sans généralisation à d'autres exemplaires de la même catégorie (poissons tropicaux de la même espèce) mais relativement insensible à des différences d'orientation ou d'échelle (Bravo & Farid, 2009). Enfin, la façon dont les associations nom-image sont apprises permettent la construction de représentations plus ou moins spécifiques de la cible, permettant une généralisation à des exemplaires de la même catégorie (apprentissage par blocs incluant tous les exemplaires d'une catégorie, *i.e.* tous les poissons d'une espèce) ou ne permettant pas cette généralisation (apprentissage par blocs incluant uniquement les images d'un exemplaire, *i.e.* les différentes images d'un poisson) (Bravo & Farid, 2012).

Un certain nombre de modèles ont implémenté les représentations de la cible comme source d'influence sur l'allocation du regard dans des tâches de recherche visuelle. En particulier, le modèle SUN (Kanan et al., 2009) implémente dans une perspective Bayésienne, l'influence des représentations de la

cible comme la probabilité conditionnelle que les caractéristiques en un point soient celles observées sachant que ce point appartient à la cible. Il parvient à des performances similaires au Contextual Guidance Model (Torralba et al., 2006) lors de la prédiction des aires communément fixées par des participants humains dans une tâche de recherche visuelle. D'autres modèles notables incluent Rao et al. (2002), qui implémente une construction progressive de cartes de saillance *top-down* suivant un décours *coarse-to-fine* et représentant cette saillance comme la similarité entre les caractéristiques observées en chaque point et à chaque instant et les caractéristiques d'une représentation stockée en mémoire. Dans le même ordre d'idée, Zelinsky (2008) propose un modèle basé sur la similarité entre l'apparence de la cible et les caractéristiques locales de l'image, excluant toute saillance *bottom-up* mais simulant notamment vision centrale et périphérique via la modélisation d'une rétine artificielle dégradant l'information captée selon l'excentricité rétinienne. Enfin, Zelinsky, Zhang, Yu, Chen et Samaras (2006) explorent les poids relatifs entre saillance *bottom-up* et *top-down* (représentation de la cible) dans la recherche visuelle et obtiennent les meilleures prédictions pour le modèle excluant toute saillance *bottom-up*. Les résultats de ces modèles soulignent à quel point la représentation de la cible est importante dans la recherche visuelle (notamment comparé à la saillance *bottom-up*), dans la mesure où elle définit directement l'objectif de cette tâche.

3.3. Résumé et conclusion

Si les mouvements des yeux peuvent être vus comme dirigés par l'attention autant que permettant la focalisation de celle-ci sur des stimuli de l'environnement, ils sont devenus un objet d'étude à part entière dans les tâches intégrant des scènes visuelles écologiques, en particulier pour la recherche visuelle d'objets. L'approche principalement employée pour modéliser l'allocation du regard dans ce type de tâche est directement liée aux modèles attentionnels de Treisman et Gelade (1980) ainsi que Wolfe (1994), et s'est développée en parallèle. Cette approche repose sur la notion de saillance désignant au sens large dans quelle mesure chaque zone se démarque de ce qui l'entoure, que ce soit par ses caractéristiques visuelles (saillance *bottom-up*) ou parce qu'elle a une pertinence particulière pour la tâche en cours (saillance *top-down*). Si la saillance *bottom-up* permet de prédire l'allocation du regard des participants humains au-dessus du hasard dans un certain nombre de tâches, il a été montré que les influences *top-down* avaient un pouvoir explicatif plus fort, en particulier pour des tâches dont les objectifs contraignent fortement l'observation, telles que la recherche visuelle. Des influences *top-down*, deux sources d'information ont été principalement étudiées : les relations entre cible et contexte et les connaissances sur l'apparence de la cible, formalisées comme représentations de la cible en mémoire de travail. Cependant, Kowler (2011) note que si l'orientation du regard se base sur la combinaison de cartes de saillance *bottom-up*, alors l'objectif du

Chapitre 1 – Traitement de l'information visuelle et perception de scènes visuelles

système oculomoteur est de fixer des régions se distinguant déjà du reste de la scène, plutôt que de fixer à chaque saccade, les régions permettant de maximiser l'information pertinente collectée. Par ailleurs, Malcolm, Nuthmann et Schyns (2014) mettent en avant la relative exclusion de la question de la contribution des mouvements des yeux à la tâche elle-même, proposant de se focaliser sur l'accumulation d'information permettant l'accomplissement de celle-ci.

Le Chapitre 1 a passé en revue, de manière non exhaustive, comment de l'intégration progressive d'information visuelle riche et complexe via l'organisation neuro-physiologique du système visuel et les mécanismes de sélection attentionnelle et oculomotrice, émerge un processus d'accumulation de l'information pertinente pour la tâche à effectuer. Ces tâches impliquant l'exploration visuelle, notamment la recherche visuelle, peuvent ainsi être vues comme des tâches de prise de décision perceptuelle dans lesquelles la fixation du regard à un endroit de la scène et la relocalisation du regard par des mouvements des yeux permettent d'accumuler de l'information pertinente. Ce chapitre a permis de mettre en avant comment les caractéristiques visuelles de l'information sensorielle et les attentes sur l'environnement visuel influencent cette accumulation d'information, mais les travaux présentés ne traitent pas de leurs influences consécutives sur les décisions elles-mêmes (permettant de mener la tâche à bien).

Chapitre 2

Prise de décision et collecte d'information pour la perception visuelle : Méthode expérimentale intégrant des méthodes de *mouse-tracking* et d'oculométrie

1. La prise de décision perceptuelle

Si la question de la contribution de l'exploration visuelle à la tâche a été peu adressée pour ce qui concerne les tâches de recherche visuelle, la prise de décision, en revanche, a bénéficié d'un intérêt sans cesse renouvelé (Gold & Shadlen, 2007; Shadlen & Kiani, 2013). Plus spécifiquement, la prise de décision perceptuelle désigne le « processus par lequel l'information collectée via les systèmes sensoriels est combinée et utilisée pour influencer la manière dont nous nous comportons dans l'environnement » (Heekeren, Marrett, & Ungerleider, 2008, p. 467). Elle repose sur l'intégration d'information sensorielle *bottom-up* avec de l'information *a priori* (connaissances, attentes) en vue de sélectionner une réponse parmi un certain nombre d'alternatives (Summerfield & de Lange, 2014). Ainsi, elle englobe un large panel de fonctions cognitives et en cela, est applicable et/ou adaptable à un grand nombre de tâches de laboratoire –telles que la détection de stimuli ou la reconnaissance et classification d'objets (Kietzmann, Geuter, & König, 2011; Smith & Ratcliff, 2004; Turner, Van Zandt, & Brown, 2011; Voss, Nagler, & Lerche, 2013)– comme de tâches se rapprochant de situations de la vie quotidienne –telles que les choix de préférences et les décisions sociales (Krajbich, Hare, Bartling, Morishima, & Fehr, 2015; Krajbich, Lu, Camerer, & Rangel, 2012; Usher, Tsetsos, Yu, & Lagnado, 2013). De manière générale, elle est formalisée comme un choix entre un nombre limité de réponses alternatives, basé sur la quantité d'évidence disponible en faveur de chacune d'entre elles

Le terme *évidence* est notre choix de traduction pour le terme anglais *evidence*. Il désigne l'information pertinente pour la prise de décision c'est à dire, l'information indiquant dans quelle mesure une hypothèse/réponse est vraie ou a de la valeur (estimée ou perçue) par rapport à son/ses alternative(s). Nous préférons cet anglicisme aux traductions telles que *support*, *preuve* ou *indice*. Le caractère indénombrable de ce terme nous semble correspondre plus directement à l'idée d'une accumulation progressive en faveur de différentes alternatives. Par ailleurs, *support* peut poser des problèmes lors de la traduction d'expressions courantes telles qu'*evidence accumulation*. Enfin, pour désigner la quantité d'évidence présente dans l'information sensorielle à intégrer (*e.g.*, rapport signal/bruit, contraste du stimulus, différence entre les stimuli à discriminer), nous emploierons le terme de *force de l'évidence*,

traduction littérale du terme *evidence strength* (ou *signal strength*, *stimulus strength...*), fréquemment employé dans la littérature (J. Palmer, Huk, & Shadlen, 2005).

La flexibilité des modèles issus du champ d'étude de la prise de décision perceptuelle permet de rechercher des mécanismes communs à diverses fonctions cognitives en jeu dans diverses tâches. Ainsi, cette section vise à proposer une vision globale bien que non exhaustive des modèles de prise de décision perceptuelle classiques, depuis la Théorie de la Détection du Signal (TDS) –approche précoce modélisant le lien entre information sensorielle et choix entre deux alternatives– aux modèles d'échantillonnage séquentiel (*sequential sampling models*) –extensions de la TDS permettant de modéliser à la fois le choix et le temps de décision, mettant ainsi en avant le caractère dynamique de la prise de décision. Via cette revue, nous mettrons en avant à la fois l'intégration d'attentes et les mécanismes permettant une accumulation active d'évidence en faveur des différentes réponses possibles dans le cadre de ces modèles.

Il est à noter que si la théorie de la détection du signal, puis les modèles d'échantillonnage séquentiel ont joué un rôle central dans l'étude de la prise de décision perceptuelle, il existe aussi d'autres types de modèles plus récents dans les domaines de la prise de décision perceptuelle comme de la cognition visuelle, notamment le codage prédictif (*predictive coding*, Feldman & Friston, 2010; Rao & Ballard, 1999; Summerfield & de Lange, 2014). Ces modèles intègrent notamment l'idée que la perception active puisse être guidée par des prédictions générées à propos de l'environnement externe. Ils visent à réduire l'erreur de prédiction issue de la comparaison entre ces prédictions et l'information sensorielle captée via une adaptation du modèle interne générant les prédictions (Catenacci Volpi, Quinon, & Pezzulo, 2014; Feldman & Friston, 2010; Friston, Adams, Perrinet, & Breakspear, 2012; Hohwy, 2012; Lepora & Pezzulo, 2015; Quinon, Catenacci Volpi, Barca, & Pezzulo, 2014). Pourtant au premier plan ces dernières années, nous avons exclu ces modèles à cause de leur complexité et de leur emploi de formalismes mathématiques variés et parfois compliqués. Ces modèles s'inscrivent dans un champ d'étude voisin des modèles échantillonnage séquentiel (Summerfield & de Lange, 2014) et dans le cadre du présent travail, il nous a semblé suffisant de ne décrire avec plus de détails que la TDS et les modèles d'échantillonnage séquentiel. Cependant, nous n'ignorons pas les apports théoriques de ces approches voisines tout au long de ce manuscrit et tentons d'y référer à chaque fois qu'il est nécessaire de le faire.

1.1. Modèles de prise de décision perceptuelle

1.1.1. *Théorie de la détection du signal*

La TDS compte parmi les théories les plus influentes, à la fois dans le domaine de la perception visuelle et de la psychologie. Un des plus important acquis qu'elle a apporté à ces champs d'étude est l'idée

que toute réponse psychophysique, même la plus simple telle que déterminer si un signal est présent dans du bruit, repose sur un processus de prise de décision (Ashby & Soto, 2015; Shadlen & Kiani, 2013). En ce sens, elle permet de rendre compte de comment l'observation d'un stimulus mène à une réponse unique (Shadlen & Kiani, 2013).

Les deux réponses alternatives sont représentées comme deux variables aléatoires gaussiennes et de même variance dans le cas du modèle classique, caractérisées sur une dimension diagnostique (la variable décisionnelle, Shadlen & Kiani, 2013), qui représente l'évidence en faveur de l'une ou l'autre des réponses. Ainsi, lorsque la tâche est de déterminer la présence d'un stimulus dans du bruit (tâche de détection), cette dimension désigne la force du stimulus (*e.g.*, rapport signal/bruit). Lorsque la tâche est de distinguer deux stimuli ou catégoriser des stimuli (tâche de catégorisation ou reconnaissance), elle intègre les caractéristiques distinguant les deux stimuli ou catégories (Macmillan & Creelman, 2005). La TDS consiste à déterminer un critère de séparation entre les deux réponses sur cette dimension. Si le stimulus observé a une valeur supérieure à ce critère, la décision est prise en faveur d'une réponse, si elle est inférieure, la décision est prise en faveur de l'autre réponse. La distance entre les distributions des réponses permet donc d'estimer dans quelle mesure les deux réponses peuvent être distinguées sur la base de l'évidence sensorielle disponible, et peut être estimée via la sensibilité –souvent résumée par la statistique d' – (Macmillan & Creelman, 2005; Summerfield & de Lange, 2014; Turner et al., 2011). Par exemple, deux catégories d'animaux très distinctes telles que des chats et des requins verront les distributions relatives à leurs réponses respectives distantes sur la dimension diagnostique, puisque de nombreux indices visuels (intégrés sur cette dimension) permettent de discriminer les exemplaires de ces catégories. À l'inverse, cette distance sera plus faible pour des dauphins vs. des requins.

Les attentes peuvent être modélisées comme un biais de réponse décalant le positionnement du critère vers l'une ou l'autre des réponses, facilitant la sélection de la réponse alternative (Summerfield & de Lange, 2014). Si le contexte dépeint un environnement terrestre, la présence d'un mammifère est *a priori* plus probable, biaisant le critère vers la réponse *poisson*, ce faisant, réduisant l'intervalle pour lequel la réponse *poisson* sera choisie et augmentant celui pour lequel la réponse *mammifère* sera choisie.

Cette théorie a donné lieu à de nombreuses extensions et généralisations, permettant notamment la discrimination sur plusieurs dimensions (Ashby & Soto, 2015) ou l'adaptation des paramètres du modèle au cours de la présentation d'un set de stimuli, permettant la construction des représentations au fur et à mesure des essais, ainsi que l'adaptation du biais de réponse (Turner et al., 2011). Si elle permet d'analyser la performance dans les tâches de prise de décision perceptuelle distinguant l'influence de variations de la représentation des stimuli vs. d'un biais de réponse sur l'exactitude des réponses, ce modèle ne rend pas compte du temps nécessaire à la complétion de la décision (J. Palmer et al., 2005; Shadlen & Kiani, 2013;

Turner et al., 2011).

1.1.2. Modèles d'échantillonnage séquentiel

Pour rendre compte à la fois du choix et du temps de réponse (lié au temps de décision), les modèles d'échantillonnage séquentiel ont été proposés. Cette catégorie de modèles peut être vue comme une extension de la TDS, ajoutant à celle-ci un aspect dynamique (Gold & Shadlen, 2007; Pleskac & Busemeyer, 2010; Smith & Ratcliff, 2004; Summerfield & de Lange, 2014). Ainsi, la TDS base la décision sur de l'information accessible dans son ensemble là où dans les modèles d'échantillonnage séquentiel, la même information est échantillonnée et intégrée progressivement, accumulant ainsi de l'évidence en faveur de chaque réponse alternative. Le caractère dynamique de ces modèles permet de filtrer le bruit associé à l'information captée au fur et à mesure que le temps passe (Kelly & O'Connell, 2013; Smith & Ratcliff, 2004). L'évidence est accumulée jusqu'à atteindre le seuil correspondant à l'une des réponses, scellant la décision en sa faveur. De nombreux modèles d'échantillonnage séquentiel ont été proposés. On peut distinguer deux grandes classes de modèles (Palmeri, Schall, & Logan, 2015; Smith & Ratcliff, 2004), la Figure 2 présente un exemple de modèle pour chacune des deux classes.

Dans la première (Figure 2.B.), les modèles à accumulateurs, l'évidence est accumulée à chaque unité de temps dans des accumulateurs distincts pour chaque alternative avec un seuil absolu. Le processus de décision est modélisé comme une course où l'alternative choisie est celle dont le total d'évidence accumulée atteint le seuil en premier. Certaines instances de cette classe de modèles (telles qu'employé pour la Figure 2.B.) implémentent un mécanisme d'inhibition mutuelle où l'évidence accumulée par une alternative réduit le total des autres, permettant la compétition entre réponses (Usher & McClelland, 2001).

Pour la seconde classe de modèles, un total unique d'évidence est accumulé, l'information accumulée en faveur d'une alternative étant de l'information accumulée contre les autres et le seuil de décision est donc relatif. Les modèles de marche aléatoire (*random-walk models*) et les modèles de diffusion (*drift-diffusion models*) sont les deux principaux types de modèles de cette classe (Smith & Ratcliff, 2004). Nous nous focaliserons sur les modèles de diffusion (Figure 2.C.), ceux-ci étant les plus répandus. La version classique de ce type de modèles décrit un processus de décision opposant deux alternatives où la quantité d'évidence dévie vers l'une ou l'autre des réponses à chaque unité de temps jusqu'au seuil lui correspondant, scellant de ce fait la décision (Ratcliff & McKoon, 2008; Ratcliff & Rouder, 1998; Ratcliff, Smith, Brown, & McKoon, 2016). Le point de départ étant fixé à 0, une quantité d'évidence positive correspond à une déviation vers le seuil de réponse A, et une quantité d'évidence négative correspond à une déviation vers le seuil de réponse B.

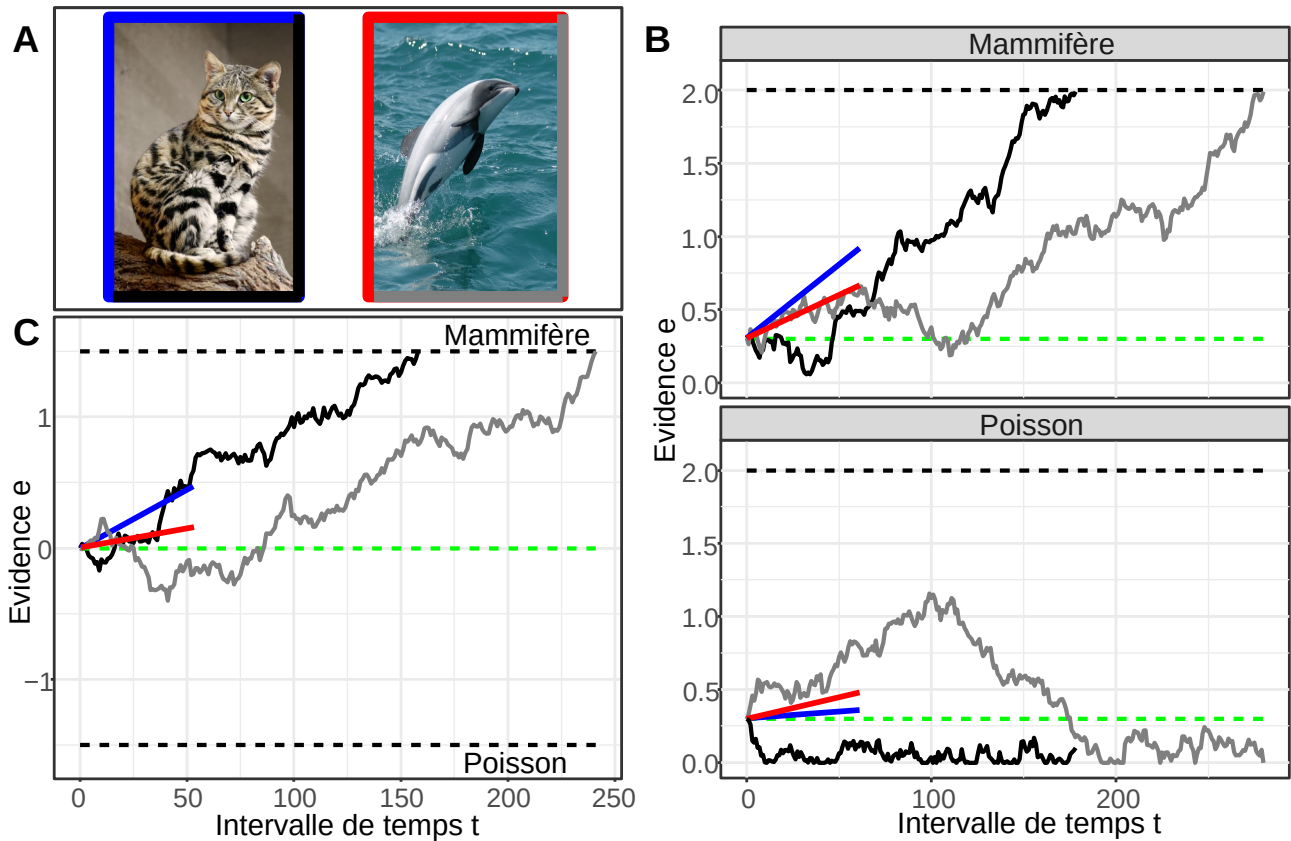


Figure 2 : Illustration de modèles à échantillonnage séquentiel dans une tâche de catégorisation de mammifères ou de poissons. La situation expérimentale est fictive et les données présentées à titre d'exemple sont simulées. **A.** Exemples de mammifères non-ambigu (chat) et ambigu (dauphin) présentés. Les cadres colorés indiquent les couleurs leur correspondant dans les graphiques. **B.** Illustration du modèle à accumulateurs. Évidence accumulée par unité de temps en faveur de chaque réponse lorsqu'un chat (courbe noire) ou un dauphin (courbe grise) est présenté. Les parties supérieure et inférieure correspondent aux accumulateurs d'évidence pour chaque réponse. Les droites pointillées noires et vertes représentent les seuils de décision et niveaux de référence, respectivement. Les segments bleus (chat) et rouges (dauphin) représentent les taux de déviation de chaque réponses, employés pour simuler les essais. Enfin, le modèle employé implémente un mécanisme d'inhibition mutuelle (une fraction de l'activation d'une réponse est soustraite à l'évidence accumulée pour l'autre réponse au temps suivant). **C.** Illustration du modèle de diffusion. Déviation de l'accumulation d'évidence en faveur d'une réponse ou d'une autre par unité de temps. Dans ce modèle, l'évidence est modélisée comme la différence d'évidence sensorielle en faveur de chaque réponse dans le stimulus. Pour refléter cela, les taux de diffusion employés dans cette simulation correspondent à la soustraction des taux de déviation de chaque réponse employés dans le modèle à accumulateurs (mais les simulations sont indépendantes, *i.e.* les valeurs de bruit générées sont différentes). Les droites pointillées noires et verte correspondent aux seuils de décision des réponses et au point de départ de l'accumulation d'évidence, respectivement.

1.1.3. Accumulation d'évidence sensorielle et attentes

Dans les modèles d'échantillonnage séquentiel, une certaine quantité d'évidence est donc accumulée à chaque unité de temps, à partir de l'information visuelle contenue dans le stimulus. La quantité d'évidence sensorielle accumulée à chaque unité de temps (force) peut varier, le stimulus pouvant être bruité ou ambigu. Cela est modélisé via le taux de déviation (*drift-rate*, segments bleus et rouges sur les figures 2.B. et 2.C.) qui, selon les implémentations, pondère la quantité d'évidence additionnée au total à chaque unité de temps (Summerfield & de Lange, 2014), ou représente le taux moyen d'évidence accumulé à chaque unité de temps sur la durée du processus d'accumulation (Forstmann, Ratcliff, & Wagenmakers, 2016; Palmeri et al., 2015; Ratcliff & Rouder, 1998). Comme la quantité d'évidence accumulée à chaque unité de temps est bruitée (de variance constante dans les modèles traditionnels), une réduction du taux de diffusion entraîne, en plus d'une durée de décision plus longue, un taux d'exactitude des réponses plus faible (J. Palmer et al., 2005). En ce sens, un lien entre activité neurophysiologique (*centroparietal positivity* mesurée par EEG) et accumulation d'évidence sensorielle reflétant la force de l'évidence a été établi, validant cet aspect des modèles d'échantillonnage séquentiel (Kelly & O'Connell, 2013).

Par ailleurs, ces modèles peuvent rendre compte des effets des attentes de différentes manières : via une contribution à la détermination du taux de déviation vers l'alternative attendue (réponse attendue) ou via l'addition d'évidence préalablement à la présentation du stimulus, équivalente à un déplacement du point de départ vers le seuil de l'alternative attendue pour les modèles de diffusion ou une augmentation de l'activation de base (niveau de référence) de l'alternative attendue pour les modèles à accumulateurs. La principale différence conceptuelle entre ces deux implémentations est que la seconde représente un biais pré-stimulus alors que la première représente un biais dynamique appliqué durant le processus de prise de décision. Dans une récente revue des principales études en neurophysiologie sur cette question, Summerfield et de Lange (2014) relèvent à la fois l'augmentation de l'activité corticale pré-stimulus en faveur de l'alternative attendue (révélée via enregistrements neuronaux chez les primates non-humains, en MEG ou IRMf) et les mécanismes par lesquels les attentes *top-down* peuvent biaiser dynamiquement la déviation vers une réponse ou une autre, et moduler l'accumulation d'évidence. Par exemple, il a été rapporté que le biais du taux de déviation généré par les attentes augmente à mesure que la décision dure, diminuant ainsi progressivement le poids relatif de l'évidence sensorielle dans le taux de déviation, lorsque la force de l'évidence sensorielle est faible (Hanks, Mazurek, Kiani, Hopp, & Shadlen, 2011). De plus, l'effet de suppression par les attentes (*expectation suppression*), caractérisé par une activité réduite au niveau des aires sensorielles lorsque le stimulus est attendu (mis en évidence par IRMf et EEG), va dans le sens de cette idée, le poids des attentes réduisant la nécessité de collecter de l'information sensorielle sur le stimulus pour prendre la décision (Summerfield & de Lange, 2014; Summerfield & Egner, 2009). Ainsi, les attentes

peuvent biaiser les traitements neuronaux relatifs à la décision à la fois avant et durant la présentation du stimulus. En résumé, les modèles d'échantillonnage séquentiel implémentent cette information comme s'additionnant à l'évidence sensorielle collectée plus tard en observant le stimulus, ou bien comme modulant la collecte d'évidence sensorielle elle-même.

Nous avons décrit les mécanismes de base des modèles d'échantillonnage séquentiel et discuté brièvement de comment l'accumulation d'évidence sensorielle et l'intégration d'attentes étaient implémentées dans ces modèles. Ces mécanismes rendent compte correctement des distributions de temps de réponse et taux d'exactitude dans les expériences comportementales (J. Palmer et al., 2005; Ratcliff & McKoon, 2008), ainsi que de mesures neurophysiologiques (EEG, MEG, trains d'impulsions neuronales) (Smith & Ratcliff, 2004). De plus, l'ajout de certains paramètres tels que des mécanismes de fuite (*leak*, oubli de l'information précoce), de palier (*gating*, inhibition de l'accumulation d'évidence jusqu'à ce que la quantité d'évidence atteigne certain seuil) ou de signaux d'urgence (*e.g.*, seuils diminuant avec le temps) peut permettre de meilleures prédictions (notamment de l'activité neurale) (Ditterich, 2006; Purcell et al., 2010; Usher & McClelland, 2001). Cependant, le processus de prise de décision résulte également de comportements actifs permettant de sélectionner l'information pertinente à la tâche (devenant alors de l'évidence) ou de moduler le processus d'accumulation d'évidence.

1.2. Mécanismes actifs d'accumulation d'information

Les modèles d'échantillonnage séquentiel reposent sur l'idée que de l'information est accumulée à mesure que le temps passe, mais dans la majorité des implémentations, cette accumulation opère passivement, alors que le regard est fixé sur le stimulus. En effet, la majeure partie des paradigmes utilisés pour étudier la prise de décision perceptuelle dans le domaine de la vision reposent sur l'observation d'un stimulus suffisamment petit pour qu'il ne soit pas nécessaire d'effectuer des saccades oculaires (Gold & Shadlen, 2007). Or des mécanismes comportementaux de vision active contribuent à cette collecte d'évidence, dans la mesure où ils permettent de sélectionner l'information pertinente pour la discrimination des réponses alternatives dans la tâche de prise de décision. Parmi les mécanismes de la perception active on peut en noter deux modulant la collecte d'information. Premièrement, comme mis en avant dans la Section 1. du Chapitre 1, différentes parties du signal visuel sont traitées selon un déroulement temporel différent. Ainsi, la composante temporelle est capitale et passer plus de temps à regarder le stimulus permet d'extraire progressivement plus d'évidence, et d'effectuer un jugement moins incertain. À l'inverse, le temps est parfois limité pour répondre, menant à un compromis où la durée de l'accumulation d'évidence est modulée pour arriver à un taux d'exactitude correct malgré la contrainte temporelle. Cet aspect de la

perception active est connu sous le terme de compromis vitesse-exactitude (CVE, *speed-accuracy trade-off*, SAT) (Heitz, 2014; Heitz & Schall, 2012). Deuxièmement, comme discuté dans les Sections 2. et 3. du Chapitre 1, l'attention visuelle et sa manifestation ouverte qu'est le comportement oculomoteur permettent de collecter prioritairement l'information visuelle pertinente pour la tâche (Carrasco, 2011; Kowler, 2011), et donc pour la décision. Bien entendu, ces mécanismes ne sont pas indépendants l'un de l'autre, passer plus de temps à collecter de l'information permettant d'effectuer plus de fixations des yeux, ou de plus longues fixations. Cependant nous allons présenter tour à tour ces deux mécanismes dans la présente section.

1.2.1. Compromis vitesse-exactitude

Lorsque les capacités de traitement sont limitées et que l'environnement est suffisamment complexe (*e.g.*, scènes naturelles), l'information visuelle doit être échantillonnée dans le temps. Certaines situations nécessitent une prise de décision rapide (*e.g.*, pour réagir à la présence d'un danger potentiel) alors que dans d'autres situations, le temps disponible n'est pas (ou est peu) limité, permettant de prendre la décision la plus exacte et/ou précise possible (*e.g.*, travail minutieux sans contrainte temporelle). Dans le premier cas, le résultat de la prise de décision aura moins de chance d'être exact ou sera moins précis que dans le second cas (Hegdé, 2008; Heitz, 2014). Le temps passé à accumuler de l'information peut donc être modulé activement pour s'adapter à la situation ou à la tâche (Bogacz, Wagenmakers, Forstmann, & Nieuwenhuis, 2010). Ce phénomène, le CVE, est un effet classique en psychologie et est souvent vu comme un marqueur du processus de prise de décision (Heitz, 2014). Un certain nombre de méthodes ont été élaborées pour manipuler le CVE vers la vitesse des réponses ou leur exactitude –*e.g.*, en manipulant les consignes, les récompenses ou imposant des latences plancher et/ou plafond en dessous et au-dessus desquelles il n'est pas possible de répondre (Wickelgren, 1977).

Enfin, les modèles d'échantillonnage séquentiel rendent compte du CVE par la manipulation de la distance entre point de départ et seuil de décision pour toutes les réponses possibles (Bogacz et al., 2010; Heitz, 2014). Plus cette distance est faible, plus le temps de réponse est bas (la quantité d'évidence à accumuler étant faible) mais plus forte est la probabilité que le bruit lié à l'accumulation d'évidence provoque un dépassement du seuil de la réponse erronée. À l'inverse, une distance élevée nécessite l'accumulation de plus d'évidence, associée à un risque faible que le bruit provoque un dépassement du seuil de la réponse erronée. La Figure 3 illustre ces modulations de seuils/points de départ. Pour les modèles à accumulateurs (Figure 3.A.), l'adaptation du CVE est implémentée via l'augmentation ou la diminution des seuils de décision ou des points de départ des accumulateurs (ou niveaux de référence, *baselines*). Pour les modèles de diffusion (Figure 3.B.), elle est implémentée via le rapprochement ou

l'éloignement des deux seuils de décision par rapport au point de départ. Au niveau neurophysiologique, un biais vers la vitesse de décision ou son exactitude se traduit par une modulation de l'activité des ganglions de la base, associée à l'augmentation ou la diminution (respectivement) du niveau d'activation de référence de structures neuronales liées à la décision et à la préparation motrice (*e.g.*, au niveau des aires prémotrices, motrices supplémentaires ou des cortex dorsolatéral préfrontal et pariétal) (Bogacz et al., 2010; Purcell et al., 2010). Cependant, de récents résultats mettent en évidence que des conditions favorisant la vitesse de décision tendent à amplifier les signaux visuels, mettant en avant que le CVE n'est pas seulement contrôlé au niveau moteur et décisionnel, mais aussi au niveau perceptif (Heitz & Schall, 2013; Schall, Purcell, Heitz, Logan, & Palmeri, 2011). Cela souligne l'activité conjointe de systèmes perceptivo-moteurs permettant à la fois d'accélérer la décision et d'augmenter l'intégration d'information sensorielle pour s'adapter à de fortes contraintes temporelles.

Ainsi, le processus de prise de décision peut-être activement modulé pour optimiser l'exactitude des décisions sous des contraintes temporelles limitant l'accumulation d'évidence et cette capacité serait universelle dans le règne animal (Heitz, 2014). Même dans une situation où le temps n'est pas limité par des circonstances extérieures, accumuler de l'évidence est coûteux à la fois en temps et effort. Il est donc nécessaire d'équilibrer le temps passé à collecter de l'information et la précision ou l'exactitude de la décision (prenant par ailleurs en compte les conséquences anticipées d'une erreur) (Drugowitsch, Moreno-Bote, Churchland, Shadlen, & Pouget, 2012). L'enseignement à tirer de ce phénomène est qu'il est important de prendre en compte à la fois une mesure temporelle et une statistique permettant d'estimer l'exactitude ou la précision d'une réponse puisque ces mesures sont liées (Wickelgren, 1977). Lors d'une tâche psychophysique impliquant un processus de prise de décision perceptuelle, il n'est pas possible de conclure qu'une condition est plus difficile (pour quelque raison que ce soit) sur la seule base que le temps de réponse de cette condition est plus élevé, sans vérifier que le taux de réponses erronées est aussi plus élevé. Ainsi, prendre en compte à la fois la vitesse et l'exactitude des réponses permet de déterminer si une condition, par exemple une information contextuelle congruente ou incongruente avant une cible à identifier, en facilite l'identification (temps de réponse courts et taux de réponse erronés faibles), la gêne (temps de réponse longs et taux de réponse erronés élevés) ou bien biaise les réponses vers la vitesse ou au contraire vers l'exactitude des réponses. Un biais vers la vitesse se traduit par l'adoption d'une stratégie de précipitation ou d'anticipation, un biais vers l'exactitude par une stratégie prudente. Cette distinction est donc importante pour l'interprétation des mécanismes psychophysiques et décisionnels en jeu.

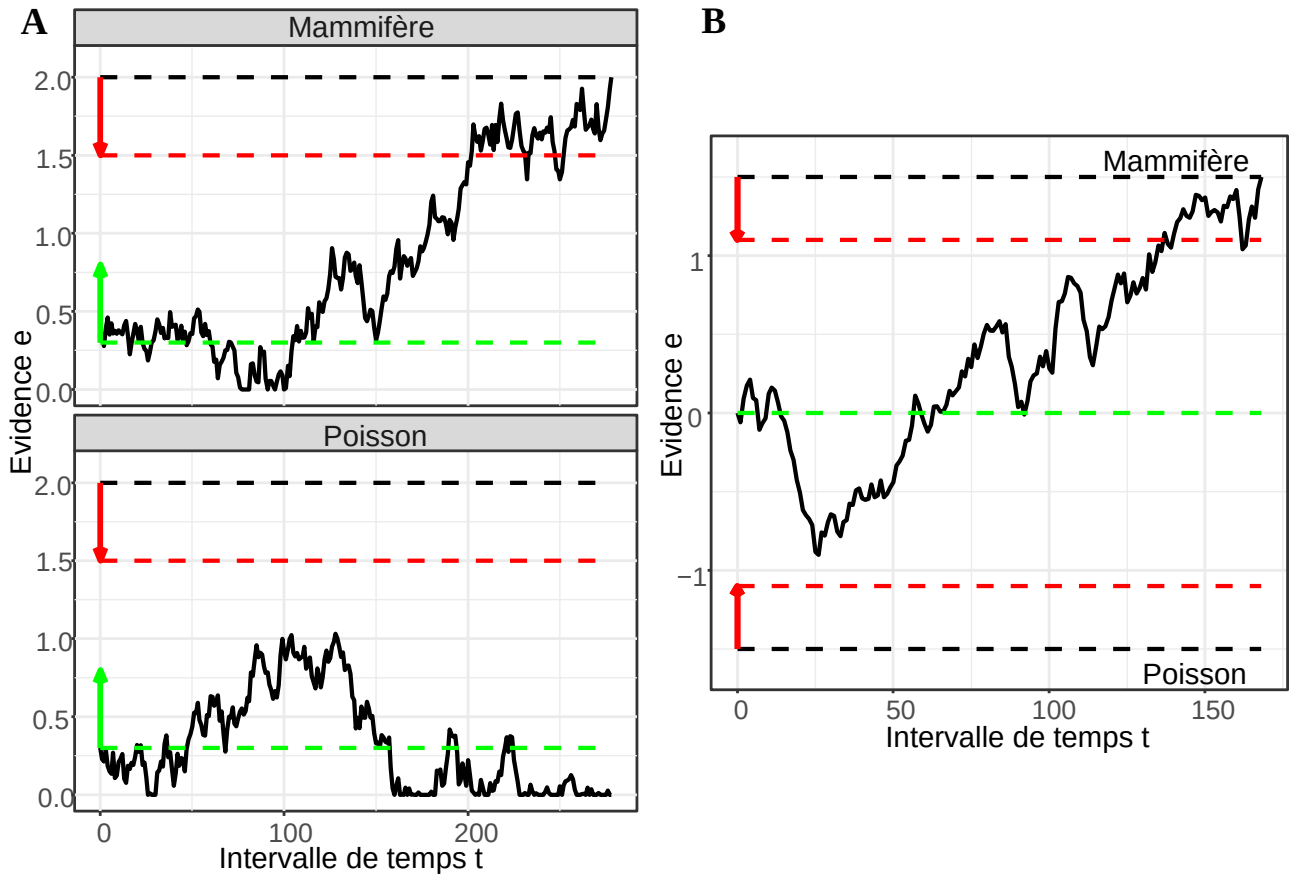


Figure 3 : Illustration des implémentations du compromis vitesse-exactitude dans les modèles à accumulateurs (A.) et les modèles de diffusion (B.). Les droites pointillées noires et vertes représentent les seuils de décision et les niveaux de référence (accumulateurs) ou point de départ (diffusion), respectivement. **A.** Pour le modèle à accumulateurs, biaiser la réponse vers la vitesse diminue les seuils de décision des deux réponses (flèches et droites pointillées rouges) ou augmente leur niveau de référence (flèche verte). **B.** Pour le modèle de diffusion, biaiser la réponse vers la vitesse rapproche les seuils de décision vers le point de départ de l'accumulation d'évidence (flèches et droites pointillées rouges). Dans les deux cas, ces manipulations diminuent la durée de la décision (moins d'évidence à accumuler) et augmentent probabilité que le seuil de décision de la réponse incorrecte soit franchi par une accumulation de bruit.

Cependant, la mise en évidence du CVE n'indique rien sur les mécanismes permettant la collecte d'information visuelle pertinente et ainsi, l'accumulation d'évidence. A ce propos, les mouvements des yeux permettent d'orienter le regard vers l'information pertinente et permettent son traitement avec une résolution plus importante. L'adaptation du CVE permet donc de laisser plus ou moins de temps pour explorer visuellement le stimulus à catégoriser ou la scène visuelle dans laquelle rechercher une cible.

1.2.2. Accumulation d'évidence via les mouvements des yeux

Dans le processus d'accumulation d'évidence, les mouvements des yeux permettent d'échantillonner l'environnement visuel en portant les localisations d'intérêt en vision fovéale et accumuler progressivement de l'information à mesure que durent les fixations (Kowler, 2011; Orquin & Mueller Loose, 2013). Ainsi, il a été montré que l'endroit où est fixé le regard sur des figures ambiguës à leur apparition (figures pouvant être reconnues comme un objet ou un autre) menait plus fréquemment à déterminer que la figure représentait l'objet pour lequel la fixation initiale captait de l'information diagnostique (Kietzmann et al., 2011). Autrement dit, l'objet identifié correspondait plus fréquemment à l'objet pour lequel de l'information diagnostique était directement disponible en vision centrale à l'apparition de la figure. De plus, les fixations suivantes étaient dévolues à confirmer que le percept initial représentait bien l'objet choisi initialement (Kietzmann et al., 2011; Kietzmann & König, 2015). Il a été proposé que le percept initial permette d'élaborer une hypothèse destinée à être testée et mise à jour au fur et à mesure de l'exécution de saccades oculaires Friston et al. (2012). Les mouvements des yeux permettent donc d'accumuler de l'évidence au regard du processus de prise de décision perceptuelle permettant la résolution de la tâche (ici, d'identification).

Ce lien a aussi été étudié au niveau neuronal. Dans une tâche où des primates non-humains devaient produire leur réponse via la modalité oculomotrice (effectuer une saccade unique vers une cible parmi des distracteurs), deux populations de neurones ont été identifiées au niveau des aires impliquées dans la programmation, le contrôle et l'exécution des saccades (Palmeri et al., 2015; Purcell et al., 2010; Purcell, Schall, Logan, & Palmeri, 2012). Des neurones dits visuels (*visually responsive neurons, visual neurons*) –présents au niveau du colliculus supérieur, champ oculaire frontal et cortex intrapariétal– seraient impliqués dans l'intégration d'évidence visuelle. Les neurones dits du mouvement (*movement-related neurons, movement neurons*) –présents au niveau du champ oculaire frontal et du colliculus supérieur– seraient impliqués dans l'accumulation de l'évidence intégrée par les neurones visuels en faveur des réponses alternatives, permettant *in fine* le déclenchement de la saccade de réponse. Les auteurs se servent de la distinction entre ces populations de neurones pour construire un modèle d'accumulateurs à palier (*gated accumulator model*) où l'évidence n'est accumulée en faveur d'une réponse (neurones du mouvement) qu'à partir d'une certaine quantité d'évidence intégrée par les neurones visuels, et rendant compte à la fois des réponses comportementales et de l'activité neuronale. Ils mettent en avant le rôle du système oculomoteur dans la prise de décision perceptuelle pour l'exécution d'une réponse (lorsque la saccade est la modalité de réponse) comme pour la détermination de l'endroit où effectuer la prochaine saccade sur un stimulus ou une scène visuelle (les auteurs suggèrent que l'activité des neurones visuels permettent la construction progressive d'une carte de saillance).

La détermination de la localisation des fixations est capitale pour extraire prioritairement l'information pertinente, de manière à minimiser la durée de la décision et ainsi, le coût associé à cette durée (Drugowitsch et al., 2012), tout en conservant un certain niveau d'exactitude. Ainsi, il a été proposé que le regard soit porté prioritairement aux endroits permettant une bonne performance à la tâche (Kowler, 2011). Les localisations ainsi priorisées permettent de maximiser l'information pertinente captée (Kowler, 2011; Renninger, Verghese, & Coughlan, 2007) ou alternativement, de minimiser l'incertitude ou la surprise sur la composition de la scène (Feldman & Friston, 2010; Friston et al., 2012; Itti & Baldi, 2009; Sprague & Ballard, 2003). Cette idée est illustrée par des résultats expérimentaux et notamment la mise en évidence du biais central de fixation (Tatler, 2007), puisque les alentours du centre d'une image sont la localisation optimale pour réaliser des fixations et maximiser l'information captée sur l'ensemble de l'image en vision périphérique tout en collectant de l'information locale à un nouvel endroit de la scène. Alternativement Renninger et al. (2007) ont montré que lors d'une tâche d'apprentissage et discrimination de formes (centrées sur l'identification plus que la recherche visuelle), les fixations du regard permettaient d'accumuler de l'information locale aux endroits les plus informatifs pour ces tâches.

Fixer longtemps le regard sur un stimulus permet donc d'accumuler de l'évidence sur ce stimulus, au regard des réponses alternatives étant l'objet du processus de prise de décision perceptuelle. Il a été montré que la durée des fixations était influencée par l'information visuelle entrante (Henderson & Smith, 2009; Nuthmann, Smith, Engbert, & Henderson, 2010), et qu'une information bruitée augmentait la durée des fixations (Harley, Dillon, & Loftus, 2004; Röhrbein, Goddard, Schneider, James, & Guo, 2015). De plus, des modèles computationnels ayant été proposés, reposent sur l'idée que l'information accumulée durant une fixation biaise un jugement de préférence en faveur du stimulus fixé (Armel, Beaumel, & Rangel, 2008; Krajbich et al., 2015, 2012). À l'inverse, effectuer de nombreuses saccades permet de récupérer de l'information à divers endroit de l'image et accumuler rapidement à la fois des échantillons d'information locale à chaque endroit fixé et des échantillons d'information globale à partir de différentes localisations (et permettant ensuite de combiner plusieurs échantillons d'information grossière) (Tatler & Vincent, 2008, 2009).

On peut ainsi distinguer deux types de comportements d'exploration visuelle. Le premier, caractérisé par de longues fixations, des saccades de faible amplitude et permettant d'extraire plus d'information locale, est désigné par Pannasch et Velichkovsky (2009) comme un mode traitement local (*local processing mode*). Le second, caractérisé par de courtes fixations, des saccades de grande amplitude et permettant d'intégrer de l'information sur l'ensemble de la scène, est désigné par ces auteurs comme un mode de traitement ambiant (*ambient processing mode*). Tatler et Vincent (2008) mettent en avant l'alternance entre ces deux types de comportements lors de l'exploration visuelle de scènes naturelles, où

une saccade de longue amplitude permet de porter le regard vers une aire pertinente, suivie de plus courtes saccades (longues fixations) qui permettent de capter l'information locale contenue dans cette aire. Lemonnier, Brémond et Baccino (2014) se sont appuyés sur ce type de distinctions pour montrer l'implication de ces deux types de comportements oculomoteurs à différents stades d'un processus de prise de décision dans la conduite routière, consistant à stopper ou non le véhicule à un croisement. Dans le cadre de la terminologie communément employée dans le domaine de la recherche visuelle (Castelhana et al., 2008; Malcolm & Henderson, 2009), le mode ambiant peut-être rapproché d'un comportement de balayage (traduction de *scanning*) et le mode local d'un comportement de vérification, que emploierons préférentiellement et que nous détaillerons par la suite. Ces comportements permettent ainsi de moduler activement la collecte d'évidence.

1.3. Résumé et conclusion

La présente section a passé en revue un pan de la littérature sur la prise de décision perceptuelle en tentant de mettre en avant un lien avec la perception visuelle active. Nous nous sommes focalisés sur deux catégories de modèles classiques, la théorie de la détection du signal, et les modèles d'échantillonnage séquentiel. Ces derniers peuvent être vus comme une extension de la théorie de la détection du signal, implémentant l'aspect continu du processus de prise de décision en intégrant la notion d'accumulation d'évidence et permettant ainsi de prédire à la fois l'exactitude des réponses et la durée des décisions. Cette accumulation d'évidence est souvent assimilée à l'accumulation d'information perceptive *bottom-up*. Par ailleurs, nous avons mis en avant comment ces modèles intégraient l'influence des attentes *a priori*, à savoir soit comme biaisant directement le choix de réponse (biais du niveau de référence), soit comme biaisant le processus d'accumulation d'évidence lui-même (modulation du taux d'accumulation/de déviation) et réduisant ainsi la quantité d'évidence à accumuler pour parvenir à la réponse attendue.

L'accumulation d'évidence n'est pas un processus passif ; un phénomène important dont ces modèles rendent compte est le compromis vitesse-exactitude. Il permet de moduler la durée allouée à l'accumulation d'évidence, de manière à prendre des décisions plus exactes sous faible contrainte temporelle ou au contraire, de s'adapter à des situations où le temps est limité (réduisant l'exactitude des réponses). Cependant si, sous leur forme traditionnelle, les modèles d'échantillonnage séquentiel rendent compte du CVE, ils sont aveugles aux mécanismes qui permettent la captation de l'information visuelle devant être traitée prioritairement. Le contrôle du regard est l'un de ces mécanismes et nous avons tenté de rendre compte de la façon dont l'exploration visuelle pouvait refléter les mécanismes cognitifs sous-tendant l'accumulation d'évidence, en particulier lors de la recherche visuelle. Nous avons conclu par la

Chapitre 2 – Prise de décision et collecte d'information : méthode expérimentale

mise en avant de deux types de comportements oculomoteurs, le comportement de balayage (ou mode ambiant) et le comportement de vérification (ou mode local). Établir cette distinction, même si elle peut être vue comme réductrice, permet de résumer simplement l'exploration visuelle de manière à étudier son influence sur la prise de décision elle-même, lors de la recherche ou de l'identification d'une cible.

De manière générale l'influence conjointe d'information sensorielle et d'attentes sur les mécanismes actifs de collecte d'information tels que le CVE et l'exploration visuelle ont été étudiés extensivement (comme passé en revue dans les sections 3. du Chapitre 1 et 1.2.1 du Chapitre 2). Par contre, l'effet de l'information sensorielle et des attentes sur la prise de décision, en particulier la sélection de la réponse, tout prenant en compte leur influence sur le CVE et l'exploration visuelle a reçu moins d'attention. Ces deux aspects (effet sur la collecte d'information et effet sur la décision elle-même) sont généralement étudiés séparément. Le présent travail cherche à les étudier de façon conjointe.

2. Problématique et méthodologie générale

2.1. Problématique générale

Cette thèse vise à éclaircir les relations entre mécanismes comportementaux permettant de recueillir de l'information visuelle provenant de l'environnement (notamment oculomoteur mais aussi via le CVE), et processus de prise de décision perceptuelle permettant d'interpréter cette information visuelle au regard de la tâche en cours d'exécution. Il est focalisé sur deux tâches classiques, la tâche de catégorisation ou reconnaissance d'objets et la tâche de recherche visuelle ou détection d'objets. Comme mis en avant précédemment, ces tâches sont des processus centraux dans la perception visuelle en cela qu'elles permettent de donner du sens au flux d'information visuelle et de structurer l'espace environnant, les objets qui le composent et leur localisation. De plus, ce sont des processus mis en jeu lors de l'exécution de tâches quotidiennes. Dans une activité visant, par exemple, à préparer du thé, la recherche visuelle et l'identification des différents ingrédients et outils sont requises.

Nous avons précédemment passé en revue les connaissances contemporaines quant à l'intégration d'information sensorielle avec les attentes dans le cadre de ces tâches, pour la perception active et l'exploration visuelle en particulier (Chapitre 1 Section 3. et Chapitre 2 Sections 1.1.3. et 1.2.). La composante décisionnelle de la catégorisation/identification et de la recherche visuelle d'objets a été prise en compte par le passé, notamment au travers de la TDS et des modèles d'échantillonnage séquentiel (que ce soit avec des tâches de détection ou de catégorisation) mais son articulation avec les aspects actifs, en particulier oculomoteurs, a fait l'objet de moins d'attention. De plus, s'il est communément admis que la prise de décision est un processus étendu dans le temps, les études expérimentales traditionnelles, ne mesurant que les taux d'erreurs et temps de réponse, sont aveugles à ses aspects continus et dynamiques.

Un certain nombre de méthodes ont été proposées pour adresser cette limitation, parmi lesquelles l'estimation de paramètres reflétant l'évolution de la décision dans le temps via des modèles ajustés à partir de données expérimentales (utilisant par exemple, des modèles d'échantillonnage séquentiel, Forstmann et al., 2016), l'usage d'EEG chez l'être humain ou l'enregistrement neuronal direct chez le primate non-humain (Heekeren et al., 2008). Alternativement, il a été proposé d'enregistrer la trajectoire des mouvements d'une souris d'ordinateur ou de la main lors d'un choix entre plusieurs alternatives pour inférer de manière continue, certains aspects du processus de prise de décision (Song & Nakayama, 2008; Spivey, Grosjean, & Knoblich, 2005). Cependant, peu d'études focalisées sur la catégorisation et recherche visuelle d'objets employant ces méthodes intègrent en parallèle des analyses relative à la perception active telles que celle des mouvements des yeux ou du CVE (néanmoins, nous avons précédemment mis en avant

Chapitre 2 – Prise de décision et collecte d'information : méthode expérimentale

Friston et al., 2012; Kietzmann et al., 2011; Kietzmann & König, 2015; Purcell et al., 2012).

Le principal objectif de ce travail est donc d'étudier expérimentalement les effets des attentes et de la force de l'évidence sensorielle présente dans l'information visuelle (*e.g.*, signal/bruit) à la fois sur la dynamique du processus de collecte active d'information visuelle (via le CVE et les mouvements des yeux) et sur celle du processus de prise de décision. Nous porterons un intérêt particulier à l'influence consécutive que le processus de collecte d'information a sur le processus de prise de décision (notamment la sélection de la réponse). Ainsi, il est attendu que la force de l'évidence sensorielle et les attentes modulent la collecte active d'information (*e.g.*, une faible force demande une collecte d'information plus longue, mesurable via les mouvements des yeux), et que cette modulation affecte à son tour la sélection des réponses.

Le modèle ainsi décrit est représenté par la Figure 4.A. Cette démarche générale permet aussi, à l'inverse, d'évaluer l'influence directe de la force de l'évidence *bottom-up* et des attentes sur la sélection de réponse (ne passant pas par la modulation du processus actif d'accumulation d'évidence via les mouvements des yeux), telle que représentée par les connections en pointillés sur la Figure 4.A.. À titre d'illustration, cette distinction entre effet direct des attentes sur la sélection de réponse et effet indirect des attentes sur la sélection de réponse via leur modulation du comportement de collecte d'information peut être rapprochée (très globalement), dans le cadre des modèles d'échantillonnage séquentiel, d'une modulation du niveau de référence (effet direct) ou du taux de diffusion (effet indirect) (voir Section 1.1.3.).

Le présent travail se focalise donc sur la contribution de mécanismes de collecte d'évidence dans la sélection de la réponse. Cependant, il est important de souligner l'influence de l'état de la prise de décision (*i.e.*, quelle hypothèse est la plus probable sachant l'évidence accumulée jusqu'à présent) sur la stratégie de collecte d'évidence (Friston et al., 2012; Kietzmann & König, 2015). La Figure 4.B. schématise ce lien bidirectionnel entre collecte d'évidence et sélection de réponse, bien que les différentes études présentées par la suite ne le testent pas directement.

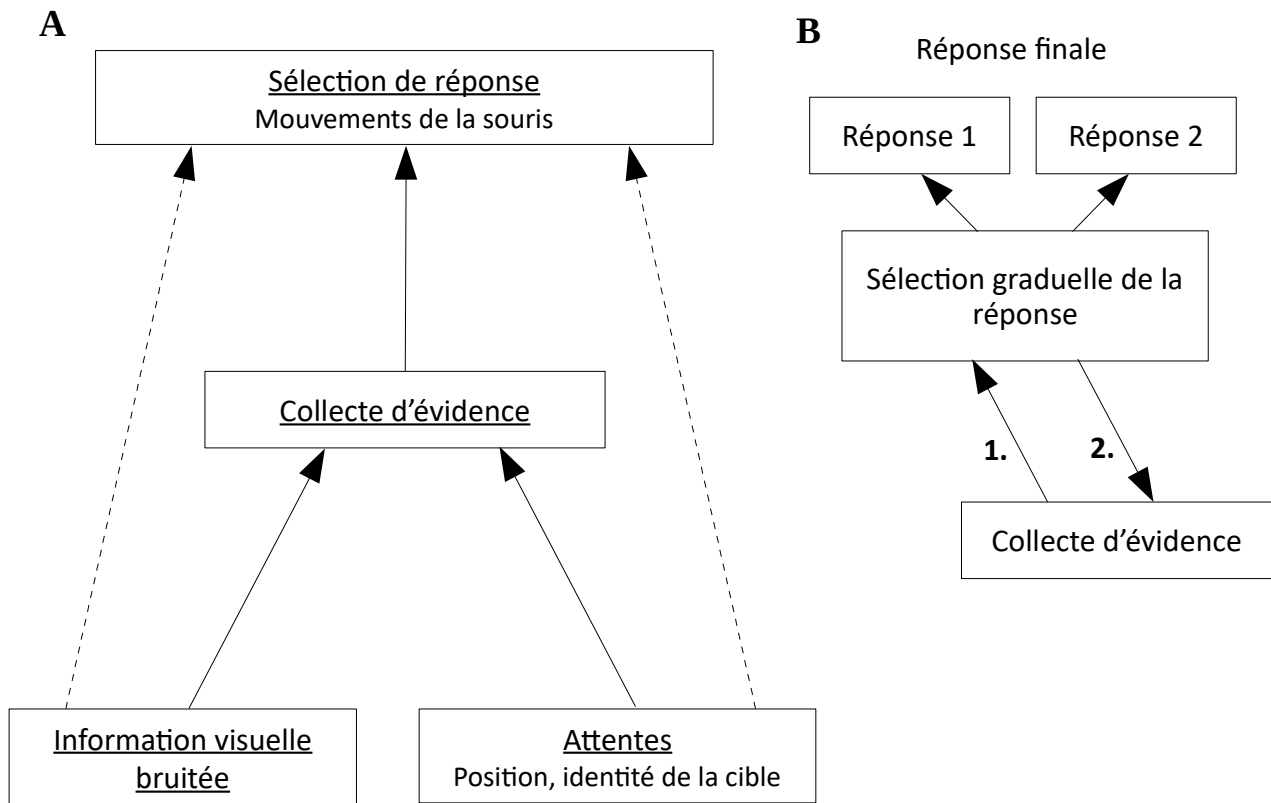


Figure 4 : **A.** Opérationnalisation du processus de prise de décision dans les Études 1, 2 et 3. L'information visuelle bruitée et les attentes influencent la décision via la modulation de la collecte active d'évidence (effet indirect, flèches pleines), et en affectant directement la sélection de la réponse (effet direct, flèches pointillés). Dans l'Étude 1, la collecte d'évidence est modulée via le compromis vitesse-exactitude. Le CVE affectant à la fois l'accumulation d'évidence et la sélection de réponse, les effets directs et indirects (*i.e.*, effet d'une variable sur la décision par la modulation de la collecte d'évidence) ne sont pas estimés mais ce décours est inféré. Dans les Études 2 et 3, la collecte d'évidence est estimée via les temps de détection et de vérification et les effets directs et indirects sont estimés.

B. Précisions concernant la relation entre collecte d'évidence et sélection de réponse dans le processus de prise de décision. **1.** Comme présenté dans la partie A., les mécanismes de collecte active d'évidence permettent son intégration et son accumulation et ainsi, la sélection de la réponse. **2.** L'état de l'accumulation d'évidence permet de mettre à jour les hypothèses sur ce qui est perçu et adapter les comportements de collecte d'évidence en conséquence (*e.g.*, le comportement oculomoteur, Friston et al., 2012; Kietzmann et al., 2011; Kietzmann & König, 2015). Cet aspect n'est pas traité dans le cadre des Études 1, 2 et 3, bien qu'il soit pris en compte dans la discussion de nos travaux.

La collecte d'évidence étant vue comme progressive, nous avons opté pour une méthodologie reposant sur les mouvements de la souris comme modalité de réponse, permettant d'estimer des aspects continus de la sélection de réponse et plus globalement, du processus de prise de décision. Dans l'optique où les mouvements des yeux permettent d'échantillonner l'évidence disponible en faveur des différentes décisions alternatives et où le CVE module le temps alloué à son accumulation, nous les avons utilisés comme estimation de la collecte d'évidence sensorielle. Estimer le CVE est une estimation indirecte de stratégies de collecte d'évidence alors que les mouvements des yeux mesurent plus directement la collecte d'évidence. Nous avons relevé précédemment que les mouvements des yeux pouvaient viser différents buts. La métrique des saccades et fixations oculaires permet par exemple la distinction entre comportements oculomoteurs de balayage de la scène visuelle et de vérification de l'objet.

Ce travail vise donc à explorer d'une part, l'intégration de l'information *bottom-up* et des attentes sur chacun de ces deux types de comportements et d'autre part, leurs contributions respectives au processus de prise de décision lors de la recherche visuelle. En effet, différentes tâches de recherche, ou différents types d'attentes, telles que les attentes sur l'identité d'une cible ou sur sa localisation dans l'espace, peuvent affecter différemment le rôle des comportements de balayage et de vérification dans le processus de prise de décision. Par exemple, des attentes sur la localisation de la cible affectent le balayage de manière plus consistante qu'elles n'affectent la vérification.

2.2. Mesurer la prise de décision : le paradigme de *mouse-tracking*

La méthodologie employée dans les différentes études composant ce travail vise donc à mettre en correspondance d'une part la prise de décision, en particulier en rendant compte de son caractère continu et d'autre part, la collecte d'information sensorielle présidant à l'accumulation d'évidence en mesurant les mouvements des yeux ou en prenant en compte le compromis vitesse-exactitude. De manière à évaluer le processus de prise de décision, nous avons employé une méthode expérimentale, le paradigme de *mouse-tracking*, initialement introduite en psychologie par Spivey et al. (2005) mais ayant pris un essor important notamment en cognition sociale, suite au développement d'un outil facilitant son emploi : le logiciel Mouse-Tracker (Freeman & Ambady, 2010). Ce paradigme, présenté dans la Figure 5, permet de mesurer la dynamique des mouvements de réponse. Il repose sur une tâche de classification ou de jugement à choix forcé binaire où les participants doivent effectuer un mouvement de souris (d'ordinateur) vers le bouton correspondant à la réponse la plus appropriée pour le stimulus leur étant présenté. Classiquement, pour démarrer chaque essai, ils doivent cliquer sur un bouton *start* en bas de l'écran, centré sur l'axe x. Le stimulus –*e.g.*, une image ou un mot– apparaît alors au centre de l'écran. Généralement, il est plus ou

moins ambigu, c'est à dire qu'une part plus ou moins grande de ses caractéristiques (*e.g.*, visuelles, sémantiques) correspondent à la réponse incorrecte. Les participants doivent alors cliquer le plus rapidement possible sur le bouton correspondant à la réponse correcte (*e.g.*, la catégorie du stimulus), située dans le coin supérieur gauche ou droit de l'écran. La trajectoire de leur mouvement de réponse est enregistrée. Une hypothèse fondamentale de ce paradigme est que lorsqu'un stimulus ambigu est présenté, la trajectoire devrait être attirée en direction de la réponse incorrecte : la compétition entre les deux réponses étant plus forte, les participants ont plus d'hésitation à l'encontre de la réponse incorrecte que lorsque le stimulus n'est pas ambigu.

En plus de mesurer les temps de réponse, cette méthode donne accès à des mesures plus riches, enregistrées durant le processus de prise de décision et permettant ainsi d'évaluer différents aspects continus et progressifs du choix (Freeman, Dale, & Farmer, 2011). Parmi les mesures habituelles employées dans une étude de *mouse-tracking*, nous avons choisi de détailler ici, l'aire sous la courbe (ASC, *area under the curve*, AUC) et le maximum de déviation (MD), bien que d'autres mesures soient présentées dans les articles inclus dans cette thèse. La première désigne l'aire géométrique entre la trajectoire de la souris et une trajectoire idéale correspondante à l'absence totale d'incertitude ou d'hésitation, *i.e.* une droite partant du bouton *start* jusqu'au bouton de réponse correct. La seconde désigne la distance maximale entre la trajectoire de la réponse et la trajectoire idéale par projection orthogonale. Ces deux mesures résument de deux manières différentes la déviation de la trajectoire : plus l'ASC et le MD sont élevés, plus la déviation de la souris vers la réponse incorrecte est forte. Ainsi, elles peuvent être utilisées pour estimer globalement le biais en faveur de la réponse alternative dans le processus de prise de décision (Freeman & Ambady, 2010). Ces mesures estiment grossièrement la déviation. Il est aussi possible d'analyser plus finement la dynamique des trajectoires, en particulier la déviation et la vitesse du mouvement en fonction du temps, apportant de l'information sur l'évolution de la compétition entre les deux réponses dans le temps (Farmer, Cargill, Hindy, Dale, & Spivey, 2007; Freeman & Ambady, 2011; Freeman, Ma, Han, & Ambady, 2013; Hehman, Stolier, & Freeman, 2015).

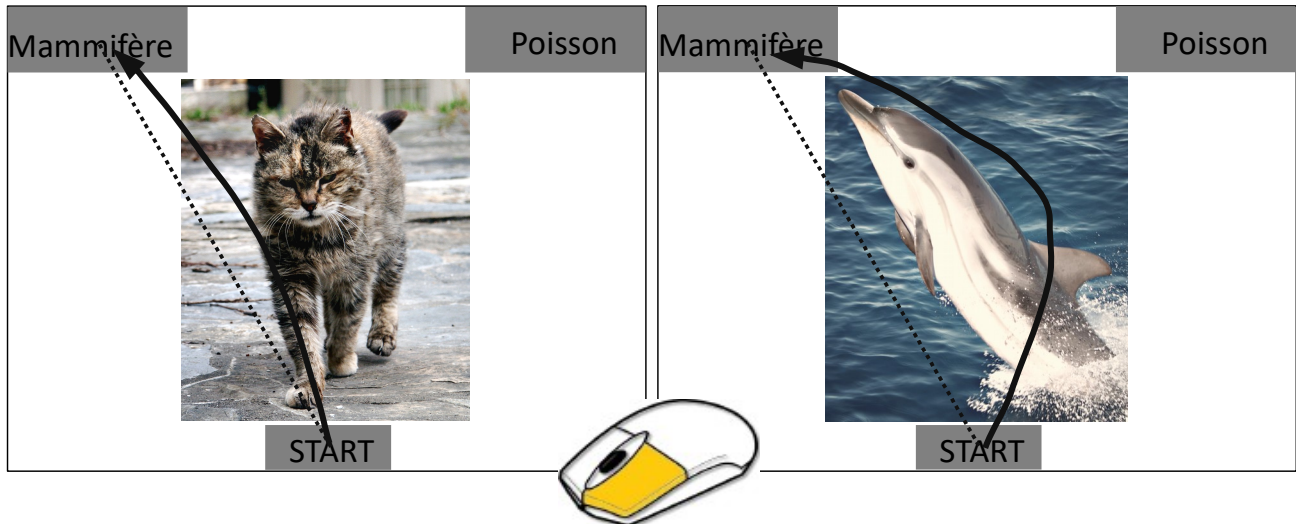


Figure 5 : Illustration du paradigme de *mouse-tracking*. Après avoir cliqué sur le bouton *start*, le stimulus est présenté et les participants doivent effectuer un mouvement de souris pour aller cliquer sur la réponse correspondante au stimulus. Ici, le stimulus représente un animal non ambigu (gauche) ou ambigu (droite) au regard de la tâche de catégorisation. Le segment noir en pointillés indique une trajectoire de réponse théorique idéale. Pour chaque image, la courbe noire représente un exemple de trajectoire de réponse en conformité avec les hypothèses classiques du paradigme : les trajectoires de réponse des stimuli ambigus devraient plus dévier vers la réponse incorrecte que celles des stimuli non ambigus. Ici, un dauphin partage de nombreuses caractéristiques avec la catégorie des poissons.

Ces interprétations reposent sur l'idée que la dynamique du mouvement est directement liée à l'accumulation d'évidence. Il a été argumenté que le système moteur soit modulé en continu par les systèmes cognitifs impliqués dans la prise de décision (Quinton et al., 2014; Song & Nakayama, 2009; Spivey et al., 2005). Il a aussi été proposé que le système cognitif de contrôle du mouvement ne puisse accéder au processus décisionnel (*e.g.*, état de l'accumulation d'évidence) qu'à certains moments (Friedman, Brown, & Finkbeiner, 2013). Ces derniers auteurs montrent qu'un modèle de diffusion ayant un accès intermittent à l'information visuelle (mouvement de points plus ou moins cohérent), bien que plus parcimonieux qu'un modèle de diffusion ayant un accès continu, est suffisant pour rendre compte des trajectoires produites par les participants. Cependant, notre objectif n'est pas de discriminer ces deux hypothèses, celles-ci s'accordant sur le lien entre accumulation d'évidence et contrôle du mouvement, ce qui est suffisant dans le cadre de ce travail.

La tâche de *mouse-tracking* a été employée dans de nombreux domaines d'étude, notamment la cognition numérique (Faulkenberry, 2014; Marghetis, Núñez, & Bergen, 2014), la psycholinguistique (Barca, Benedetti, & Pezzulo, 2016; Barca & Pezzulo, 2012, 2015; Farmer et al., 2007; Morett & MacWhinney, 2013) et la mémoire (Papesch & Goldinger, 2012). Dans le domaine de la catégorisation sur la base de matériel visuel, cette méthode a été employée par Dale, Kehoe, & Spivey (2007) dans une tâche où des

exemplaires plus ou moins ambigus au regard de catégories d'animaux (*e.g.*, une baleine pour un choix entre mammifère et poisson) provoquaient des trajectoires déviant en direction de la réponse alternative. Ce type de résultat a été répliqué dans une série d'études portant sur la reconnaissance de traits de visages tels que le genre (Freeman, Ambady, Rule, & Johnson, 2008; Johnson, Freeman, & Pauker, 2012) et l'ethnie (Freeman & Ambady, 2011; Freeman, Pauker, Apfelbaum, & Ambady, 2010), via la présentation de visages plus ou moins ambigus au regard de ces tâches de discrimination. De plus les auteurs ont montré une influence *top-down* des stéréotypes dans la reconnaissance de l'ethnicité de visages (Freeman et al., 2013; Freeman, Penner, Saperstein, Scheutz, & Ambady, 2011) ou la formation d'impressions sur des visages ayant des traits masculins ou féminins (Freeman & Ambady, 2009). Enfin, Quinton et al. (2014) ont utilisé des dessins d'animaux formés de segments, dont l'ambiguïté au regard de leur catégorie d'appartenance (*e.g.*, chien ou girafe) était manipulée en faisant varier des caractéristiques précises (*e.g.*, longueur du segment dessinant le cou, angle entre les segments associés aux membres). Ils ont construit un modèle computationnel implémentant des principes issus du codage prédictif, déterminant d'une part où porter l'attention sur le stimulus (en combinant saillance visuelle et influences *top-down*), et produisant d'autre part des trajectoires de réponses analogues aux trajectoires de la souris produites par des participants humains. Le modèle reproduit ainsi l'effet de l'ambiguïté des stimuli.

Néanmoins, à notre connaissance, aucune étude expérimentale n'a directement étudié le lien entre dynamique du choix telle que mesurée via la tâche de *mouse-tracking* et l'allocation du regard sur le stimulus. De même, la question du CVE dans le cadre du paradigme de *mouse-tracking* n'a pas encore été abordée.

2.3. Présentation des études expérimentales

Le présent travail est composé de trois études (séries d'expériences). La méthodologie générale que nous avons employée consiste à mesurer différents aspects du processus de prise de décision perceptuelle, notamment son évolution temporelle via l'analyse de la dynamique des mouvements de la souris en utilisant le paradigme de *mouse-tracking*, mettant en évidence la dynamique du choix entre deux alternatives. Nous proposons de mettre en relation cette dynamique avec des mécanismes actifs de collecte d'évidence (CVE ou exploration visuelle), l'accumulation d'évidence étant un aspect fondamental des modèles de prise de décision. Nous examinons l'intégration d'attentes avec la collecte d'une évidence visuelle de force variable (manipulée au moyen de bruit visuel ou de brouillard) au travers de deux tâches classiques pouvant être vues comme impliquant un processus de prise de décision perceptuelle : la catégorisation de cible et la recherche visuelle dans des scènes visuelles naturelles.

À noter que par la suite (notamment dans les articles présentés dans la section suivante mais écrits antérieurement à la thèse), nous employons la terminologie « force de l'évidence visuelle » (ou sensorielle, ou *bottom-up*) ainsi que la terminologie « incertitude de l'évidence visuelle » (ou sensorielle, ou *bottom-up*) pour référer à la même idée : que l'information visuelle du stimulus puisse contenir plus ou moins d'évidence (ou de l'évidence plus ou moins facilement accessible) en faveur de chaque alternative, et qui requiert donc d'adapter la quantité d'information à accumuler (*i.e.* moduler la durée de l'accumulation d'évidence). Ainsi, l'ajout de bruit visuel diminue la force de l'évidence visuelle et augmente son incertitude.

Dans l'Étude 1 (Article 1) nous proposons trois expériences employant une tâche de catégorisation où la force de l'évidence visuelle contenue dans les stimuli est manipulée via l'adjonction de bruit visuel gaussien et les attentes sont manipulées via la congruence d'une image contextuelle avec la cible. Cette étude met en avant la modulation de l'accumulation d'évidence via le CVE en présence d'information contextuelle préalable. Les Études 2 et 3 se focalisent sur le rôle du comportement oculomoteur dans l'accumulation d'évidence, l'exploration visuelle étant un moyen plus direct de mesurer la collecte active d'information que l'évaluation du CVE. L'Étude 2 (Article 2) propose une expérience employant une tâche de recherche visuelle où la force de l'évidence visuelle est manipulée via l'adjonction de bruit gaussien, et les attentes par indiçage contextuel de la localisation de la cible. Enfin, l'Étude 3 (Articles 3 et 4) propose deux expériences employant une tâche combinant identification de cible et recherche visuelle où la force de l'évidence visuelle est manipulée par la quantité de brouillard présent dans la scène visuelle et les attentes par la probabilité qu'un stimulus cible ou un distracteur fasse partie de la scène. De plus, cette dernière étude intègre thématiquement des aspects applicatifs sur la vision en situation de conduite dans le brouillard, reproduisant un environnement routier statique et manipulant des attentes pertinentes dans ce domaine.

Cette présente section vise donc à détailler nos méthodes pour évaluer le CVE et l'exploration visuelle via les mouvements des yeux, et comment nous les avons opérationnalisés dans leurs études respectives, au regard de nos objectifs théoriques.

2.3.1. Évaluer la collecte active d'évidence via le compromis vitesse-exactitude

Nous avons vu précédemment (Section 1.2.1.) que lorsque l'on mesure à la fois la vitesse d'une décision et son exactitude, il est possible de distinguer des conditions où la tâche de prise de décision est facilitée, dégradée, biaisée en faveur de la rapidité de la décision ou biaisée en faveur de son exactitude (Wickelgren, 1977). Dans une tâche traditionnelle où les temps de réaction sont mesurés (vitesse de décision), l'exactitude des réponses est déterminée à la fin de chaque essai, par son caractère *correct* ou *erroné*. La difficulté de la tâche doit être échelonnée de manière à produire un nombre minimum d'erreurs

même dans les conditions les plus simples. Le paradigme de *mouse-tracking* présente l'avantage de permettre l'évaluation de la réponse de manière dynamique, au cours de son exécution. Cela permet de mesurer la vitesse et l'exactitude des réponses de manière plus fine.

La vitesse peut être estimée via le temps de réponse final, mais aussi par exemple, via le temps d'initiation du mouvement ou sa vélocité, permettant d'évaluer plus finement l'avance ou le délai du processus de prise de décision et d'estimer la précipitation de la réponse. La relation entre initiation du mouvement et vélocité dans le cas où le temps pour répondre est limité (comme c'est le cas dans une tâche de *mouse-tracking*), apporte une complexité supplémentaire. Ainsi, un temps d'initiation court peut mener à une vélocité plus faible qu'un temps d'initiation long, étant donné qu'un mouvement commencé plus tôt a plus de temps pour être terminé avant la limite. De plus, dans le paradigme de *mouse-tracking*, le mouvement de réponse doit débiter avant la fin du processus de prise de décision. Ainsi, si son initiation est avancée, la programmation initiale du mouvement repose sur moins d'évidence, la décision est plus éloignée de sa complétion et la vélocité du mouvement est en conséquence faible (Battaglia & Schrater, 2007). L'initiation du mouvement et par extension, dans quelle mesure le mouvement de réponse global est anticipé ou retardé est donc une mesure plus pertinente de la vitesse de décision que la vélocité du mouvement.

Une tâche de *mouse-tracking* présentant des réponses plus continues, les mouvements de réponse peuvent être corrigés, rendant l'occurrence d'erreurs très rare. Par contre, une réponse incertaine ou une décision précoce en faveur de la réponse incorrecte ont tendance à dévier la trajectoire du mouvement vers celle-ci (Buc Calderon, Verguts, & Gevers, 2015; Friedman et al., 2013; Resulaj, Kiani, Wolpert, & Shadlen, 2009). Dans ce cas, la réponse incorrecte est vue comme potentiellement juste plus longtemps, influençant en conséquence le mouvement (Buc Calderon et al., 2015). La déviation du mouvement peut donc être vue comme une mesure d'erreur dans le processus de décision. Une réponse biaisée vers la vitesse devrait donc se traduire par un mouvement effectué plus tôt, avec une déviation vers la réponse incorrecte plus forte et une réponse biaisée vers l'exactitude devrait, à l'inverse, être caractérisée par un mouvement retardé, mais plus direct vers la réponse correcte.

Dans l'Étude 1, nous avons étudié l'influence du contexte visuel et de la force de l'évidence sensorielle d'une cible à catégoriser (*vivant vs. non vivant*) sur la prise de décision, l'évaluation d'un biais du CVE permettant d'en inférer l'accumulation d'évidence. Dans ces expériences, la congruence du contexte (ainsi que son absence ou présence) était manipulée pour générer des attentes sur la catégorie de la cible et la force de l'évidence via l'addition de bruit Gaussien. Employer cette méthode nous permet donc de distinguer le biais du processus de décision vers une réponse ou une autre (facilitation ou dégradation des réponses sur les mesures de vitesse comme d'exactitude) de l'adaptation du CVE à l'information présentée.

Elle nous permet ainsi d'éclairer quels aspects du processus de prise de décision (*e.g.*, biais de réponse, CVE) sont influencés par la dégradation de la cible (bruit), les attentes sur la catégorie de la cible (congruence du contexte), ou la présentation d'un contexte lors d'une tâche de catégorisation. Cependant, évaluer le CVE tel que proposé dans cette étude ne constitue pas une mesure directe de la collecte active d'évidence, séparée de la mesure du processus de prise de décision. Les Études suivantes proposent donc d'évaluer les mouvements des yeux, et de les mettre en relation avec la prise de décision, inférée via le paradigme de *mouse-tracking*.

2.3.2. Évaluer la collecte active d'évidence via les mouvements des yeux

Dans les Études 2 et 3, nous avons cherché à évaluer directement le lien entre collecte d'information visuelle via les mouvements des yeux et processus décisionnel. L'enregistrement des mouvements des yeux rend disponible une grande diversité de mesures, donnant accès à différents processus impliqués dans la collecte d'information visuelle (Kietzmann et al., 2011; Kietzmann & König, 2015). Ainsi, nous avons précédemment mis en avant deux types de comportements oculomoteurs : le comportement de balayage permet d'intégrer de l'information sur l'ensemble de la scène par d'amples saccades et de courtes fixations et le comportement de vérification permet d'intégrer de l'information locale par de courtes saccades et de longues fixations.

Dans le cadre plus restreint de la tâche de recherche visuelle, cette distinction renvoie à la division du temps d'exploration visuelle en différentes périodes. Deux principales divisions ont été proposées dans la littérature. Castelhana et al. (2008) distinguent deux périodes : la latence de la cible, désignant le temps passé à explorer visuellement la scène avant la première fixation sur la cible, et le temps de vérification, désignant la première séquence de fixations sur la cible avant que le regard ne se porte à un autre endroit. Alternativement, Malcolm & Henderson (2009) divisent l'exploration visuelle en trois périodes : la période d'initiation (*initiation epoch*) comprend la latence de la première saccade après apparition de la scène visuelle, la période de balayage (*scanning epoch*) s'étend de cette première saccade à la première saccade terminant sur la cible et la période de vérification (*verification epoch*), est définie de la même manière que par Castelhana et al. (2008). Ainsi, la latence de la cible correspond à la somme des périodes d'initiation et de balayage. La latence de la cible correspond ainsi au temps passé pour détecter la cible alors que le temps de vérification correspond au temps d'identification de celle-ci.

Dans la suite de ce manuscrit, nous désignerons la latence de la cible par période de détection (s'il est question de l'intervalle de temps durant lequel des saccades et fixations du regard sont effectuées pour détecter la cible) et temps de détection (s'il est question de la durée de la détection elle-même). De même nous distinguerons période de vérification et temps de vérification. Cette distinction entre temps et

Chapitre 2 – Prise de décision et collecte d'information : méthode expérimentale

période, introduite par Malcolm et Henderson (2009), nous sera utile pour définir certaines mesures focalisées plus particulièrement sur une période (*e.g.*, nombre de fixations durant la période de détection), mais aussi car la notion de période introduit une distinction fonctionnelle.

En effet, Nuthmann (2014) a mis en avant une distinction fonctionnelle entre périodes de détection et de vérification, à propos des rôles respectifs de la vision centrale et périphérique lors de ces phases. Ainsi, une vision périphérique bruitée dégrade les performances en période de détection, alors qu'une vision centrale bruitée dégrade les performances en période de vérification, mettant en avant l'importance de la captation de la cible en vision périphérique pour la période de détection et de l'identification de la cible en vision centrale pour la période de vérification. Cette distinction peut être rapprochée de la distinction neurophysiologique entre circuits magnocellulaire (information majoritairement issue des bâtonnets et vision périphérique) et parvocellulaire (information majoritairement issue des cônes et vision centrale), présentés dans le Chapitre 1 Section 1.1..

Il a aussi été montré que des attentes affectaient à la fois les temps de détection et de vérification. Ainsi, il a été mis en évidence que les attentes sur l'apparence de la cible (*i.e.* représentation de la cible plus ou moins précise, manipulée par exemple en indiquant la cible à rechercher via son nom ou une image plus ou moins précise, voir Chapitre 1 Section 3.2.2.) facilitent à la fois la détection et la vérification (Castelhana & Heaven, 2010; Castelhana et al., 2008; Malcolm & Henderson, 2009, 2010, Spotorno et al., 2014, 2015), indiquant que la représentation de la cible peut-être utilisée pour la détecter en vision périphérique et guider ainsi le balayage de la scène d'une part, et d'autre part, identifier la cible plus rapidement une fois qu'elle est en vision centrale. Concernant les attentes sur la localisation de la cible, il a été montré de manière consistante qu'une cible présentée à une localisation habituelle réduit le temps de détection, indiquant que le regard est dirigé plus rapidement vers la cible (Castelhana & Heaven, 2010, 2011; Malcolm & Henderson, 2010; Pereira & Castelhana, 2014; Spotorno et al., 2014, 2015). Ces résultats ont été observés en considérant la latence de la cible (temps de détection) comme le temps de balayage uniquement. De manière générale, cet effet n'a pas été observé sur le temps d'initiation à l'exception d'une étude employant un design et des scènes visuelles adaptées spécifiquement pour étudier la période d'initiation (Spotorno et al., 2014). Par contre, l'effet des attentes sur la localisation de la cible sur le temps de vérification est plus nuancé. Certaines études rapportent une facilitation lorsque la cible est localisée à un endroit habituel (Castelhana & Heaven, 2011; Malcolm & Henderson, 2010; Vö & Henderson, 2009) alors que d'autres échouent à observer cet effet (Castelhana & Heaven, 2010; Pereira & Castelhana, 2014; Spotorno et al., 2014).

Notons néanmoins que la distinction entre période de détection d'une part et période de vérification d'autre part, ne correspond pas parfaitement à la distinction entre modes de traitement

ambient et local (définis dans la Section 1.2.). En effet, si la période de vérification implique un mode de traitement local, il est aussi possible d'adopter un mode de traitement local avant fixation de la cible (durant la période de détection), par exemple pour déterminer que la cible n'est pas dans une zone précise de la scène.

Enfin, dans l'Étude 2, nous avons analysé le nombre de fixations et la durée des fixations durant la période de balayage, ces mesures étant fréquemment rapportées dans la littérature d'une part, mais aussi car ces mesures ciblent plus spécifiquement l'effet du bruit visuel, qui tend à augmenter la durée des fixations (Röhrbein et al., 2015). La première mesure estime le nombre de régions fixées avant fixation de la cible et la seconde estime le temps passé pour déterminer que la région fixée ne contient pas la cible. Mais de manière générale, nous avons préféré mesurer l'exploration visuelle prioritairement via les temps de détection et de vérification, pour conserver un nombre limité de mesures des mouvements des yeux, de manière à simplifier les résultats et analyser de manière plus riche la prise de décision via les mouvements de la souris.

Si, comme nous l'avons souligné dans la présente section, les effets des attentes et de la dégradation de l'information visuelle sur les temps de détection et de vérification de la cible lors de la recherche visuelle ont été étudiés extensivement, ce n'est pas le cas de leur influence sur le processus de prise de décision. Les mécanismes d'accumulation d'évidence en jeu dans le processus de prise de décision relatif à la recherche visuelle (décision *cible absente/présente*) ont bénéficié de peu d'attention.

Les Études 2 et 3 testent directement l'idée que la détection et la vérification de la cible permettent (au moins en partie) cette accumulation d'évidence. Les deux études manipulent toutes deux la force de l'évidence sensorielle (bruit Gaussien dans l'Étude 2, brouillard dans l'Étude 3) et elles se focalisent chacune sur une forme d'attentes différente. L'Étude 2 se focalise sur les attentes sur la localisation de la cible dans une tâche de recherche visuelle où les participants doivent trouver des objets dans ces scènes d'intérieur. L'Étude 3 se focalise sur l'attente d'une cible dont la probabilité de faire partie de la scène est plus élevée (son apparence est connue, rapprochant cette forme d'attentes des connaissances sur l'apparence de la cible), dans une tâche combinant recherche visuelle et identification de la cible (piéton vs. cerf/chevreuil). Les effets conjugués de la force de l'évidence sensorielle et des attentes sont évalués à la fois sur les temps de détection et vérification, et sur des mesures de la prise de décision via les mouvements de souris.

Pour mettre en relation les mesures résumant l'exploration visuelle et la prise de décision, nous avons directement analysé des modèles de médiation/modération, dont la Figure 4 présente un schéma général (où la collecte d'évidence est remplacée par les temps de détection et de vérification, alors vus comme médiateurs et la sélection de la réponse par des mesures résumant les mouvement de réponse telles que les ASC). Nous avons par ailleurs analysé les trajectoires de la souris de manière continue,

Chapitre 2 – Prise de décision et collecte d'information : méthode expérimentale

mettant en relation la dynamique des mouvements et les époques de détection et vérification lorsque les attentes et la force de l'évidence varient. Cela permet de mettre en évidence la modulation de l'exploration visuelle par les attentes et la force de l'évidence, et l'effet consécutif de cette modulation sur la prise de décision. En d'autres termes, cela permet d'expliquer (une part de) l'effet des attentes et de la force de l'évidence sur la prise de décision, par l'influence de ces variables sur la détection et la vérification de la cible et ainsi éclairer les mécanismes d'accumulation d'évidence lors de la recherche visuelle. Par exemple, rechercher un objet cible à un endroit inhabituel devrait augmenter le temps de détection (accumulation d'évidence plus lente), et ainsi dégrader la décision (temps plus long, trajectoire déviée) car la réponse *cible absente* reste plausible plus longtemps et/ou de l'évidence est accumulée en sa faveur. Voir les introductions des Études 2 et 3 pour des résumés plus spécifiques des hypothèses de ces études.

Chapitre 3

Études expérimentales

1. Étude 1 : Effets conjugués de l'incertitude perceptive et des attentes générées par le contexte dans une tâche de choix forcé binaire

L'Étude 1 est composée d'un article soumis au journal *Acta Psychologica*, présentant une série de trois expériences. Les tâches employées sont des tâches de catégorisation où les participants, après avoir appuyé sur le bouton *start* pour commencer l'essai, doivent indiquer le caractère *vivant* ou *non vivant* de cibles représentant des animaux vs. des outils (Expériences 1 et 2) ou des véhicules (Expérience 3) en cliquant sur les boutons de réponse correspondants. Un contexte sémantiquement relié à la cible ou non relié (relié à une cible de la catégorie opposée), ou une image grise (aucun contexte), était présenté préalablement de manière à générer des attentes sur la catégorie de la cible. Originellement, cette étude visait à reproduire un effet d'indigence contextuel sur la reconnaissance de la cible et sur le processus de prise de décision en jeu lors de sa catégorisation. Nous nous intéressons en particulier à l'effet du contexte dans des conditions où la force de l'évidence visuelle est plus ou moins élevée.

Nous attendions donc une attraction des trajectoires de la souris vers la réponse congruente au contexte, soit une déviation plus faible, des temps d'initiation du mouvement et des temps de réponse plus courts lorsque le contexte était congruent avec la cible (facilitation de la décision) et inversement lorsqu'il était incongruent. De plus, la cible pouvait être plus ou moins bruitée dans le but de moduler la force de l'évidence sensorielle qu'elle convoie, et ainsi provoquer des temps de réponse et d'initiation plus élevés et une déviation en direction de la réponse erronée ; l'accumulation d'évidence étant plus longue, cette réponse est considérée plus longtemps. L'intégration des attentes avec l'incertitude de l'évidence sensorielle était aussi examinée. Nous attendions ainsi une augmentation de l'influence des attentes lorsque l'incertitude *bottom-up* est élevée, indice d'une intégration optimale d'information dans le processus de prise de décision (Summerfield & de Lange, 2014), où lorsque l'évidence sensorielle est difficile à extraire, les attentes ont plus de poids de manière à optimiser la durée de la prise de décision.

Par ailleurs, examiner le temps d'initiation et les profils de vélocité des mouvements de réponse permet d'évaluer dans quelle mesure le contexte provoque une anticipation ou une précipitation des réponses (forte déviation, temps d'initiation courts et augmentation précoce de la vélocité) et dans quelle mesure le bruit retarde les réponses. Comme décrit dans le Chapitre 2 (Section 2.3.1), la mise en relation de

Chapitre 3 – Études expérimentales

ces mesures permet de rendre compte du CVE de manière plus fine, l'erreur de réponse (la déviation de la souris par rapport à une trajectoire idéale) pouvant être mesurée pour chaque essai (en même temps que différentes mesures temporelles). Par contraste, les tâches traditionnelles doivent adapter finement leur difficulté pour obtenir un nombre d'erreurs suffisant. Ainsi, une condition dans laquelle les réponses sont biaisées vers la vitesse devraient provoquer une déviation plus élevée, des temps de réponse plus courts et des mouvements commençant plus tôt.

Dans cette expérience, tenir compte du CVE permettait donc d'évaluer la modulation de l'accumulation d'évidence, et distinguer ainsi des réponses facilitées/gênées par les attentes ou le bruit, de réponses biaisées vers la vitesse ou l'exactitude. Il nous est difficile de formuler des hypothèses *a priori* claires, cet aspect de l'étude ayant été ajouté tardivement. Dans une version antérieure de cet article n'incluant alors que les Expériences 1 et 2, et sans analyses de vélocité, un reviewer d'*Acta Psychologica* a pointé une variation du CVE entre les conditions avec et sans contexte. La présence d'un contexte (congruent comme incongruent) provoquait un biais du CVE vers la vitesse (nous relevions alors ce résultat mais parlions de « perturbation par la présence du contexte »). Cette interprétation en terme de CVE a donc été faite *a posteriori* pour les études 1 et 2, mais a été confirmée par la suite, lors de la révision de l'article, avec l'ajout de l'Expérience 3 (visant notamment à répliquer cet effet ainsi que l'effet de congruence) et les analyses de vélocité effectuées sur les trois expériences.

Dans l'Expérience 1, le contexte était amorcé avant l'apparition de la cible. Dans l'Expérience 2, le contexte était présenté seul, puis la cible apparaissait au centre après un court délai. La séquence de stimuli proposée dans l'Expérience 2 était, par rapport à l'Expérience 1, plus fluide et évitait la demande implicite d'attendre que la cible apparaisse pour démarrer le mouvement de réponse. Cette manipulation permettait d'éviter la discontinuité dans la présentation du contexte et de la cible. Cependant, ces deux protocoles ne permettaient pas de déterminer si un biais des trajectoires en direction de la réponse congruente au contexte, était dû à un biais de la reconnaissance de la cible (cible plus facile à reconnaître si contexte congruent présenté) ou au fait que le contexte provoque un biais moteur direct vers la réponse qui lui est reliée. Ainsi, dans l'Expérience 3, le contexte était présenté avant que les participants appuient sur le bouton *start*, l'excluant ainsi de la séquence du mouvement de réponse. Cette manipulation, demandée par un reviewer d'*Acta Psychologica*, devait permettre de faire disparaître l'effet de congruence du contexte si ce biais est d'origine motrice (biais de réponse), mais le maintenir si le biais est représentationnel.

Article 1

Étude 1, expériences 1-3

**Tracking the time course of perceptual decisions reveals the effects of contextual and noisy
sensory information on the speed and accuracy of the responses**

Boris Quétard,

Jean-Charles Quinton,

Laura Barca,

Giovanni Pezzulo,

Marie Izaute,

Michèle Colomb,

Martial Mermillod,

Journal: Acta Psychologica

Statut: Version révisée (soumise)

Tracking the time course of perceptual decisions reveals the effects of contextual and noisy sensory information on the speed and accuracy of the responses

Boris Quétard ^{a *},

Jean-Charles Quinton ^b,

Laura Barca ^c,

Giovanni Pezzulo ^c,

Marie Izaute ^a,

Michèle Colomb ^d,

Martial Mermillod ^e,

^a Université Clermont Auvergne, CNRS, LAPSCO, F-63037 Clermont-Ferrand, France

^b Université Grenoble Alpes, CNRS, LJK, F-38000 Grenoble, France

^c Institute of Cognitive Science and Technologies, CNR, Rome, Italy

^d CEREMA, Département Laboratoire de Clermont-Ferrand, F-63017 Clermont-Ferrand, France

^e Université Grenoble Alpes, CNRS, LPNC, F-38040, Grenoble, France

*Corresponding author. LAPSCO 34 avenue Carnot 63037 Clermont Ferrand Cedex1. E-mail:

boris.quetard@uca.fr.

Secondary contacts: G. Pezzulo: giovanni.pezzulo@isct.cnr.it; M. Izaute: marie.izaute@uca.fr.

Abstract

We explored the integration of bottom-up noisy sensory information and (top-down) contextual information during a perceptual decision-making task. In a dual forced-choice task participants had to judge whether a visual target stimulus (an animal or a tool/vehicle) which was altered by varying levels of noise, was a living or a non-living entity. The contextual prior information was manipulated by presenting a context shortly before the target (Experiment 1), over the target (Experiment 2), or in a separate phase in which participants controlled the context offset and target onset by pressing a *start* button (Experiment 3). The context image was grey (no context), congruent or incongruent with the target. To study the on-line dynamics of the decision process, we tracked the participants' mouse movements during the choice. We found evidence of anticipatory influences in the processing of both congruent and incongruent contextual information, which both biased the speed-accuracy trade-off towards speed and were moderated by noise levels. Furthermore, mouse trajectories were biased towards responses congruent with the context when context and target were in the same stimulus sequence (Experiments 1 and 2). Overall, the results of these experiments contribute to shed light on the competition between choice alternatives, highlighting the permissive role of contextual information.

Keywords: Perceptual decision-making, Categorization, Contextual information, Noise, Speed accuracy trade-off, Mouse-tracking

1. Introduction

The identification of objects or features can be conceptualized as a decision process that considers all the available information. For instance, recognizing an apple requires to integrate visual information and extract diagnostic features such as shape or colour, allowing to discard possible alternatives (*e.g.* resembling objects such as a balloon or a tomato) and ultimately decide that the percept represents an apple. Diverse visual features are integrated in parallel but not over the same time course and thus, visual information is progressively processed and accumulated (Bar, 2009; Greene & Oliva, 2009; Hegdé, 2008). In keeping with a large body of literature, the decision process can be modelled as a stochastic accumulation of evidence at a certain rate (*i.e.* drift rate), in favour of different candidate representations up to a threshold where the decision is made, *i.e.* *accumulation-to-bound* (Churchland, Kiani, & Shadlen, 2008; Gold & Shadlen, 2007; Ratcliff & McKoon, 2008; Ratcliff & Rouder, 1998; Ratcliff, Smith, Brown, & McKoon, 2016). The rate of the evidence accumulation depends on the strength of the integrated evidence (Drugowitsch, Moreno-Bote, Churchland, Shadlen, & Pouget, 2012; Palmer, Huk, & Shadlen, 2005). Thus, continuous integration of information about the target in a bottom-up manner within a decision making process may lead, for instance, to the recognition of objects.

However, to apply the *accumulation-to-bound* framework to object recognition, it is necessary to consider that evidence accumulation and the subsequent decision-making process are not only based on a passive bottom-up extraction of visual information. They also involve active strategies for accumulating evidence. Thus, the active visual exploration of the stimulus allows the system to gather evidence from different locations within the stimulus. Eye fixations allow to verify hypotheses about the stimulus' identity and thus, disambiguate it up to its final recognition (Donnarumma, Costantini, Ambrosini, Friston, & Pezzulo, 2017; Friston, Adams, Perrinet, & Breakspear, 2012; Kietzmann, Geuter, & König, 2011; Kietzmann & König, 2015; Quinton, Volpi, Barca, & Pezzulo, 2014).

In perceptual decisions that do not require performing eye movements (such as identifying a stimulus of small size), a marker of active evidence accumulation can be assessed through the Speed-Accuracy Trade-Off (SAT) (Bogacz, Wagenmakers, Forstmann, & Nieuwenhuis, 2010; Heitz, 2014; Heitz & Schall, 2012, 2013). In these tasks, when evidence is easily accessible, both speed and accuracy are increased (facilitation effect) and conversely, task difficulty and stimulus uncertainty tend to decrease them (impairment effect) (Wickelgren, 1977). Depending on the tasks goals, trading speed for accuracy allows to mitigate the effects of task difficulty and stimulus uncertainty (Bogacz, Brown, Moehlis, Holmes, & Cohen, 2006; Drugowitsch et al., 2012; Wickelgren, 1977). Biasing the SAT towards accuracy increases the time dedicated to accumulate more evidence, thus increasing the probability to make an accurate decision, and

Article 1 (Étude 1) – Contexte et bruit dans le déroulement temporel des décisions perceptuelles

increasing the response time. Conversely, biasing the SAT towards speed precipitates the response (*e.g.* when not enough time is available or when one wants to be quick). Consequently, the probability to take an erroneous decision increases since a sufficient amount of evidence could not be accumulated in time.

This active accumulation of evidence can be led by prior knowledge about the environment and the stimuli composing it (Gilbert & Li, 2013). Thus, prior knowledge about the stimulus can enhance the decision-making process since the beginning of the visual accumulation process by biasing it. Such influence can be due to the expectations one can have about which object to attend to, or which response will likely be the correct one. It is usually modelled as an increase of the evidence accumulation baseline of the expected response (alternatively, an increase of the drift rate of the expected response) (Summerfield & de Lange, 2014; Summerfield & Egnér, 2009).

A type of prior information that has been the object of considerable study is the link between a target object and the context in which it can be found. In real-world scenes, objects are not distributed randomly and relationships among objects and between these objects and their specific contexts can be learned. This knowledge can be used as an additional source of information in perceptual decision-making processes. It has been shown that context can be highly predictive of a given object or class of objects (Bar, 2009; Fenske, Aminoff, Gronau, & Bar, 2006; Kveraga et al., 2011) and helps the visual search of objects by guiding gaze to attended locations (Biederman, 1972; Brockmole, Castelhano, & Henderson, 2006; Brockmole & Võ, 2010; Malcolm & Henderson, 2010; Spotorno, Malcolm, & Tatler, 2014).

A small number of studies have directly addressed the effect of target-context consistency on identification by presenting the target objects in coherent or jumbled scenes (Biederman, Rabinowitz, Glass, & Stacy, 1974), among congruent or incongruent objects (Auckland, Cave, & Donnelly, 2007), in the foreground of a congruent or incongruent visual scene (Davenport & Potter, 2004) or preceded by an appropriate or inappropriate scene (Palmer, 1975). In these experiments, there was no probabilistic cuing, as incongruent trials were as frequent as congruent trials. In such case, the facilitation of target recognition by congruent contexts may be interpreted as the integration of contextual prior as an additional source of information inducing expectations about the target by activating representations which are both congruent with the context and linked to the target object. Additionally, Hollingworth and Henderson (1998, 1999) showed that context can directly bias the response of the identification task without actually facilitating target recognition. Therefore, the context can provoke the early activation of representations which can facilitate target perception and recognition (Auckland et al., 2007; Biederman et al., 1974; Davenport & Potter, 2004), and/or bias the decision towards the response congruent with the context (Hollingworth & Henderson, 1998, 1999).

These influences of contextual information on the decision-making process raise questions about

Article 1 (Étude 1) – Contexte et bruit dans le déroulement temporel des décisions perceptuelles

how context is integrated in the process of accumulating evidence and how this interacts with the continuous accumulation of bottom-up evidence. To investigate this question, we focused on the combined effects of bottom-up uncertainty, which modulates the sensory evidence strength and thus its accumulation, and contextual prior information on the decision making process. One major paradigm traditionally used to study perceptual decision making is a forced-choice stimulus discrimination task. In this task, the participant is asked to choose between alternatives and uncertainty about the correct response due, for example, to a reduction in stimulus strength, increases both error rates and response times (Palmer, Huk, & Shadlen, 2005). The main limitation of this kind of tasks is that it provides measures which are collected after the decision has been made, and therefore poorly account for the continuous nature of the accumulation process underlying decision-making (*i.e.* accuracy is a binary variable and time is the duration of the overall process).

A recent methodology allows to measure the effect of uncertainty or ambiguity on the perceptual decision-making components of a forced-choice categorization task in a more continuous way, by tracking the hand movements (Buc Calderon, Verguts, & Gevers, 2015; Resulaj, Kiani, Wolpert, & Shadlen, 2009; Song & Nakayama, 2009) or mouse movements (Barca & Pezzulo, 2015; Dale, Kehoe, & Spivey, 2007; Freeman & Ambady, 2010; Spivey, Grosjean, & Knoblich, 2005; Wang, Taylor, & Brunyé, 2012) while reaching the response. The continuous nature of the response movement occurring during the perception of the stimulus allows to infer components of the decision-making process underpinning the response. In such task, the response buttons are placed on the top corners of the screen and the deviation of the mouse trajectory towards the incorrect response as well as its time course, can reflect the competition between the responses (Buc Calderon et al., 2015; Hehman, Stolier, & Freeman, 2015). Thus, ambiguity or uncertainty relative to the identity or category of the target stimulus should increase this deviation (Dale et al., 2007).

Interestingly, this method also permits to study in more details the influence of the conditions on the SAT. Compared with traditional reaction times tasks, mouse-tracking allows to estimate the decision error for each trials, whether they are correct or erroneous although the error is not based on the final decision. Thus, speed can be evaluated through time measures such as movement initiation times (IT), global response times (RT) and velocity profiles. On the other hand, accuracy can be evaluated through the mouse deviation from the ideal trajectory as it can reflect erroneous first decisions or uncertain decisions. This is made possible by the fact that in this kind of task, participants can quickly initiate a response while not being certain of the correct response, or based on early guesses, directing the mouse towards the centre of the screen or towards the guessed response (Buc Calderon et al., 2015).

In the present study, participants had to judge whether the target, an animal or a tool (vehicle in Experiment 3), was *living* or *non-living*. We manipulated contextual information by priming the targets with *congruent*, *incongruent* or *control* (grey) context images –with congruent images depicting a scene in which the target could be found (natural scene representing the habitat of the animal, indoor scene where the tool can usually be found), and incongruent images depicting a scene from the opposite category. We made three main hypotheses.

First, congruent contexts should facilitate the decision-making process compared to incongruent contexts, as they prime both the target category and the response label. Hence, we should observe increased RT and increased trajectory deviation towards the incorrect response for incongruent contexts. Additionally, IT and mouse velocity profiles should indicate whether incongruent contexts delay the response. As competition between the response alternatives is expected to be higher, velocity should begin to increase later and the movement delayed (Hehman et al., 2015).

Second, examining these measures also allows us to evaluate how the SAT is adapted across context conditions, as it can be seen as a mechanism of active perception adapting the response to the available information (Chen & Perona, 2015; Heitz, 2014). In the absence of a context both RT and trajectory deviation should lie between the congruent and incongruent conditions as it provides no information on either responses. But it is also a condition where less information is available early, which should delay the moment at which the decision and its associated response movement start. Hence, the early availability of contextual information would precipitate the beginning of the accumulation of evidence. In this case, we should observe a SAT biased towards accuracy for no-context conditions (*i.e.* delayed choice and smaller deviation), and biased towards speed for context conditions.. Contextual information, whether it is congruent or incongruent, would be considered as additional evidence which need to be accumulated in a limited time.

Third, we manipulated the uncertainty of bottom-up information by adding noise in the target pictures, thereby reducing the strength of the available information (low signal/noise ratio). We hypothesized that increased uncertainty would increase the RT, the IT and the trajectory deviations as well as reduce the overall mouse velocity, because in the presence of uncertainty, separating the targets from noise and accumulating evidence require more time (Drugowitsch et al., 2012; Gold & Shadlen, 2001; Palmer et al., 2005; Röhrbein, Goddard, Schneider, James, & Guo, 2015). In accordance with predictive coding schemes in which the integration of information depends on the relative uncertainty of the visual bottom-up evidence and top-down priors, we expected the effect of contextual information to be bigger when the visual information is highly degraded by noise (Friston, 2005; Summerfield & de Lange, 2014). This uncertainty-based modulation would allow for expectations to compensate for the uncertainty of a noisy

environment, by increasing the weight (gain) of anticipations generated by prior knowledge of the context in target category identification. In our experiment, the facilitation effect by a congruent prime (compared to an incongruent one) should be bigger with higher levels of noise degrading the target.

2. Experiment 1

2.1. Method

2.1.1. Participants

Forty-five students (31 women, mean age = 20.22, SD = 2.3) participated in the experiment. All of them were first-year psychology students (n = 21) and second-year psychology students (n = 24) at Clermont Auvergne University in Clermont-Ferrand, France. Their vision was either normal or corrected-to-normal (glasses or contact lenses) and they were asked to respond with their hand of preference when using a computer mouse. They received course credits as compensation.

The participants completed a consent form before beginning the experiment. This study was approved by the French Committee for the Welfare of Persons “South-East VI”.

2.1.2. Stimuli

We built two sets of stimuli: one set of targets and one set of contextual primes. The target images were taken from the POPORO database (Kovalenko, Chaumon, & Busch, 2012): We chose 22 wild animals and 22 tools which can be found around the house including in the bathroom, kitchen, living-room, bedroom and office. All the target pictures had a 400 x 400 pixel resolution and the non-object parts of the images were filled with a white background. The contextual primes were 44 real-world visual images taken from the SUN database (Xiao, Hays, Ehinger, Oliva, & Torralba, 2010). We adapted the size of the images to a resolution of 762 x 425 pixels by resizing and cropping them. Each of the images depicted a scene in which it is common to find one of the targets. For animals, the congruent contextual images consisted of a scene of their habitual environment (e.g. ice field scene for a polar bear) whereas for tools, the congruent images depicted the room where the tool in question is commonly found (e.g. an office scene for a pen).

For each target, two congruent context images were chosen and we performed a pretest in order to select the better of these. To associate each target with non-congruent context scenes, we formed 22 pairs of targets including an animal and a tool. Therefore, the congruent context scenes for a target animal were also the non-congruent context scenes for a target tool and *vice versa* (see Figure 1.A.). We avoided forming pairs of related targets. A relationship among targets or between target and context could indeed exist at the semantic or linguistic levels. We therefore pre-tested this material in order to verify the association

Article 1 (Étude 1) – Contexte et bruit dans le déroulé temporel des décisions perceptuelles

between the target and context images, and asked 21 students (first and second year in psychology) to rate whether the target image was related to the context image presented next to it on a continuous scale (0 to 100). We used the E-Prime 2 software suite to perform this pretest. Each target was presented four times (with its 2 congruent and 2 incongruent contexts) and the positions of the target and context images were counterbalanced across subjects. Target-context pairs were randomized.

We selected the best context for each target by minimizing the outcome of a function averaging the context's association score with its incongruent target and 100 minus the association score of its congruent target. At this point, each target of a category was paired with a target of the other category and one congruent context for each target remained. To ensure that all the targets were correctly linked with their congruent and non-congruent contexts, we used the above function to eliminate the target pairs with poor target-context associations (highest scores on the function). Three target pairs were removed using this method. The minimum association score for congruent associations included in the final stimulus set was 86.47 out of 100 and the maximum association score for incongruent associations was 7.42 out of 100.

To manipulate bottom-up uncertainty, we applied two levels of additive noise to the target pictures. The noise distributions were uniform across pixels (covering the entire target), and Gaussian for each pixel colour component (RGB). The zero mean Gaussian distributions used to alter the intensity of pixels had a variance of 0.6 and 2 for the low and high-noise targets, respectively. We excluded two additional target pairs because one of the targets was not recognizable after adding noise to the image. After applying this procedure, we retained 17 targets of each category together with the associated contexts. Finally, we built a uniform grey 762 x 425 pixels image (RGB value of 90) as a no-context prime.

2.1.3. Apparatus

We used Freeman's MouseTracker software to build and run the experiment. This allowed us to record and process mouse movements. We used a wired optical mouse. Stimuli were presented on a 17-inch Dell D1025HE display, with a resolution of 1024 x 768 pixels. The refresh rate was 60 Hz.

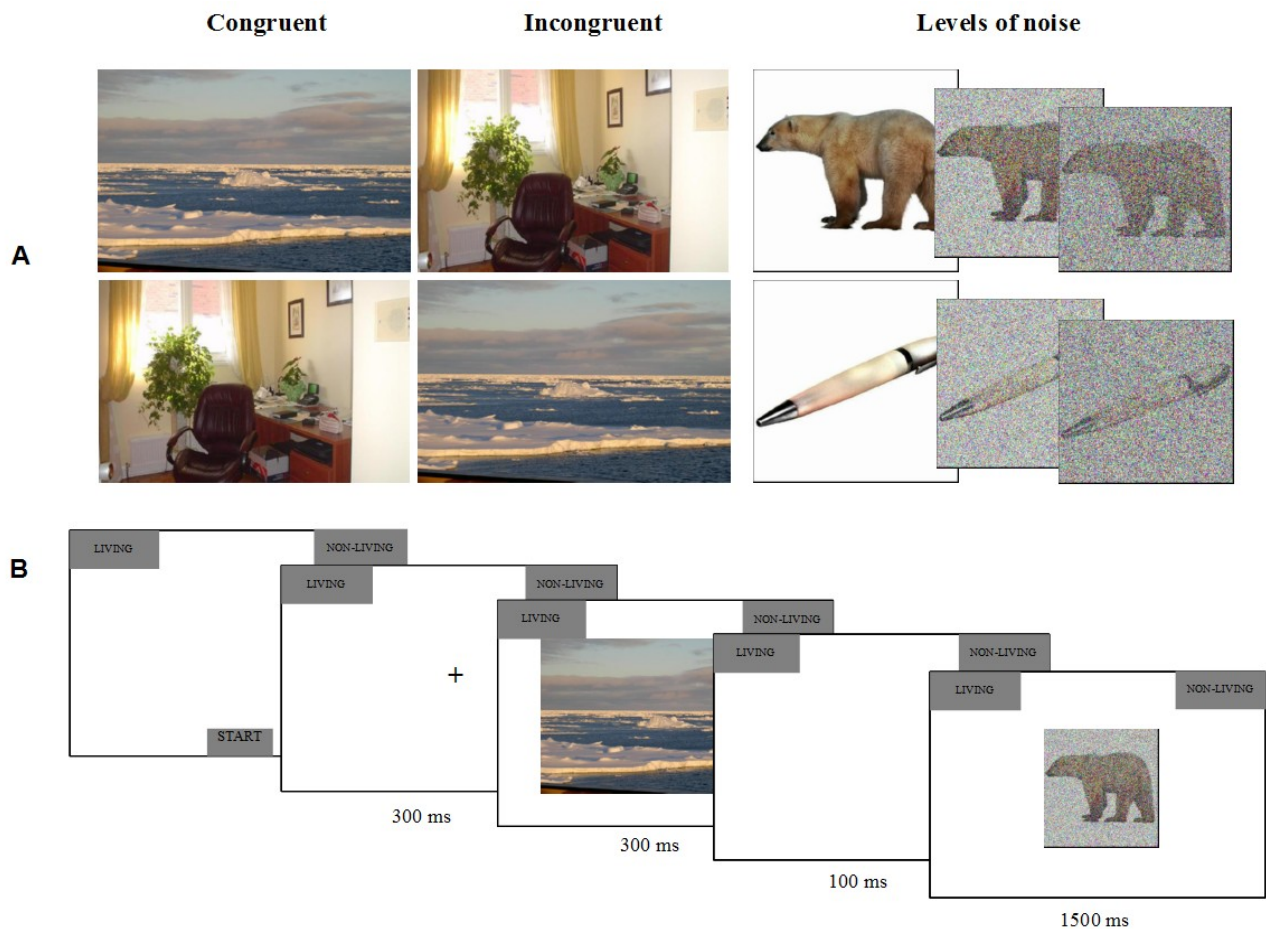


Figure 1: Overview of the stimuli and procedure of Experiment 1. Part A. shows an example association between an animal and a tool with their congruent and incongruent contexts, as well as their different levels of noise. Part B. shows the time course of the stimulus presentation

2.1.4. Procedure

Figure 1.B. shows the time course of the presentation of the stimuli. They had to click on a /START/ button centred at the bottom of the screen to begin each trial. This button and the two response buttons, located at the top-left and top-right of the screen, remained visible for the entire duration of the experiment. After the initial click on the /START/ button, a fixation cross was presented at the centre of the screen for 300 ms and the participant was required to fixate on this. Next, the prime image, which was congruent, incongruent or grey (no-context), appeared for 300 ms and was followed by a 100 ms blank screen interval. Finally, the target was presented and remained on the screen for 1500 ms. The target was an image of an animal or a tool presented with no noise, low noise or high noise. The participants were instructed to click on the /LIVING/ or /NON LIVING/ response button corresponding to the target's identity as quickly as possible. Participants had to begin their mouse movement as soon as possible. The trial ended

when the subject clicked on a response button or 1600 ms after onset of the target. At this point a /TIME OUT/ message informed the participant that the trial had ended with no response. If the participant gave an erroneous response, a red cross was displayed in both the practice and test blocks.

First, the participants performed 12 practice trials with stimuli which were not used in the test phase but were built using the same method. The test block included the 306 trials (34 targets \times 3 contexts \times 3 noise levels), which were displayed to the participant in a random order. Thus, each target was presented 9 times. Participants were invited to take two short breaks during the experiment, namely after 102 and 204 trials. Half of the participants were presented with /LIVING/ response buttons on the left and /NON LIVING/ response buttons on the right and the presentation was reversed for the other half.

2.2. Results

2.2.1. Data processing

In order to compare mouse trajectories, we applied standard pre-processing steps as described in the source article on the MouseTracker software (Freeman & Ambady, 2010). All the trajectories were remapped rightward and rescaled into a common coordinate system with the x axis ranging from -1 to 1 and the y axis ranging from 0 to 1.5.

We excluded trials where the /TIME OUT/ message was triggered (4.14% of total trials) and trials where the wrong response button was pushed (0.73% of total trials). As the proportion of error responses was very low, we did not analyse them.

To assess the effects of congruency and noise on response times and trajectories, we used Linear Mixed-Effects Models (LMM). This type of model allows us to test within-subjects designs by taking into account simultaneously both by-participants and by-targets variability. Hence, they allow to model the error generated by the effect of individual differences between participants and items on the measure relative to the experimental condition (Baayen, Davidson, & Bates, 2008). As the Noise and Congruency variables both had three levels, we recoded them. Although Noise could be considered as an ordinal variable, recoding it using a linear approximation (a single numeric variable such as no noise = -1, low noise = 0 and high noise = +1) was sufficient to correctly test our hypothesis while simplifying our statistical models (using one instead of two contrasts). For Congruency, we computed a set of two orthogonal contrasts. Congruency contrast 1 refers to the congruency effect and compared congruent and incongruent conditions (cong = -0.5, inc = +0.5, no-context = 0). Congruency contrast 2 refers to the context presence effect and compared the congruent and incongruent conditions on one hand, and the no-context condition on the other (cong = -0.25, incong = -0.25, no-context = +0.5). We considered Noise, Congruency contrasts 1 and 2 and the interactions between Noise and Congruency contrasts as fixed effects factors. We specified a maximal

random effect structure allowing the model to converge, as suggested by Barr, Levy, Scheepers, & Tily (2013). We therefore included variations in random intercepts, variations in random slopes (excluding the interactions) and allowed random correlations.

We resumed the mouse movements using 6 dependent variables based on the response times, cursor location and deviation and mouse velocity (presented in the following sections), each analysed with a separate model, and we analysed the velocity as a function of time across the conditions. Initially, we also included another dependent variable calculated by default by the MouseTracker software, the times of maximum deviation, but we discarded it; The reasons are that the results were redundant with the RT and more importantly, that the precise decision mechanisms connected to it are not clear in the mouse-tracking literature. We used the R's lme4 package (Bates, Maechler, & Bolker, 2012) to run the analyses and p-values were estimated using the Satterthwaite approximation. For each model, we removed the trials with a standardized residual greater than 3 standard deviations.

2.2.2. Time measures

Time measures include the responses times (RT) and movement initiation times (IT). The response times were defined as the time that elapsed from the appearance of the target to the subject's response. The initiation times were defined as the time elapsed from the appearance of the target to the moment when the mouse cursor leaved the area of the START button, as calculated by (Buc Calderon et al., 2015). Figure 2.A. shows the mean RT and IT as a function of noise level for congruent, incongruent and no-context trials. Analysis using LMM revealed a significant main effect of Noise level for both RT ($b = 26.69$, $SE = 2.53$, $t(34) = 10.57$, $p < .001$) and IT ($b = 12.56$, $SE = 1.81$, $t(40) = 6.95$, $p < .001$). As the levels of the Noise factor are converted into numeric values (from -1 to +1), our data seem consistent with a model showing an increase of RT as noise increases. The difference between congruent and incongruent conditions was not significant for both RT (Congruency contrast 1, $b = 4.56$, $SE = 4.37$, $t(44) = 1.04$, $p = .30$) and IT ($b = -0.39$, $SE = 3.40$, $t(34) = -0.12$, $p = .91$). We did not observe an interaction between this contrast and Noise level either (RT: $b = -3.04$, $SE = 3.28$, $t(12656) = -0.93$, $p = 0.35$; IT: $b = 1.04$, $SE = 3.09$, $t(12357) = 0.34$, $p = 0.74$). Finally, when a contextual prime was presented (either congruent or incongruent), the participants were faster than where they saw a grey, no-context prime. This was the case for the RT (Congruency contrast 2, $b = 41.14$, $SE = 5.75$, $t(45) = 7.16$, $p < .001$) and the IT ($b = 78.51$, $SE = 8.83$, $t(45) = 8.89$, $p < .001$). The interaction between Congruency contrast 2 and Noise level was significant for the RT ($b = -8.01$, $SE = 3.83$, $t(12663) = -2.09$, $p = .037$); The difference between trials with contextual primes and with no-context prime was diminished as noise increased. However, there was no interaction for the IT ($b = -2.15$, $SE = 3.62$, $t(12356) = -0.59$, $p = 0.55$).

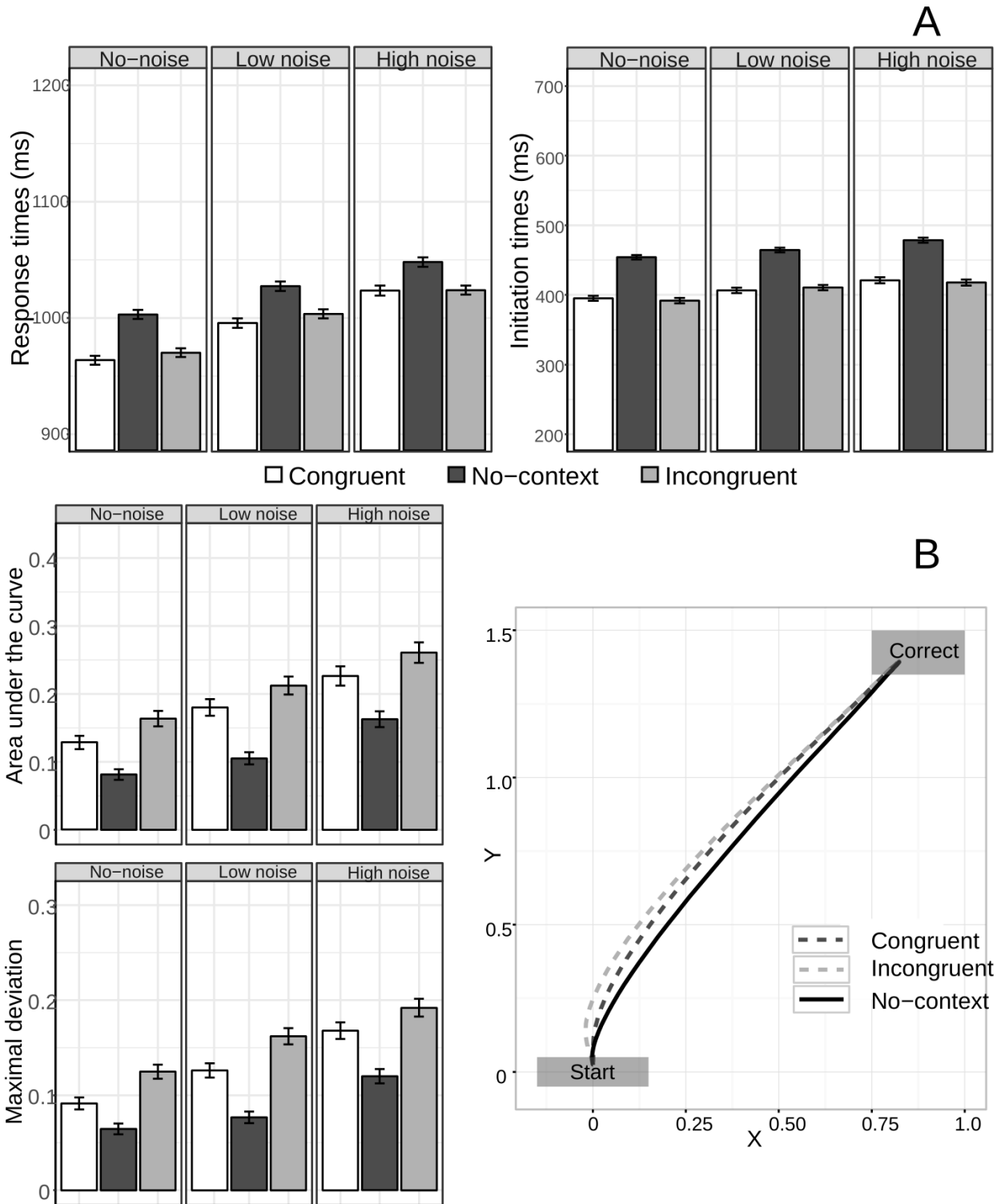


Figure 2: **A.** Response time as a function of context congruency and noise. **B.** Left: Area under the curve and maximal deviation as a function of context congruency and noise. White: Congruent context. Light grey: Incongruent context, Dark grey: No-context (grey background). Errors bars represent the standard error of the mean. Right: Mean mouse trajectory by context congruency conditions. Dark grey dashed curve: Congruent context, Light grey dashed curve: Incongruent context, black solid curve: No-context.

2.2.3. Analysis of trajectory deviation

Two trajectory deviation measures were analysed: The Area-Under-Curve (AUC) was defined as the geometric area between the ideal response curve and the response trajectory and the Maximal Deviation (MD) as the largest perpendicular deviation from the ideal trajectory. In order to compute the AUC, the trajectories were time-normalized into 101 points using linear interpolation. Figure 2.B. shows mean standardized AUC and MD by condition as well as mean trajectory by context congruency condition.

Noise level significantly increased the AUC ($b = 0.047$, $SE = 0.009$, $t(48) = 5.48$, $p < .001$) and MD ($b = 0.033$, $SE = 0.006$, $t(44) = 5.74$, $p < .001$) of the trajectories. Both Congruency contrasts were significant for each measure. The AUC and MD of trajectories were higher for incongruent context primes than for congruent contexts (AUC: $b = 0.036$, $SE = 0.018$, $t(45) = 2.03$, $p = 0.048$; MD: $b = 0.032$, $SE = 0.012$, $t(46) = 2.74$, $p = .009$). Compared to no-context conditions, context conditions had significantly higher AUC ($b = -0.109$, $SE = 0.017$, $t(49) = -6.24$, $p < .001$) and MD ($b = -0.073$, $SE = 0.011$, $t(46) = -6.84$, $p < .001$). None of the interactions between Noise level and any Congruency contrast was significant (all $|t| < 2$, $p > .05$).

2.2.4. Analysis of velocity

For each trajectory, we calculated the velocity at each 16 ms time-step (60 Hz temporal resolution) and smoothed the velocity curves by using a convolution with a triangular kernel over a sliding window of 21 time-steps. This smoothing step is required by the irregularity of mouse event sampling on most computer devices. The maximal velocity peak was detected for each trajectory and we analysed its amplitude (Maximal Velocity MV) and the time when it occurred (MV-time) using LMM with the same method we applied for RT, AUC and MD.

Figure 3.A. displays these measures as a function of Noise and Context congruency. Only Noise had an effect on MV; As the noise level increased, MV was reduced ($b = -0.061$, $SE = 0.021$, $t(34) = -5.43$, $p < .001$) which may indicate uncertainty about the decision. There was no significant difference between congruent and incongruent contexts ($b = 0.033$, $SE = 0.021$, $t(112) = 1.53$, $p = .128$) nor between the context conditions and the no-context condition ($b = 0.054$, $SE = 0.03$, $t(45) = 1.76$, $p = .085$). None of the interactions were significant either ($|t| < 1$, $p > 0.05$). The latency of the maximal velocity peak significantly increased with Noise level ($b = 22.54$, $SE = 2.07$, $t(35) = 10.88$, $p < .001$). We also observed that MV-time was delayed by incongruent primes compared to congruent primes ($b = 10.05$, $SE = 3.19$, $t(45) = 3.15$, $p = .003$) and this effect did not interact with noise ($b = -2.14$, $SE = 2.18$, $t(12524) = -0.98$, $p = 0.327$). The MV-time was also delayed by no-context primes compared to contextual primes ($b = 41.81$, $SE = 4.41$, $t(44) = 9.47$, $p < .001$) and this interacted with Noise level ($b = -6.52$, $SE = 2.55$, $t(12536) = -2.56$, $p = .011$); Thus MV-time occurred sooner for contextual primes but this effect slightly decreased when adding noise. However, it is

unclear whether the delayed MV-time is due to the global movement being delayed when presenting no-context or to a distortion of the movement dynamics towards longer time taken to get to the maximal speed.

In order to distinguish these two explanations, we examined the global velocity profiles of each condition. Trajectories finishing before the maximal duration were prolonged with 0 velocity values to ensure that all trajectories had the same number of time-steps. We estimated the velocity profiles for each experimental condition using LMM. For each time-step, we computed one model per noise condition, recoding the Context congruency variable with dummy coding (control condition: no-context). Each of the models included the recoded variable as fixed parameters and the variance of the intercept were included as random effects for both participants and targets grouping variables. For each time-step, we estimated the velocity per condition based on the fixed parameters. Figure 3.B. displays the velocity profiles obtained. This method allows to compute more accurate estimates of the velocity per conditions compared to calculating the means by time-step since it takes into account the variability across both participants and targets. It also provides a direct comparison of the no-context condition with the congruent and incongruent conditions at each time-step, for each noise level (we set the threshold for significant difference at $|t| > 2$, which corresponds to a slightly more conservative threshold than $\alpha = 0.05$).

Whatever the noise condition, we found a significantly higher velocity for congruent and incongruent contexts compared to no-context for the earliest time-steps, from target onset to 624-672 ms, depending on experimental conditions. In the latest time steps this difference was reversed; from 672-736 ms to 1136-1488 ms. This pattern is consistent with a delay of the global response movement when presenting no-context compared to congruent and incongruent contexts. We further estimated whether there is a global delay using cross-correlation, which allows to estimate the lag (with steps of 16 ms) for which two velocity time series correlate best. All correlations were superior to .98 at the optimal lag (where the curves were the most correlated). No-context velocity profiles were delayed by a 16 ms step (highest correlation) compared to congruent contexts (except without noise where it was delayed by 32 ms); They were also delayed by a 16 ms step compared to incongruent contexts. Between congruent and incongruent contexts, the correlation was the highest with no lag.

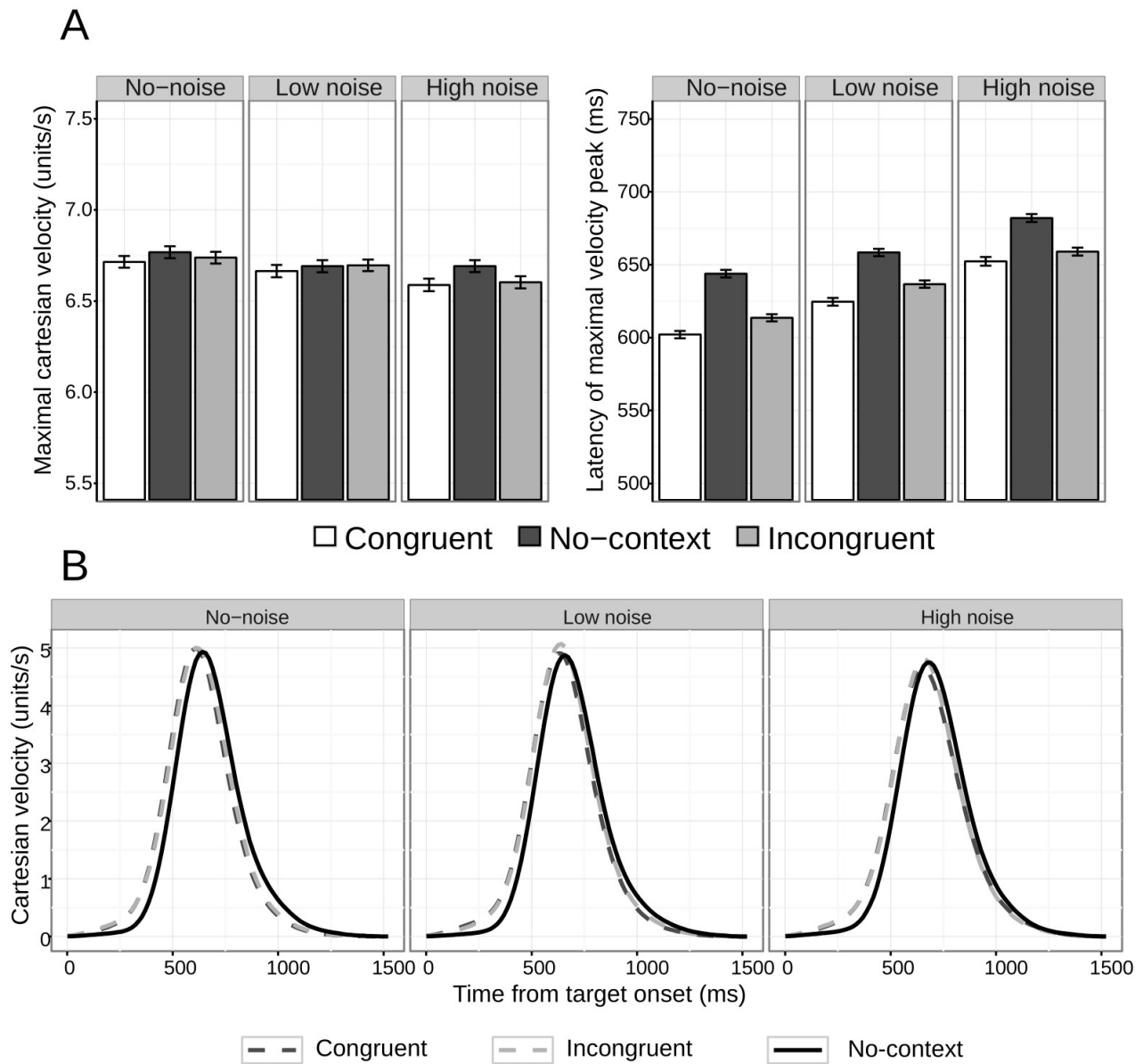


Figure 3: **A.** Maximal cartesian velocity and Latency of the maximal velocity peak as a function of context congruency and noise. White: Congruent context. Light grey: Incongruent context, Dark grey: No-context (grey background). Errors bars represent the standard error of the mean. **B.** Cartesian velocity as a function of time. Dark grey dashed curve: Congruent context, Light grey dashed curve: Incongruent context, black solid curve: No-context. The velocity for each noise condition and at each time-step (16 ms intervals) was estimated using separate LMM. Context congruency was recoded as dummy coding with the No-context condition being the control; Variations of intercepts across participants and targets were specified as random effects.

2.3. Discussion

For all measures, we found an effect of Noise consistent with an impairment of recognition when the noise level increased. As expected, in the presence of noise, RT and IT increased, response movements were globally delayed and their velocity was slower, trajectories deviated more towards the incorrect response button. Manipulating the Context congruency had the expected effect on trajectories, with incongruent primes eliciting more attraction by the alternative response than congruent primes and delaying the velocity peak. However, no effect of this kind was observed on RT and IT, nor did we find a lag of the global movement for incongruent contexts. Furthermore, the no-context primes biased the SAT towards accuracy, while context primes biased it towards speed. In response to the no-context primes, the participants exhibited slower RT and higher IT, but straighter trajectories towards the correct response button, as well as delayed response movement with the mouse reaching its velocity peak later in the trial.

We did not find a significant interaction between noise and context congruency (contrast 1). However, it may be erroneous to draw conclusions about this absence of interaction at this point. We cannot conclude either, that the context prime had no influence on the time-related characteristics of mouse responses in this experiment, despite it affected the target selection and biased trajectories towards the related response. Some elements of our methodology could have perturbed the time-related aspects of the participants' responses. The delay between when the /START/ button was pressed and the onset of the target was quite long (700 ms), with the result that the movement was less likely to begin early in the decision-making process. Because of this, the participants were implicitly asked to wait for the target and ignore the prime. The preparation of the movement initiation prior target onset can induce such pattern of results, where the responses are initiated at similar times and executed within the same time frame but with a larger deviation and later velocity peak for incongruent contexts.

Another interesting point comes from the comparison between no-context primes and context primes. A difference in behaviour can be observed across these categories of primes. When no context was presented, the SAT was biased towards accuracy, when a context was presented, it was biased towards speed. This bias towards speed for context primes could be due to their semantic content, their meaningfulness, or their visual complexity (compared to grey pictures primes), that can be seen as additional information which can be accumulated. The early appearance of a complex task-related picture may have triggered the early initiation of the decision process (and accumulation of evidence) compared to trials in which no-context primes were presented.

It might not have been a straightforward task for the participants to have to make a decision on the second stimulus that was presented, especially if the first stimulus was itself semantically related to one of the response buttons. In Experiment 2, we tried to reduce this effect by diminishing the delay between the

Article 1 (Étude 1) – Contexte et bruit dans le déroulé temporel des décisions perceptuelles

beginning of a trial and the appearance of the target, making the transition between context and target smoother and by simultaneously showing the context and target. This should reduce the risk that the preparation of the movement would interfere with congruency effects (revealing these effects on time measures) and with context presence effects in the decision-making process while making the stimulus sequence smoother and more continuous, thus simpler to handle for the participants.

3. Experiment 2

3.1. Method

3.1.1. Participants

Thirty students (26 women, mean age = 19.55, $SD = 1.59$) participated in the experiment. They were first-year psychology students ($n = 5$) and second-year psychology students ($n = 25$) at Clermont Auvergne University in Clermont-Ferrand, France. Their vision was either normal or corrected-to-normal (glasses or contact lenses) and they were asked to respond with their hand of preference when using a computer mouse. They received course credits as compensation.

The participants completed a consent form before beginning the experiment. This study was approved by the French Committee for the Welfare of Persons “South-East VI”.

3.1.2. Stimuli

To create the stimuli for this experiment, we used the same targets and context images as in Experiment 1. We built the target images by cutting out the objects from the original POPORO images. These images were then blended in at the centre of the context images without modifying their original resolution. Finally, we applied a Gaussian noise function on the image. The Gaussian noise was uniformly applied over a 400 x 400 pixel window which covered the complete target object and the noise was progressively attenuated outside this window, leaving the sides of the final image free of noise.

Since the noise level used in Experiment 1 was too high for these new stimuli (the contours of the targets being more salient on a white background), we adapted it. We set the variance of the noise in the central window to 0.4 and 1 for the low and high-noise pictures, respectively. See Figure 4.A. for visual instances. Our complete stimulus set included 35 context pictures (17 natural contexts, 17 indoor contexts and 1 grey background context), thus resulting in 306 pictures when contexts, target objects and noise levels (none, low or high) were combined.

3.1.3. Apparatus

As in Experiment 1, we used MouseTracker software. The experiment was presented on a laptop with a 15.6 inches display and a resolution of 1600 x 900 pixels. The refresh rate was 60 Hz. We positioned the participant's chair at a viewing distance of approximately 57 cm and asked the participant to sit up straight against the back of the chair.

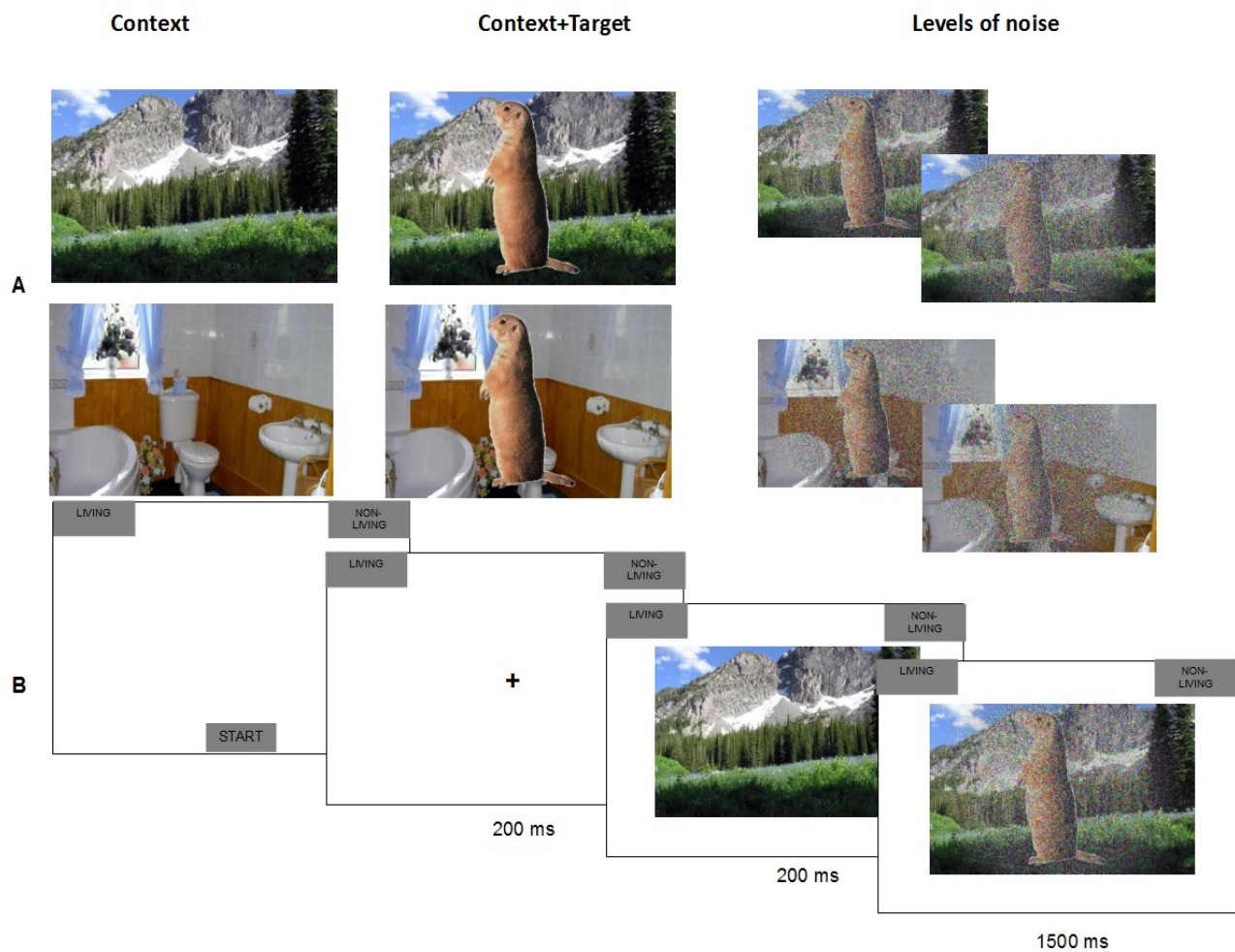


Figure 4: Overview of the stimuli and procedure of Experiment 2. Part A. shows examples of an animal blended into a congruent (top) and an incongruent context (bottom), with their different levels of noise. Part B. shows the time course of the stimulus presentation.

3.1.4. Procedure

The procedure was approximately the same as in Experiment 1. The main difference lays in the time course of the presentation of the stimuli (see Figure 4.B.). After the participants clicked on the /START/ button, the fixation cross was presented for 200 ms. Then, the context image appeared for 200 ms and was immediately followed by the stimulus containing both the context and the target, which was displayed for 1500 ms with one of the three levels of noise applied. Consequently, both the target and noise appeared to be present in the context. The /TIME OUT/ message appeared after an additional 1100 ms. We gave the same instructions to the participants as in Experiment 1 and also administered the same pauses. We counterbalanced the response buttons in the same way as in the previous experiment.

3.2. Results

3.2.1. Data processing

We used exactly the same procedure as in Experiment 1. This led us to exclude 0.84% of the trials because the /TIME OUT/ message was triggered and 0.73% for eliciting an incorrect response. As for Experiment 1, for the analysis of each measure, we excluded the trials whose residual in the corresponding LMM exceeded 3 standard deviations.

3.2.2. Time measures

Figure 5.A. shows mean RT and IT as a function of Context congruency and Noise level. Analyses with LMM showed that the RT and IT increased with Noise level (RT: $b = 34.69$, $SE = 4.15$, $t(38) = 8.36$, $p < .001$; IT: $b = 18.36$, $SE = 2.86$, $t(35) = 6.43$, $p < .001$). There was an effect of Congruency contrast 1 on both measures where the participants' RT and IT were shorter for congruent than incongruent contexts (RT: $b = 25.51$, $SE = 5.46$, $t(28) = 4.68$, $p < .001$; IT: $b = 11.57$, $SE = 4.05$, $t(29) = 2.86$, $p = .008$). As in Experiment 1, movements were initiated significantly earlier for context conditions (congruent and incongruent), than for no-context conditions ($b = 39.08$, $SE = 6.64$, $t(29) = 5.89$, $p < .001$). However, distinctively from Experiment 1, the effect of congruency contrast 2 was not significant for RT ($b = -7.1$, $SE = 8.55$, $t(37) = -0.83$, $p = .412$). When both contrast 1 and 2 are taken together, the results are consistent with the hypothesis of a gradual increase in RT between congruent, no-context and incongruent conditions. As our set of contrasts did not allow us to directly analyse the facilitatory effect of congruent context and the impairment effect of incongruent context compared to no-context, we performed an additional LMM analysis using dummy coding comparing the congruent and incongruent conditions to the control no-context condition. While there was no difference of RT between congruent contexts and no-context trials ($b = -7.43$, $SE = 6.94$, $t(35) = -1.07$, $p = .292$), incongruent contexts elicited larger RT than no-context ($b = 18.08$, $SE = 6.995$, $t(36) = 2.59$, $p = .014$).

Finally, as in Experiment 1, the Noise level by Congruency contrast 1 interaction was not significant for both RT ($b = -3.48$, $SE = 5.02$, $t(8906) = -0.69$, $p = .49$) and IT ($b = 3.20$, $SE = 4.20$, $t(8612) = 0.76$, $p = .45$). However, the Noise level by Congruency contrast 2 interaction was marginally significant for the RT ($b = -10.92$, $SE = 5.91$, $t(8906) = -1.88$, $p = .06$) and significant for the IT ($b = -10.02$, $SE = 4.86$, $t(8614) = -2.06$, $p = .0039$). Thus, in the presence of a context, participants precipitated their responses, but this effect diminished as noise level increased.

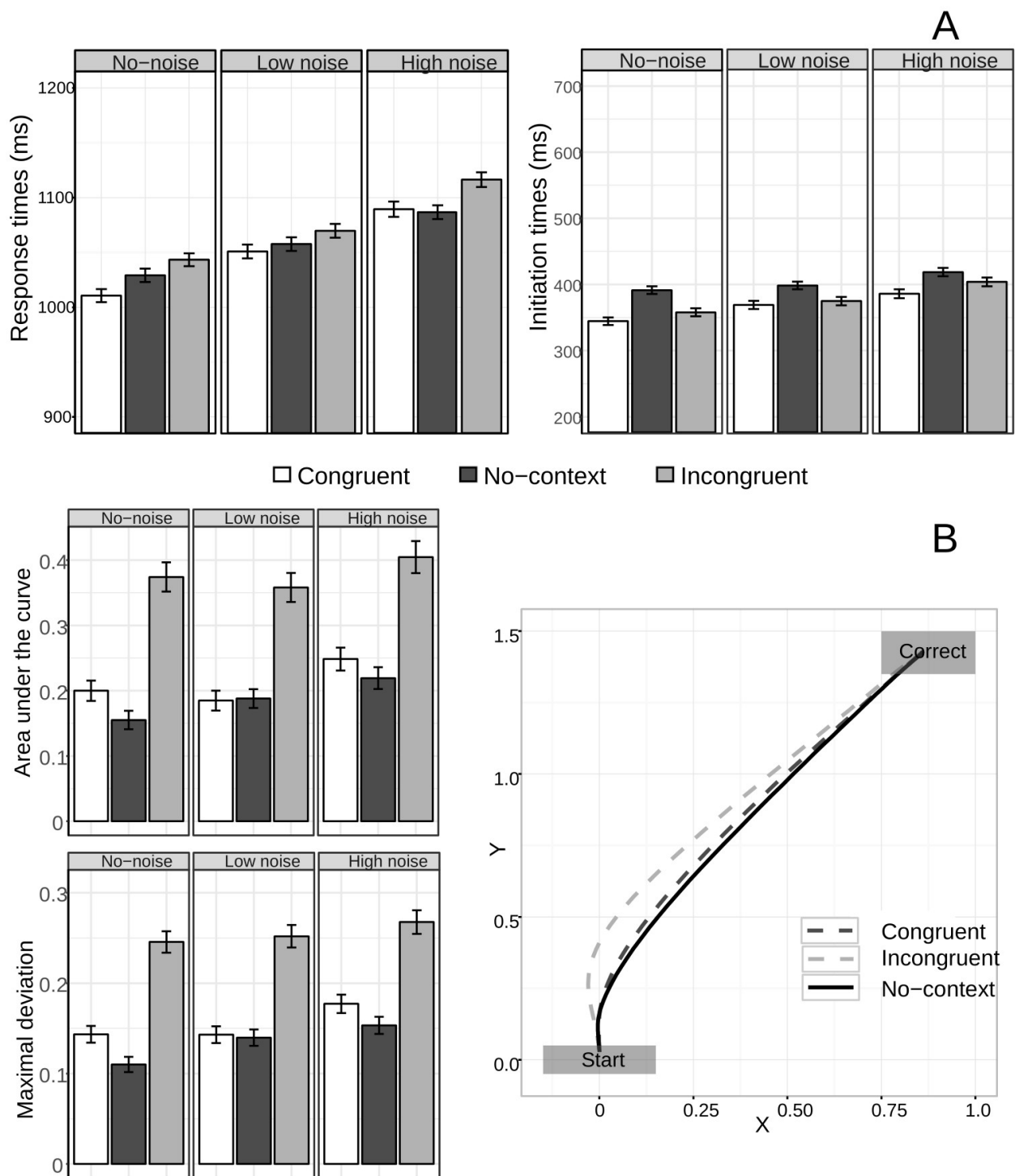


Figure 5: **A.** Response time as a function of context congruency and noise. White: Congruent context. Light grey: Incongruent context, Dark grey: No-context (grey background). Errors bars represent the standard error of the mean. **B.** Left: Area under the curve and maximal deviation as a function of context congruency and noise. Right: Mean mouse trajectory by context congruency conditions. Dark grey dashed curve: Congruent context, Light grey dashed curve: Incongruent context, black solid curve: No-context.

3.2.3. Analysis of trajectory deviation

Figures 5.B. indicate the mean AUC and MD in each condition, as well as mean trajectories by context congruency conditions. The analyses confirmed the effects of Noise level and Context congruency. Noise level led to both higher AUC ($b = 0.026$, $SE = 0.009$, $t(29) = 2.99$, $p = .006$) and higher MD ($b = 0.017$, $SE = 0.005$, $t(37) = 3.12$, $p = .003$). Congruent contexts elicited significantly reduced attraction by the alternative response on both these measures compared to incongruent contexts (AUC: $b = 0.172$, $SE = 0.035$, $t(47) = 4.85$, $p < .001$; MD: $b = 0.099$, $SE = 0.019$, $t(46) = 5.34$, $p < .001$). The effect of congruency contrast 2 was also significant (AUC: $b = -0.147$, $SE = 0.03$, $t(40) = -5.0$, $p < .001$; MD: $b = -0.094$, $SE = 0.017$, $t(37) = -5.6$, $p < .001$), showing that AUC and MD were smaller for the no-context conditions than the context conditions. This effect can primarily be explained in terms of reduced scores for the no-context compared to the incongruent conditions but since the current orthogonal contrast do not allow to verify that, we ran a LMM with dummy coding testing the differences between each context condition and the control no-context condition. Incongruent contexts led to a larger deviation than no-context both on AUC ($b = 0.197$, $SE = 0.037$, $t(46) = 5.31$, $p < .001$) and MD ($b = 0.12$, $SE = 0.025$, $t(46) = 5.95$, $p < .001$). There was no difference between congruent contexts and no-context trials for AUC ($b = 0.025$, $SE = 0.015$, $t(61) = 1.61$, $p = .113$) but the MD were larger for congruent contexts than no-contexts ($b = 0.021$, $SE = 0.009$, $t(28) = 2.29$, $p = .03$). Finally, there was no interaction of Noise level with any Context contrast (both $|t| < 2$, $p > .05$).

3.2.4. Analysis of velocity

To analyse velocity, we applied the same preprocessing steps, measured the same variables and fitted the same LMM than in Experiment 1. Figure 6.A. displays mean maximal velocity (MV) and times of maximal velocity (MV-time) by conditions.

The analysis of MV indicated an effect of the noise level, reducing the maximal movement velocity ($b = -0.068$, $SE = 0.018$, $t(28) = -3.72$, $p < .001$). Contrasting with Experiment 1, there was a significant difference between congruent and incongruent contexts ($b = 0.078$, $SE = 0.033$, $t(37) = 2.4$, $p = .022$) where speed was higher for incongruent contexts. No-context conditions also elicited movements with higher speed than context conditions ($b = 0.145$, $SE = 0.046$, $t(28) = 3.14$, $p = .004$). None of the interactions were significant ($|t| < 1$, $p > 0.05$). As for Experiment 1, the latency of the maximal velocity peak was significantly delayed with Noise level ($b = 29.32$, $SE = 2.83$, $t(38) = 10.36$, $p < .001$). We also observed that MV-time was increased by incongruent primes compared to congruent primes ($b = 32.94$, $SE = 3.44$, $t(26) = 9.57$, $p < .001$) and this effect did not interact with noise ($b = -1.11$, $SE = 3.14$, $t(8516) = -0.353$, $p = 0.724$). Thus, the delay for incongruent contexts was constant across noise levels. The MV-time was not significantly delayed by no-context conditions compared to contextual conditions ($b = 3.97$, $SE = 6.12$, $t(41) = 0.649$, $p = .52$). However

there was an interaction between the Congruency contrast 2 and Noise ($b = -16.22$, $SE = 3.62$, $t(8513) = -4.48$, $p < .001$); Hence, the maximal velocity peak was delayed by noise in a lesser degree for no-context condition than for contextual conditions.

We analysed the velocity profiles using the same method than for Experiment 1. Figure 6.B. displays the velocity profiles for each condition. In the early time-steps, there was a significantly higher velocity for congruent and incongruent contexts compared to no-context. This difference was significant from 80-192 ms to 464-544 ms after target onset, depending on the experimental conditions. The velocity was higher for no-context trials compared to both congruent contexts from 576-608 ms to 768-1072 ms. For no-context trials compared to incongruent contexts, the former increased the velocity from 544 to 912 ms when the target was not degraded by noise; However, when the target was degraded by noise, the velocity for no-context trials was significantly higher from 528 ms to 672-688 ms, but significantly lower from 752 ms to 1008-1136 ms. These results indicates that although the mouse movements began earlier for both congruent and incongruent contexts, the response movements were globally delayed for incongruent contexts when noise degraded the targets.

Further evidence can be found when calculating the cross-correlations between the estimated velocity profiles across conditions. All correlations were superior to .99 at the optimal lag values, where the curves were the most correlated. For targets non-degraded by noise, both incongruent and no-context velocity profiles were delayed by two 16 ms time-steps compared to congruent velocity profile. For targets degraded by low noise, the maximal correlation between congruent and incongruent profiles was observed when the incongruent profile was delayed by one 16 ms time-step; There was no lag between no-context profile and both congruent and incongruent profiles. Finally, for target degraded with high noise, the incongruent profile had a one time-step delay compared to both no-context and congruent profiles. Thus, movements were executed earlier when congruent contexts were presented but with increasing noise, the delay between no-contexts and congruent contexts was reduced while incongruent contexts increasingly delayed the movement compared to no-context. These results are consistent with the previous analysis of MT-time, where no-context trials were less affected by noise than trials where a context was presented.

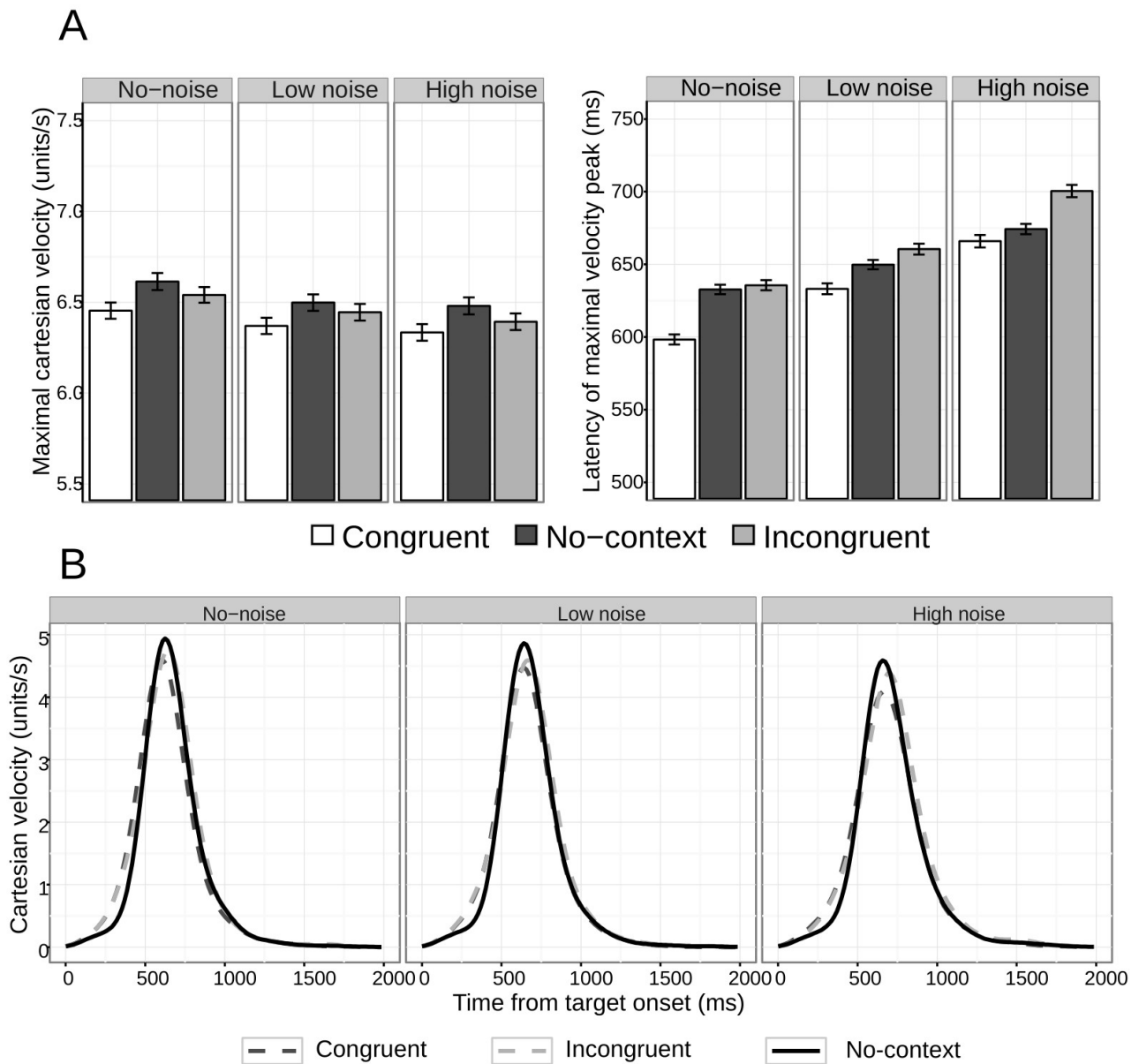


Figure 6: A. Maximal cartesian velocity and Latency of the maximal velocity peak as a function of context congruency and noise. White: Congruent context. Light grey: Incongruent context, Dark grey: No-context (grey background). Errors bars represent the standard error of the mean. **B.** Cartesian velocity as a function of time. Dark grey dashed curve: Congruent context, Light grey dashed curve: Incongruent context, black solid curve: No-context. The velocity for each noise condition and at each time-step (16 ms intervals) was estimated using separate LMM. Context congruency was recoded as dummy coding with the No-context condition being the control; Variations of intercepts across participants and targets were specified as random effects.

3.3. Discussion

As in Experiment 1, we found an effect of Noise level on all our measures. As expected, increasing the noise led to degraded performances in the perceptual decision-making task. Consequently, our manipulation of bottom-up uncertainty seems to be effective.

In Experiment 1, we did not observe any difference between congruent and incongruent context in movement initiation time and global response time. Furthermore, mouse movements were not delayed by incongruent contexts. In Experiment 2, we reduced the time that elapsed before the target appeared to 400 ms (instead of 700 ms in Experiment 1) and attenuated the sequential presentation of the contexts and targets, aiming to avoid interference of the movement preparation with the congruency effect. We expected a clearer effect of context congruency on temporal measures. In accordance with this hypothesis, we observed higher IT and RT as well as delayed mouse movements and MV-time for incongruent compared to congruent contexts. Additionally, we reproduced the congruency effects on AUC which we found in Experiment 1, with higher deviation towards the incorrect response for incongruent conditions.

Despite our manipulations were designed to prevent response preparation (in part by shortening the delay between context and target onsets), we replicated and extended the effect of the presence of a context reported in Experiment 1. Hence, no-context trials were biased towards accuracy whereas trials with a congruent or incongruent context were biased towards speed as deviation was lower, mouse movement began later and velocity began increasing later for no-context. This pattern of results could be interpreted as an anticipatory mechanism triggered by the availability of potentially helpful prior information (whether it leads to the correct or incorrect response). This availability of very early information would trigger the response process earlier during the decision-making process.

The putative anticipatory process did not interfere with the congruency effect. Hence, while the response movements began earlier when presenting contexts than no-context, the global RT, maximal velocity peak and the global velocity profile were still delayed for incongruent contexts compared to congruent contexts. In incongruent conditions, early movement initiation but delayed movement and longer response associated with a higher deviation towards the incorrect response highlight a stronger competition between the two alternative responses. For congruent contexts, results can be interpreted in terms of an early competition with movements beginning early and higher trajectory deviations than when there is no context (albeit lower than for incongruent contexts), triggered by the context early integration. This competition is resolved faster due to the correct initial assumptions given by the congruency of the context. For no-context conditions, since no information is available early, the competition would begin later and so would the response movement, allowing to accumulate more visual information on the target identity, and respond with a straighter mouse trajectory.

However, the same findings could be explained by an alternative hypothesis that appeals generically to the presentation of stimuli rather than specifically to the availability of relevant prior information. In this second experiment, we reduced the delay between context and target onsets to make the stimulus presentation smoother. This was made to prevent that the participants feel they have to wait for the target to appear and prepare their response during this time. But we did not completely remove this delay. This anticipation could be triggered not only by available contextual information (congruent or incongruent) but also by any elaborated stimulus giving a feeling of discontinuity and cluttering, due to the superposition of two distinct stimuli. This situation was not present in the no-context condition where only a uniform grey picture was presented. Additionally, the present design cannot disentangle two explanations of the context congruency effect, that is, context pre-activates representations congruent or incongruent with the target (Auckland et al., 2007; Bar, 2009; Biederman et al., 1974) or context biases the movement towards the congruent responses (*e.g.*, indoor context biases towards *non living*) (Hollingworth & Henderson, 1998, 1999).

One possibility to address both these questions is to shift the moment at which the /START/ button has to be pressed just before target appearance, thus changing the time-course such that pressing a /CONTEXT/ button starts the trial, triggering the presentation of the context image. Then, pressing the /START/ button triggers the onset of the target and participants can immediately begin their response. If congruent and incongruent contexts still bias the SAT towards speed, this should be due to the integration of additional prior information before /START/ button is pressed, instead of an effect of motor preparation while waiting for the target to appear since target onset is triggered voluntarily. If the congruency effect remains, it is likely due to the semantic influence of the context pre-activating the target category rather than a motor bias towards the context-congruent response, since the response movement is clearly separated from the context presentation by the /START/ button.

4. Experiment 3

4.1. Method

4.1.1. Participants

Thirty two students (27 women, mean age = 19.72, $SD = 1.17$) participated in the experiment. They were first-year psychology students ($n = 4$) and second-year psychology students ($n = 28$) at Clermont Auvergne University in Clermont-Ferrand, France. Their vision was either normal or corrected-to-normal (glasses or contact lenses) and they were instructed to use their hand of preference when using a computer mouse. They received course credits as compensation.

The participants completed a consent form before beginning the experiment. This study was approved by the French Committee for the Welfare of Persons “South-East VI”.

4.1.2. Stimuli

In this experiment, we modified half of the material. In order to decrease a potential effect of animate targets (*vs.* inanimate) on target recognition (Bonin, Gelin, Laroche, Méot, & Bugajska, 2015; Mermillod, Droit-Volet, Devaux, Schaefer, & Vermeulen, 2010), and to control for the indoor/outdoor characteristics of the context scenes, we used animals and vehicles (instead of tools) as target categories as they were conceptually closer (both able to move and found in outdoor contexts), although clearly *living* or *non-living*. Controlling for these allows for a better control of variability across target-context combinations and of the potential confound of the animacy effect (Bonin et al., 2015). Thus, their congruent contexts were man-made outdoor scenes in which the vehicle could be likely be found (*e.g.*, farm for a tractor, construction site for a digger...).

All targets had a size of 400 x 400 pixels and the contexts of 600 x 450 pixels. We selected 16 vehicle targets with their congruent context and paired each of them with an animal and his congruent context previously used in Experiment 1 and 2. Thus, as in the previous experiments, the vehicle's context is used as the incongruent context of the paired animal and *vice-versa*. We were especially careful that there was no semantic link between a target and its incongruent context when pairing them (*e.g.* we avoided pairing a swan with a tractor since the swan is linked with a farm, or a seagull with a boat).

We verified that the target were highly linked with their respective congruent context while being linked as weakly as possible with their incongruent context using a similar method as described for Experiment 1. Ten students rated the relation between each target and its congruent or incongruent context from 0 (not related) to 100 (completely related), presented side-by-side in a random order. To detect the pairs of targets which needed a rework (because of high link scores with incongruent contexts or low for

Article 1 (Étude 1) – Contexte et bruit dans le dérours temporel des décisions perceptuelles

congruent contexts), we averaged the link scores of the participants by targets and contexts and used a function maximizing high congruent link scores and low incongruent link scores for each target. Two pairs were suppressed and three pairs were modified by changing a problematic context or pairing them to different targets. We finally re-tested the three modified pairs to ensure the modifications improved their link scores. In the end of the procedure we had 14 targets of each category with their congruent and incongruent contexts. For the minimum link score for congruent contexts was 74 out of 100 and the maximal link score for incongruent contexts was 15 out of 100. The minimal score difference between congruent and incongruent for a target was 67.9.

To degrade the targets with low and high noise level, we used the same additive noise levels as in Experiment 1. The no-context prime was a uniform grey rectangle of 600 x 450 pixels (RGB: 90;90;90). Figure 7.A. displays examples of animals and vehicles with their congruent and incongruent contexts, at different noise level.



Figure 7: Overview of the stimuli and procedure of Experiment 3. Part A. shows an example of a vehicle target with its associated congruent and incongruent contexts, as well as its different levels of noise. Part B. shows the time course of the stimulus presentation.

4.1.3. Apparatus

Contrary to Experiments 1 and 2, we used a Java implementation of the mouse-tracking paradigm, since Freeman's MouseTracker software could not easily be exploited to design this experiment (due to the conditional appearance of stimuli and buttons, as well as user decided delays). The Java program was previously pretested against the MouseTracker software (Quinton et al., 2014) and later used for designing and running all experiments in (Smeding, Quinton, Lauer, Barca, & Pezzulo, 2016). Contrary to MouseTracker which relies on regular sampling of the mouse state, this program recorded all events generated by the mouse device (timestamps and cursor position) instead of the cursor position at a given frame-rate. The experiment was presented on a laptop with a 15.6-inch display and a resolution of 1680 x 1050 pixels. The refresh rate was 60 Hz. We positioned the participant's chair at a viewing distance of approximately 57 cm and asked the participant to sit up straight against the back of the chair.

4.1.4. Procedure

The main goals of this Experiment was to both remove any delay before target presentation, and to temporally separate the context and target presentations with a mouse click to display the target and start the response movement. To do so, we adapted the time course of the presentation of the stimuli (see Figure 7.B.). To begin each trial, the participants had to click on the /CONTEXT/ button localized at the bottom centre of the screen, triggering the presentation of the no-context image or a context image congruent or incongruent with the upcoming target. After a 400 ms delay ensuring a minimal presentation time, the /START/ button appeared where the /CONTEXT/ button was previously. This delay was chosen consistently with Experiments 1. Then, the context was maintained until the participants pressed the /START/ button; We used this extended context presentation time (> 400 ms) to ensure the context could not be easily ignored, and was properly integrated. As soon as this happened, the target picture was presented with one of the three levels of noise applied.

The participants were instructed to click on the response button as fast as possible and to begin their mouse movement as soon as possible. The /TIME OUT/ message appeared after a 2000 ms delay after target onset, ending the trial and the target was presented for this duration. A red cross indicated an erroneous response after immediately after the response. Before beginning the experiment, the participants had 15 training trials displaying different animals and vehicles than in the test session. The test block included 252 trials (28 targets × 3 contexts × 3 noise levels). During the test session, the participants were proposed two short breaks. We counterbalanced the response buttons in the same way as in the previous experiments.

4.2. Results

4.2.1. Data processing

We excluded 0.72% of the trials because the /TIME OUT/ message was triggered and 1.01% for eliciting an incorrect response. As for Experiments 1 and 2, for the analysis of each measure, we excluded the trials whose residual in the corresponding LMM exceeded 3 standard deviations.

4.2.2. Analysis of latencies

Figure 8.A. shows mean RT and IT as a function of Context congruency and Noise level. Analyses with LMM indicated that the Noise level significantly increased the RT ($b = 21.12$, $SE = 2.85$, $t(31) = 7.4$, $p < .001$) and the IT ($b = 7.78$, $SE = 1.91$, $t(30) = 4.08$, $p < .001$). There was no difference between congruent and incongruent contexts for the RT (contrast 1: $b = -3.04$, $SE = 4.90$, $t(22) = -0.62$, $p = .542$) and IT ($b = -0.29$, $SE = 3.92$, $t(28) = -0.07$, $p = .94$). However, compared with no-context conditions, context conditions induced lower RT (contrast 2: $b = 23.82$, $SE = 6.92$, $t(31) = 3.44$, $p = .002$) and earlier IT ($b = 39.84$, $SE = 5.09$, $t(28) = 7.83$, $p < .001$), consistently with the previous experiments. None of the interactions between congruency contrasts and noise were significant neither for the RT (contrast 1: $b = 6.41$, $SE = 4.72$, $t(7545) = 1.36$, $p = .175$; contrast 2: $b = -1.98$, $SE = 5.47$, $t(7542) = -0.36$, $p = .718$) nor for the IT (contrast 1: $b = 3.01$, $SE = 3.50$, $t(7558) = 0.86$, $p = .39$; contrast 2: $b = 2.06$, $SE = 4.06$, $t(7556) = 0.51$, $p = .61$).

4.2.3. Analysis of trajectory deviation

Figure 8.B. indicates the mean AUC and MD in each condition as well as the mean trajectories by context congruency condition. Analyses with LMM indicated that the Noise level significantly increased both the AUC ($b = 0.012$, $SE = 0.002$, $t(42) = 5.64$, $p < .001$) and the MD ($b = 0.026$, $SE = 0.005$, $t(40) = 5.64$, $p < .001$). There was no difference between congruent and incongruent contexts neither for AUC ($b = -0.007$, $SE = 0.005$, $t(27) = -1.32$, $p = .196$) nor for MD ($b = -0.014$, $SE = 0.010$, $t(28) = -1.46$, $p = .156$). Compared to context conditions, no-context condition induced significantly lower AUC ($b = -0.013$, $SE = 0.005$, $t(30) = -2.39$, $p = .024$) and MD ($b = -0.024$, $SE = 0.010$, $t(32) = -2.43$, $p = .021$). None of the interactions between Congruency contrast 1 and Noise were significant ($|t| < 1$, $p > .05$ for both AUC and MD). The interaction between Congruency contrast 2 and Noise was not significant for AUC ($b = -0.008$, $SE = 0.005$, $t(7543) = -1.55$, $p = .121$) but was significant for MD ($b = -0.022$, $SE = 0.010$, $t(7544) = -2.29$, $p = .022$). With increasing noise level, the no-context condition decreased the MD in a larger extent than the context conditions.

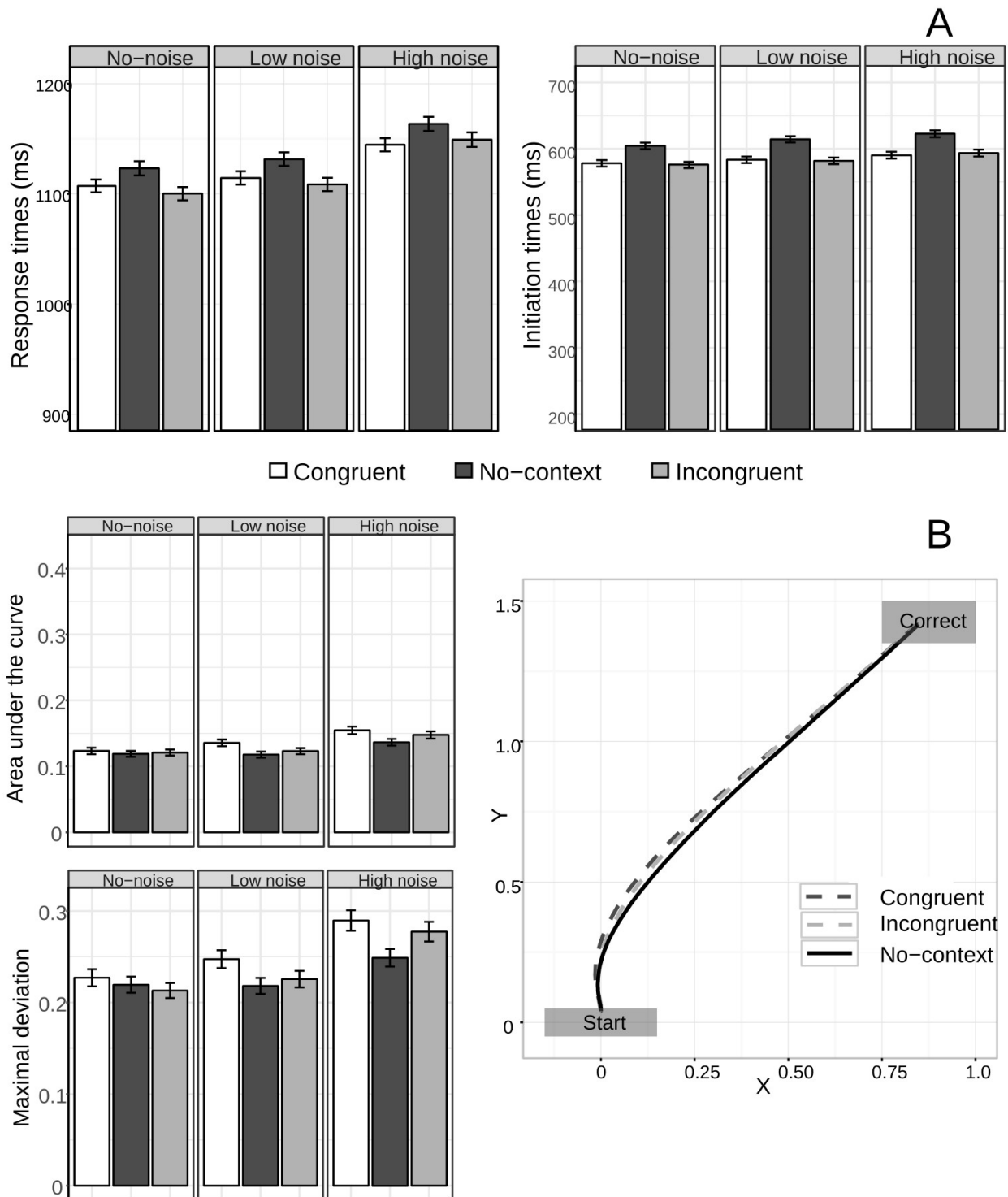


Figure 8: **A.** Response time as a function of context congruency and noise. White: Congruent context. Light grey: Incongruent context, Dark grey: No-context (grey background). Errors bars represent the standard error of the mean. **B.** Left: Area under the curve and maximal deviation as a function of context congruency and noise. Right: Mean mouse trajectory by context congruency conditions. Dark grey dashed curve: Congruent context, Light grey dashed curve: Incongruent context, black solid curve: No-context.

4.2.4. Analysis of velocity

The Java mouse-tracking program we used to run this experiment recorded the mouse events, not the cursor position at a regular framerate. While improving the spatial and temporal accuracy of the data, this format is not directly compatible with the procedure used in Experiments 1 and 2. First, we thus pre-processed each mouse trajectory using linear interpolation to obtain mouse coordinates at each 16-ms time-step. From then, we applied the same preprocessing steps, as for the previous experiments and calculated the MV and MV-time. Figure 9.A. displays the means of these measures by conditions.

Similarly to Experiment 1, only Noise had an effect on MV; Increasing noise level reduced the MV ($b = -0.031$, $SE = 0.01$, $t(104) = -3.07$, $p = .003$). There was no significant difference between congruent and incongruent contexts ($b = -0.001$, $SE = 0.022$, $t(51) = -0.06$, $p = .953$) nor between context conditions and no-context conditions ($b = 0.051$, $SE = 0.029$, $t(29) = 1.77$, $p = .088$). None of the interactions were significant either ($|t| < 1$, $p > 0.05$). Noise level significantly increased MV-time ($b = 17.76$, $SE = 1.82$, $t(30) = 9.76$, $p < .001$). Incongruent contexts did not significantly delay MV-time compared to congruent contexts ($b = -3.21$, $SE = 3.53$, $t(27) = -0.91$, $p = .371$) but MV-time was delayed by no-context primes compared to contextual primes ($b = 28.37$, $SE = 4.71$, $t(29) = 6.02$, $p < .001$). There was no interaction (both $|t| < 1$, $p > 0.05$).

Analysis of the velocity profiles using the same method as in the previous experiments confirmed that not only the velocity peak was delayed when no-context primes were presented but also the global mouse movement. Figure 9.B. displays the velocity profiles for each condition. Congruent and incongruent contexts increased velocity compared to no-context from 192-448 ms after target onset (and thus, /START/ button press) to 768-832 ms (depending on experimental conditions) and decreased it from 864-1040 ms to 1168-1696 ms. Similarly to Experiment 1, these results indicate a delay of the global response movement when no context was presented. Cross correlation of the velocity profiles are in accordance with this interpretation. All maximal correlations were above .99 at the optimal lag. While no lag is observed between congruent and incongruent velocity profiles for each noise level, no-context velocity profiles are delayed by one 16 ms time step compared to congruent and incongruent conditions. There is one exception: With a low noise level, congruent and no-context profiles have no lag. Yet, no-context profile was delayed by one time step compared to incongruent profile, and incongruent profile had no lag with congruent profile, the failure to detect a delay between congruent and no-context could be due to the low precision of our method, with a rather low temporal resolution (60 Hz) along with broad-range cross-correlations.

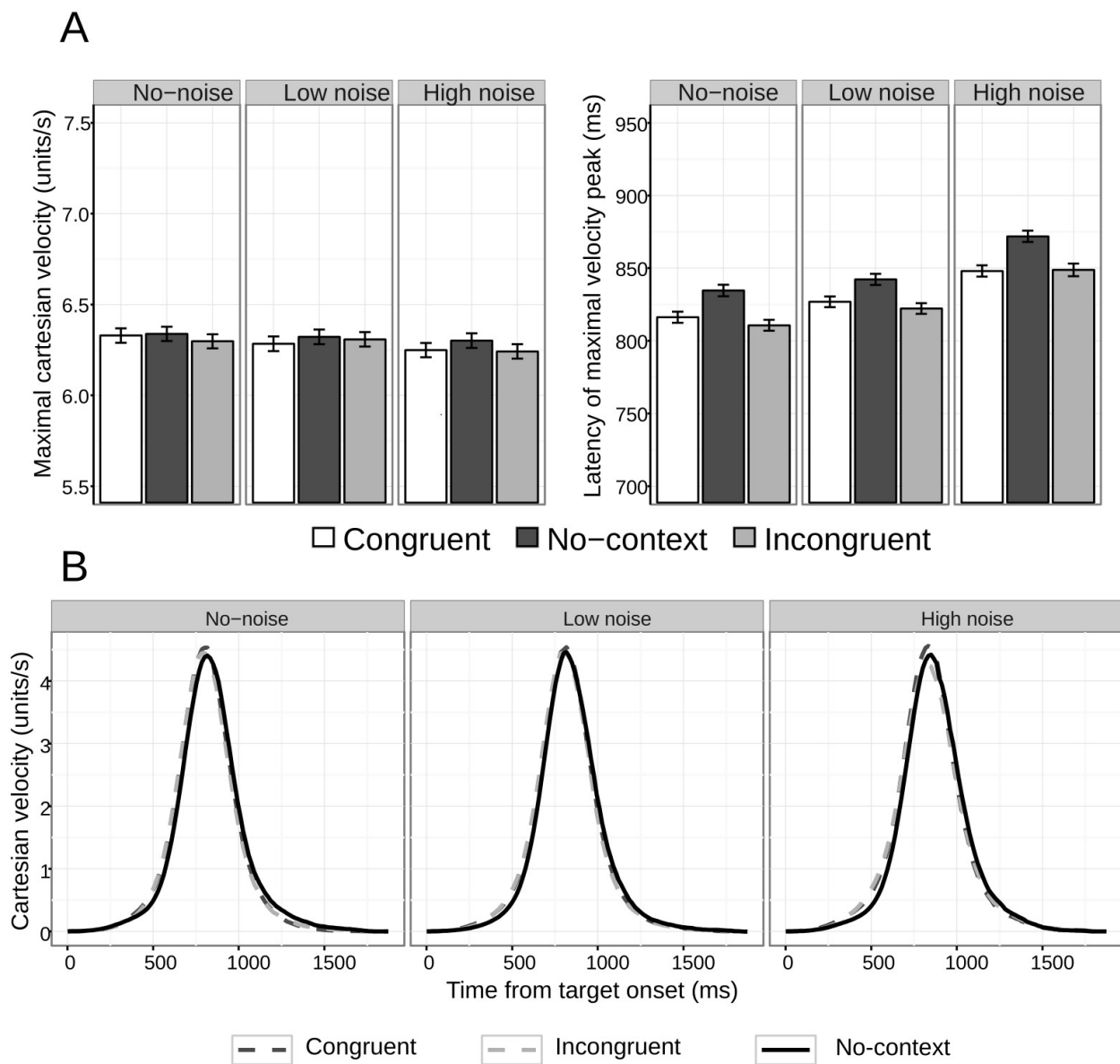


Figure 9: **A.** Maximal Cartesian velocity and Latency of the maximal velocity peak as a function of context congruency and noise. White: Congruent context. Light grey: Incongruent context, Dark grey: No-context (grey background). Errors bars represent the standard error of the mean. **B.** Cartesian velocity as a function of time. Dark grey dashed curve: Congruent context, Light grey dashed curve: Incongruent context, black solid curve: No-context. The velocity for each noise condition and at each time-step (16 ms intervals) was estimated using separate LMM. Context congruency was recoded as dummy coding with the No-context condition being the control; Variations of intercepts across participants and targets were specified as random effects.

4.3. Discussion

As in Experiment 1 and 2, we confirmed the effects of noise, which globally impaired the decision-making process by increasing the RT and IT, decreasing the mouse velocity and delaying the velocity peak, as well as increasing the mouse deviation.

To replicate the results of Experiment 1 and 2, we hypothesized responses biased towards speed when a context image (congruent or incongruent) was presented, and responses biased towards accuracy when no context was presented. This would mean that this pattern of results is not only permitted by a delay before target onset, where the participant waits for the target to appear. Evidence from the context can also be accumulated before the /START/ button is pressed. In this experiment, context presentation was clearly separated from target presentation and the response properly began at the target onset. Thus, this manipulation isolated the effects of prior contextual information on the SAT from potential effects of the stimulus sequence in generating anticipation.

Results showed that both response times and initiation times were longer for no-context conditions, that mouse deviations towards the incorrect response were reduced and that both the maximal velocity peaks and global response movements were delayed compared to the context conditions. When participants had an early access to potentially relevant prior information (conditions with context images presented), the decision process was triggered earlier compared to no-context conditions. As we interpret it, the context images generated the early competition between the two responses with no certainty about which one is correct, which led to early movement initiation and increased trajectory deviation just after target onset. Thus, before pressing the /START/ button, a certain amount of contextual information was already integrated. In the case where no such information was available (no-context conditions), the response process was not triggered until a sufficient amount of bottom-up information was accumulated, therefore leading to more accurate trajectories.

However, context congruency had no specific effect on the response dynamics for none of the measures, as we found no differences between congruent and incongruent contexts. This could indicate that when contextual prior information is provided during a time-frame clearly separated from the target on which a decision has to be made, its specific content is not taken into full account in the decision-making process. Note that this would apply when there is no certainty that the context will provide an accurate information –which is the case in this study where the occurrence rate of congruent contexts is equal to the occurrence rate of incongruent contexts.

Taken together, these results indicate that while the presence of a context did affect the time course of the response and its spatial bias, the specificity of its content was not taken into account in the decision process. This can be interpreted in terms of accumulation of evidence within the decision-making process,

where the context was taken as a potentially useful amount of information thus advancing the initiation of the decision process in order to integrate this information, but not reliable enough to bias the response towards a specific alternative.

5. General discussion

In this research, we presented a paradigm that makes it possible to measure and, to a certain extent, infer perceptual decision-making processes underlying visual object category recognition, thus providing a way of studying how contextual information is integrated with bottom-up information in the process of accumulating evidence. Throughout the three experiments, we distinguished two different effects of contextual information: context congruency (*vs.* incongruency) and context presence (*vs.* absence).

5.1. Influence of context congruency on response selection

The response movements were directed straighter towards the context-congruent response buttons than towards the incongruent ones in Experiments 1 and 2. The adaptations we implemented to the task in Experiment 2, also permitted to observe shorter overall response times (RT) and earlier response movements for congruent contexts than for incongruent contexts. However, this particular effect was not reproduced in Experiment 3. These results are better explained by a response bias induced by contexts that are congruent with the response labels (Hollingworth & Henderson, 1998, 1999) rather than by the contexts activating the representations linked with the target, thus facilitating its recognition (Auckland et al., 2007; Biederman et al., 1974; Davenport & Potter, 2004; Palmer, 1975). However, our main interest was not the identification itself but the global decision-making process underlying both the target identification and the response selection. Consequently, our methodology differed from that used in the experiments cited above. Whereas these studies focused on short stimulus presentations, mostly using a backward mask (except for Palmer, 1975) and proper identification of the target (by naming the target or indicating whether the target was present or absent for example), we used a design in which the target was presented until a response specifying the category of the target was made. By doing so, we allowed the emergence of a forced-choice decision-making process with the mouse movements permitting to measure the dynamics of the responses towards the accurate and alternative buttons.

What we can observe from the comparison between Experiment 1-2 and Experiment 3, is that the congruency effect was present when the context and target were presented in a single stimulus sequence,

whereas when the context was presented outside the response time-frame and the target onset was controlled by the participant, there was no congruency effect. In the former case, it seems that both context identity and target identity exerted an influence on response selection and this would be allowed by the context being presented during a stimulus sequence, even if response movements began after the target onset. This could highlight how the influence of the context specificity on a decision process is embodied within the motor response, which is the main assumption of mouse-tracking paradigms (Dale et al., 2007; Freeman, Dale, & Farmer, 2011; Song & Nakayama, 2009), and especially in the case of prior knowledge about the target (Beffara et al., 2012).

5.2. Speed-accuracy trade-off and evidence accumulation in the presence of a context

Contextual information had a second effect on the decision-making process. Regardless the identity of the context, its presentation biased the SAT towards speed, whereas the presentation of a uniform grey picture biased it towards accuracy. We found evidence of this in the results of all three experiments, no-context trials inducing smaller deviation, delayed movement initiation and delayed global movements. To our knowledge, such effect has not been found before in the field of contextual cueing. However, it is consistent with the *accumulation-to-bound* framework for decision-making (Gold & Shadlen, 2007; Ratcliff & Rouder, 1998; Ratcliff et al., 2016; Summerfield & de Lange, 2014), where the context can be considered as potential evidence, triggering the decision process initiation earlier. As a result, the competition between the two alternative responses begins earlier, and increases the deviation from a straight trajectory.

Experiment 2 illustrates how the context presence effect co-exists with the congruency effect. In Experiment 2, where the context was presented within the stimulus sequence including the decision, response trajectories began early for both context conditions, but later on, response trajectories for incongruent contexts were delayed and deviated compared with congruent contexts, indicating a correction of the initial decision. Stochastic evidence accumulation models explain response bias as a modulation of the evidence baselines, increasing the baseline of the anticipated alternative (or equivalently, a shift of the starting point in favour of the anticipated alternative for the drift diffusion models) whether it is a correct or an incorrect response (Gold & Shadlen, 2007; Mulder, Wagenmakers, Ratcliff, Boekel, & Forstmann, 2012; Ravenzwaaij, Mulder, Tuerlinckx, & Wagenmakers, 2012; Summerfield & de Lange, 2014). Furthermore, these models implement the SAT as a modulation of the global evidence baseline for all the alternative responses or similarly, of the threshold depending on the implementations) (Bogacz et al., 2010; Heitz, 2014). A higher baseline/lower threshold in speed-biased conditions results in the threshold being reached earlier, with the noise of the evidence accumulation being more influential towards reaching it. This highlights the relationship between the presence of a context biasing the SAT towards speed (global

baseline increase) and the congruency effect, in which the context identity is taken as relevant evidence (baseline increase for response congruent with context).

Another possible interpretation within the *accumulation-to-bound* framework is that when a context is presented, the integration and accumulation of visual evidence begins during its presentation and before target onset. Thus, when evidence about the target begins its accumulation, a certain amount of contextual information has already been accumulated. This interpretation can explain the difference in the results of Experiments 1-2 and Experiment 3. If the context is presented in a stimulus sequence where the target is presented and during which the movement has to be executed (Experiments 1 and 2), then evidence towards the congruent response is accumulated and can influence the response movement (Buc Calderon et al., 2015) as an additional source of information (besides target information). If the context is not presented in a time-frame when the response movement has to be executed (Experiment 3), the context still initiates the evidence accumulation, but no evidence is effectively accumulated towards the response selection (drift rate near 0) since pressing the /START/ button separates context perception from the response movement. In this perspective, movement preparation and decision process are inter-dependant. However, further studies should be conducted to discriminate between these two possible interpretations.

5.3. Integration of visual noise and context in the decision-making process

We found evidence of the context presence interacting with the level of noise. The noise was a direct manipulation of the availability or uncertainty of visual information on the target. We found evidence that increasing noise increased the RT, the IT and the deviation from the optimal response trajectory, reduced the response velocity and delayed the velocity peak. This indicates that less accessible visual information globally impaired the decision-making as measured through mouse trajectories, consistently with Palmer et al. (2005). Furthermore, in Experiment 1 and 2, we found evidence that the noise degraded the RT and MV-time in a larger extent when a context was presented compared to no-context (although the interaction is only marginally significant for RT in Experiment 2). However, this interaction was not found for the movement initiation. In these experiments, the context was presented in the same stimulus sequence and it biases the SAT towards speed (compared with no-context) and therefore, decisions on non-degraded targets began earlier and were quicker. But for increased noise levels, the speed bias does not allow to accumulate enough evidence and it is required to reduce it. Thus, these results give evidence of an active mechanism allowing to increase the response movements duration after their initiation, when more time is needed to cope with the uncertainty of visual information. Similar results can be found in the literature. Hence, the experience of difficulty (Alter, Oppenheimer, Epley, & Eyre, 2007) and in the field of attention,

conditions inducing a high perceptual load (such as crowding) (Lavie, 2010; Lavie & Tsal, 1994) can reduce the reliance on non-informative or interferent cues. In our task, a high noise inducing experience of difficulty or a high perceptual load would reduce the effect of context presence and increase the focus on the target and thus, reduce the decision speed.

By contrast, the noise did not moderate the speed bias of context presence when the context was not in the same stimulus sequence as the target (Experiment 3). Nevertheless, the fact that in Experiment 3, the SAT was still biased towards speed compared with no context indicates that the context played a permissive role in evidence accumulation. These results could suggest that when presented consecutively (Experiment 1 and 2), some of the contextual information is integrated in parallel with target information, which is not the case if the beginning of the decision sequence precedes target onset (Experiment 3). It is to be noted that this latter manipulation can also diminish perceived difficulty, as the time course of the trial is more controlled by the participants. This gives account of noise and context information being integrated, through the adaptation of the SAT, in a manner that is independent of the content of context information. In other words, whether the actual context content was congruent with the target or incongruent did not interact with noise conditions. Therefore the congruency of the context was apparently not used to compensate for a lack of visual bottom-up information, nor its effect was reduced by noise. This finding seems difficult to conciliate with the idea that the response selection process is based on a process combining optimally the accumulated evidence with prior information about the context up to a threshold, as traditionally assumed in the *accumulation-to-bound* framework (Bogacz et al., 2006; Ratcliff et al., 2016; Summerfield & de Lange, 2014).

However, it is important to note that in our experiment, contextual (prior) information was very uncertain and its presentation did not increase the probability of the appearance of congruent targets (or in other words had low precision). Participants might have learned this fact over the course of the experiment, which involved a lot of trials and included a training session with both congruent and incongruent trials. It could be argued that the risks associated with relying on the context (which may or may not induce costs on performance) under visual uncertainty overcame the costs of accumulating evidence (Drugowitsch et al., 2012). In another study, we observed such interaction between degradation by fog (on 3D stimuli) and the probabilistic expectation of the appearance of a target (among two) on mouse trajectory deviation (Quétard et al., 2015).

5.4. Perspectives and conclusion

The use of mouse-tracking tools to measure perceptual decision-making allowed us to undertake a more fine-grained analysis of our participants' behaviour than the traditional analysis of response times and

Article 1 (Étude 1) – Contexte et bruit dans le déroulement temporel des décisions perceptuelles

accuracy, especially taking into account the SAT across experimental condition. This method avoids the issue of scaling the task difficulty in order to obtain a reasonable error rate, since the mouse trajectory measures make it possible to evaluate the precision and accuracy of the correct responses. We emphasized how the available contextual information affected the accumulation of evidence through the SAT and through conflicting influences of incongruent context and target in an embodied manner.

An open question concerns the precise mechanism biasing the SAT towards speed in the presence of a context. As previously discussed, this study cannot conclude on whether the presence of a context before target onset triggers the accumulation of evidence, or the absence of a context leads to increased threshold due to uncertainty (alternatively, the presence of a context increases the evidence baseline due to less perceived uncertainty). This question presents interesting perspectives about how the decision process should be modelled in the *accumulation-to-bound* framework. The hypothesis that the absence of a context increases the threshold can be explained by basic sequential sampling models such as the drift diffusion model, the same way the SAT is implemented (Forstmann, Ratcliff, & Wagenmakers, 2016; Heitz, 2014). However, the hypothesis that contextual information is accumulated early and used (Experiment 1 and 2) or not (Experiment 3) for target selection and motor preparation implies a model including a deadline or an urgency signal mechanism (such as thresholds collapsing as decision time elapses) (Ditterich, 2006; Drugowitsch et al., 2012; Thura, Beauregard-Racine, Fradet, & Cisek, 2012). Without such mechanism, these kind of models cannot account for the results of Experiment 3, where the responses movements finished earlier in trials including a context compared with those not included one, while the contextual evidence was not used for responses selection.

Hence, further studies need to be carried on, with the interesting perspective of addressing the question of the architecture of perceptual decision-making models. Finally, mouse-tracking and reaching tasks are of particular interest as they emphasize the interdependence, and even maybe, the embodiment of decision and action (Cisek & Pastor-Bernier, 2014; Lepora & Pezzulo, 2015; Pezzulo & Cisek, 2016). Their combination with other tools such as eye-tracking (Quétard et al., 2015, 2016) could lead to a better understanding of the link between active evidence gathering (e.g. visual exploration), decision-making, and motor preparation.

Acknowledgements

This work has been sponsored by the French government research program "Investissements d'Avenir" managed by the French National Research Agency (ANR) in the framework of the IMobS3 Laboratory of Excellence (ANR-10-LABX-16-01) and the RobotEx Equipment of Excellence (ANR-10-EQPX-44), by the European Union through the Regional Competitiveness and Employment program -2014-2020- (ERDF – AURA region) and by the AURA région.

Bibliography

- Alter, A. L., Oppenheimer, D. M., Epley, N., & Eyre, R. N. (2007). Overcoming intuition: metacognitive difficulty activates analytic reasoning. *Journal of Experimental Psychology: General*, *136*(4), 569–576. <http://doi.org/10.1037/0096-3445.136.4.569>
- Auckland, M. E., Cave, K. R., & Donnelly, N. (2007). Nontarget objects can influence perceptual processes during object recognition. *Psychonomic Bulletin & Review*, *14*(2), 332–337. <http://doi.org/10.3758/BF03194073>
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. <http://doi.org/10.1016/j.jml.2007.12.005>
- Bar, M. (2009). The proactive brain: memory for predictions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1235–1243. <http://doi.org/10.1098/rstb.2008.0310>
- Barca, L., & Pezzulo, G. (2015). Tracking second thoughts: continuous and discrete revision processes during visual lexical decision. *PLoS One*, *10*(2), e0116193. <http://doi.org/10.1371/journal.pone.0116193>
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. <http://doi.org/10.1016/j.jml.2012.11.001>
- Bates, D., Maechler, M., & Bolker, B. (2012). lme4: Linear mixed-effects models using Eigen and Eigen++. <https://cran.r-project.org/web/packages/lme4/index.html>
- Beffara, B., Ouellet, M., Vermeulen, N., Basu, A., Morisseau, T., & Mermillod, M. (2012). Enhanced embodied response following ambiguous emotional processing. *Cognitive Processing*, *13*(Suppl 1), S103–S106. <http://doi.org/10.1007/s10339-012-0468-6>
- Biederman, I. (1972). Perceiving Real-World Scenes. *Science*, *177*(4043), 77–80. <http://doi.org/10.1126/science.177.4043.77>
- Biederman, I., Rabinowitz, J. C., Glass, A. L., & Stacy, E. W. (1974). On the information extracted from a

Article 1 (Étude 1) – Contexte et bruit dans le déroulement temporel des décisions perceptuelles

glance at a scene. *Journal of Experimental Psychology*, 103(3), 597–600.

<http://doi.org/10.1037/h0037158>

- Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological Review*, 113(4), 700–765. <http://doi.org/10.1037/0033-295X.113.4.700>
- Bogacz, R., Wagenmakers, E. J., Forstmann, B. U., & Nieuwenhuis, S. (2010). The neural basis of the speed-accuracy tradeoff. *Trends in Neurosciences*, 33(1), 10–16. <http://doi.org/10.1016/j.tins.2009.09.002>
- Bonin, P., Gelin, M., Laroche, B., Méot, A., & Bugajska, A. (2015). The “how” of animacy effects in episodic memory. *Experimental Psychology*, 62(6), 371–384. <http://doi.org/10.1027/1618-3169/a000308>
- Brockmole, J. R., Castelano, M. S., & Henderson, J. M. (2006). Contextual cueing in naturalistic scenes: Global and local contexts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(4), 699–706. <http://doi.org/10.1037/0278-7393.32.4.699>
- Brockmole, J. R., & Võ, M. L.-H. (2010). Semantic memory for contextual regularities within and across scene categories: Evidence from eye movements. *Attention, Perception & Psychophysics*, 72(7), 1803–1813. <http://doi.org/10.3758/APP.72.7.1803>
- Buc Calderon, C., Verguts, T., & Gevers, W. (2015). Losing the boundary: Cognition biases action well after action selection. *Journal of Experimental Psychology: General*, 144(4), 737–743. <http://doi.org/10.1037/xge0000087>
- Chen, B., & Perona, P. (2015). Speed versus accuracy in visual search: Optimal performance and neural architecture. *Journal of Vision*, 15(16):9, 1–29. <http://doi.org/10.1167/15.16.9>
- Churchland, A. K., Kiani, R., & Shadlen, M. N. (2008). Decision-making with multiple alternatives. *Nature Neuroscience*, 11(6), 693–702. <http://doi.org/10.1038/nn.2123>
- Cisek, P., & Pastor-Bernier, A. (2014). On the challenges and mechanisms of embodied decisions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1655), 20130479–20130479. <http://doi.org/10.1098/rstb.2013.0479>
- Dale, R., Kehoe, C., & Spivey, M. J. (2007). Graded motor responses in the time course of categorizing atypical exemplars. *Memory & Cognition*, 35(1), 15–28. <http://doi.org/10.3758/BF03195938>
- Davenport, J. L., & Potter, M. C. (2004). Scene Consistency in Object and Background Perception. *Psychological Science*, 15(8), 559–564. <http://doi.org/10.1111/j.0956-7976.2004.00719.x>
- Ditterich, J. (2006). Evidence for time-variant decision making. *European Journal of Neuroscience*, 24(12), 3628–3641. <http://doi.org/10.1111/j.1460-9568.2006.05221.x>
- Donnarumma, F., Costantini, M., Ambrosini, E., Friston, K. J., & Pezzulo, G. (2017). Action perception as hypothesis testing. *Cortex*, 89, 45–60. <http://doi.org/10.1016/j.cortex.2017.01.016>
- Drugowitsch, J., Moreno-Bote, R., Churchland, A. K., Shadlen, M. N., & Pouget, A. (2012). The Cost of Accumulating Evidence in Perceptual Decision Making. *Journal of Neuroscience*, 32(11), 3612–3628.

Article 1 (Étude 1) – Contexte et bruit dans le déroulement temporel des décisions perceptuelles

<http://doi.org/10.1523/JNEUROSCI.4010-11.2012>

- Fenske, M. J., Aminoff, E. A., Gronau, N., & Bar, M. (2006). Top-down facilitation of visual object recognition: object-based and context-based contributions. *Progress in Brain Research*, *155*, 3–21. [http://doi.org/10.1016/S0079-6123\(06\)55001-0](http://doi.org/10.1016/S0079-6123(06)55001-0)
- Forstmann, B. U., Ratcliff, R., & Wagenmakers, E.-J. (2016). Sequential Sampling Models in Cognitive Neuroscience: Advantages, Applications, and Extensions. *Annual Review of Psychology*, *67*, 641–66. <http://doi.org/10.1146/annurev-psych-122414-033645>
- Freeman, J. B., & Ambady, N. (2010). MouseTracker: Software for studying real-time mental processing using a computer mouse-tracking method. *Behavior Research Methods*, *42*(1), 226–241. <http://doi.org/10.3758/BRM.42.1.226>
- Freeman, J. B., Dale, R., & Farmer, T. A. (2011). Hand in Motion Reveals Mind in Motion. *Frontiers in Psychology*, *2*:59. <http://doi.org/10.3389/fpsyg.2011.00059>
- Friston, K. J. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1456), 815–836. <http://doi.org/10.1098/rstb.2005.1622>
- Friston, K. J., Adams, R. A., Perrinet, L., & Breakspear, M. (2012). Perceptions as hypotheses: saccades as experiments. *Frontiers in Psychology*, *3*, 1–20. <http://doi.org/10.3389/fpsyg.2012.00151>
- Gilbert, C. D., & Li, W. (2013). Top-down influences on visual processing. *Nature Reviews Neuroscience*, *14*(5), 350–363. <http://doi.org/10.1038/nrn3476>
- Gold, J. I., & Shadlen, M. N. (2001). Neural computations that underlie decisions about sensory stimuli. *Trends in Cognitive Sciences*, *5*(1), 10–16. [http://doi.org/10.1016/S1364-6613\(00\)01567-9](http://doi.org/10.1016/S1364-6613(00)01567-9)
- Gold, J. I., & Shadlen, M. N. (2007). The Neural Basis of Decision Making. *Annual Review of Neuroscience*, *30*(1), 535–574. <http://doi.org/10.1146/annurev.neuro.29.051605.113038>
- Greene, M. R., & Oliva, A. (2009). The briefest of glances: the time course of natural scene understanding. *Psychological Science*, *20*(4), 464–472. <http://doi.org/10.1111/j.1467-9280.2009.02316.x>
- Hegd , J. (2008). Time course of visual perception: coarse-to-fine processing and beyond. *Progress in Neurobiology*, *84*(4), 405–439. <http://doi.org/10.1016/j.pneurobio.2007.09.001>
- Helman, E., Stoller, R., & Freeman, J. (2015). Advanced Mouse-Tracking Analytic Techniques for Enhancing Psychological Science. *Group Process & Intergroup Relations*, *8*(3), 384–401. <http://doi.org/10.1177/1368430214538325>
- Heitz, R. P. (2014). The speed-accuracy tradeoff: History, physiology, methodology, and behavior. *Frontiers in Neuroscience*, *8*(8 JUN), 1–19. <http://doi.org/10.3389/fnins.2014.00150>
- Heitz, R. P., & Schall, J. D. (2012). Neural Mechanisms of Speed-Accuracy Tradeoff. *Neuron*, *76*(3), 616–628. <http://doi.org/10.1016/j.neuron.2012.08.030>
- Heitz, R. P., & Schall, J. D. (2013). Neural chronometry and coherency across speed-accuracy demands reveal lack of homomorphism between computational and neural mechanisms of evidence accumulation.

Article 1 (Étude 1) – Contexte et bruit dans le déroulement temporel des décisions perceptuelles

Philosophical Transactions of the Royal Society B: Biological Sciences, 368(1628), 20130071.
<http://doi.org/10.1098/rstb.2013.0071>

Hollingworth, A., & Henderson, J. M. (1998). Does consistent scene context facilitate object perception? *Journal of Experimental Psychology: General*, 127(4), 398–415.

Hollingworth, A., & Henderson, J. M. (1999). Object identification is isolated from scene semantic constraint: evidence from object type and token discrimination. *Acta Psychologica*, 102(2–3), 319–343.
[http://doi.org/10.1016/S0001-6918\(98\)00053-5](http://doi.org/10.1016/S0001-6918(98)00053-5)

Kietzmann, T. C., Geuter, S., & König, P. (2011). Overt visual attention as a causal factor of perceptual awareness. *PLoS One*, 6(7), e22614. <http://doi.org/10.1371/journal.pone.0022614>

Kietzmann, T. C., & König, P. (2015). Effects of contextual information and stimulus ambiguity on overt visual sampling behavior. *Vision Research*, 110, 76–86. <http://doi.org/10.1016/j.visres.2015.02.023>

Kovalenko, L. Y., Chaumon, M., & Busch, N. A. (2012). A pool of pairs of related objects (POPORO) for investigating visual semantic integration: behavioral and electrophysiological validation. *Brain Topography*, 25(3), 272–284. <http://doi.org/10.1007/s10548-011-0216-8>

Kveraga, K., Ghuman, A. S., Kassam, K. S., Aminoff, E. A., Hämäläinen, M. S., Chaumon, M., & Bar, M. (2011). Early onset of neural synchronization in the contextual associations network. *Proceedings of the National Academy of Sciences of the United States of America*, 108(8), 3389–3394.
<http://doi.org/10.1073/pnas.1013760108>

Lavie, N. (2010). Attention, Distraction, and Cognitive Control Under Load. *Current Directions in Psychological Science*, 19(3), 143–148. <http://doi.org/10.1177/0963721410370295>

Lavie, N., & Tsal, Y. (1994). Perceptual load as a major determinant of the locus of selection in visual attention. *Perception & Psychophysics*, 56(2), 183–197. <http://doi.org/10.3758/BF03213897>

Lepora, N. F., & Pezzulo, G. (2015). Embodied choice: How action influences perceptual decision making. *PLoS Computational Biology*, 11(4), e1004110. <http://doi.org/10.1371/journal.pcbi.1004110>

Malcolm, G. L., & Henderson, J. M. (2010). Combining top-down processes to guide eye movements during real-world scene search. *Journal of Vision*, 10(2):4, 1–11. <http://doi.org/10.1167/10.2.4>

Mermillod, M., Droit-Volet, S., Devaux, D., Schaefer, A., & Vermeulen, N. (2010). Are coarse scales sufficient for fast detection of visual threat? *Psychological Science*, 21(10), 1429–37.
<http://doi.org/10.1177/0956797610381503>

Mulder, M. J., Wagenmakers, E., Ratcliff, R., Boekel, W., & Forstmann, B. U. (2012). Bias in the brain: A diffusion model analysis of prior probability and potential payoff. *Journal of Neuroscience*, 32(7), 2335–2343. <http://doi.org/10.1523/JNEUROSCI.4156-11.2012>

Palmer, J., Huk, A. C., & Shadlen, M. N. (2005). The effect of stimulus strength on the speed and accuracy of a perceptual decision. *Journal of Vision*, 5, 376–404. <http://doi.org/10.1167/5.5.1>

Palmer, S. E. (1975). The effects of contextual scenes on the identification of objects. *Memory & Cognition*,

Article 1 (Étude 1) – Contexte et bruit dans le déroulement temporel des décisions perceptuelles

3(5), 519–526. <http://doi.org/10.3758/BF03197524>

Pezzulo, G., & Cisek, P. (2016). Navigating the Affordance Landscape: Feedback Control as a Process Model of Behavior and Cognition. *Trends in Cognitive Sciences*, 20(6), 414–424. <http://doi.org/10.1016/j.tics.2016.03.013>

Quéward, B., Quinton, J.-C., Colomb, M., Pezzulo, G., Barca, L., Izaute, M., ... Mermillod, M. (2015). Combined effects of expectations and visual uncertainty upon detection and identification of a target in the fog. *Cognitive Processing*, 16(S1), 343–348. <http://doi.org/10.1007/s10339-015-0673-1>

Quéward, B., Quinton, J. C., Mermillod, M., Barca, L., Pezzulo, G., Colomb, M., & Izaute, M. (2016). Differential effects of visual uncertainty and contextual guidance on perceptual decisions: Evidence from eye and mouse tracking in visual search. *Journal of Vision*, 16(11):28, 1-28. <http://doi.org/10.1167/16.11.28>

Quinton, J.-C., Volpi, N. C., Barca, L., & Pezzulo, G. (2014). The Cat is on the Mat. or is it a Dog? Dynamic Competition in Perceptual Decision Making. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 44(5), 539–551. <http://doi.org/10.1109/TSMC.2013.2279664>

Ratcliff, R., & McKoon, G. (2008). The Diffusion Decision Model: Theory and Data for Two-Choice Decision Tasks. *Neural Computation*, 20(4), 873–922. <http://doi.org/10.1162/neco.2008.12-06-420>

Ratcliff, R., & Rouder, J. N. (1998). Modeling Response Times for Two-Choice Decisions. *Psychological Science*, 9(5), 347–356. <http://doi.org/10.1111/1467-9280.00067>

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion Decision Model: Current Issues and History. *Trends in Cognitive Sciences*, 20(4), 260–281. <http://doi.org/10.1016/j.tics.2016.01.007>

Ravenzwaaij, D. Van, Mulder, M. J., Tuerlinckx, F., & Wagenmakers, E. (2012). Do the dynamics of prior information depend on task context? An analysis of optimal performance and an empirical test. *Frontiers in Psychology*, 3:132. <http://doi.org/10.3389/fpsyg.2012.00132>

Resulaj, A., Kiani, R., Wolpert, D. M., & Shadlen, M. N. (2009). Changes of mind in decision-making. *Nature*, 461(7261), 263–266. <http://doi.org/10.1038/nature08275>

Röhrbein, F., Goddard, P., Schneider, M., James, G., & Guo, K. (2015). How does image noise affect actual and predicted human gaze allocation in assessing image quality? *Vision Research*, 112, 11–25. <http://doi.org/10.1016/j.visres.2015.03.029>

Smeding, A., Quinton, J., Lauer, K., Barca, L., & Pezzulo, G. (2016). Tracking and Simulating Dynamics of Implicit Stereotypes: A Situated Social Cognition Perspective. *Journal of Personality and Social Psychology*, 111(6), 817–834. <http://doi.org/10.1037/pspa0000063>

Song, J.-H., & Nakayama, K. (2009). Hidden cognitive states revealed in choice reaching tasks. *Trends in Cognitive Sciences*, 13(8), 360–366. <http://doi.org/10.1016/j.tics.2009.04.009>

Spivey, M. J., Grosjean, M., & Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences*, 102(29), 10393–10398. <http://doi.org/www.pnas.org/cgi/doi/10.1073/pnas.0503903102>

Article 1 (Étude 1) – Contexte et bruit dans le déroulement temporel des décisions perceptuelles

Spotorno, S., Malcolm, G. L., & Tatler, B. W. (2014). How context information and target information guide the eyes from the first epoch of search in real-world scenes. *Journal of Vision, 14*(2):7, 1–21. <http://doi.org/10.1167/14.2.7>

Summerfield, C., & de Lange, F. P. (2014). Expectation in perceptual decision making: neural and computational mechanisms. *Nature Reviews Neuroscience, 15*(11), 745–756. <http://doi.org/10.1038/nrn3838>

Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences, 13*(9), 403–409. <http://doi.org/10.1016/j.tics.2009.06.003>

Thura, D., Beauregard-Racine, J., Fradet, C.-W., & Cisek, P. (2012). Decision making by urgency gating: theory and experimental support. *Journal of Neurophysiology, 108*(11), 2912–2930. <http://doi.org/10.1152/jn.01071.2011>

Wang, Q., Taylor, H. A., & Brunyé, T. T. (2012). When going the right way is hard to do: Distinct phases of action compatibility in spatial knowledge development. *Acta Psychologica, 139*(3), 449–457. <http://doi.org/10.1016/j.actpsy.2012.01.006>

Wickelgren, W. A. (1977). Speed-Accuracy Tradeoff and Information Processing Dynamics. *Acta Psychologica, 41*, 67–85.

Xiao, J., Hays, J., Ehinger, K. A., Oliva, A., & Torralba, A. (2010). SUN database: Large-scale scene recognition from abbey to zoo. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 3485–3492). IEEE. <http://doi.org/10.1109/CVPR.2010.5539970>

1.1. Résumé des résultats de l'Étude 1

L'Étude 1 n'a pas permis de mettre en évidence d'effet représentationnel du contexte sur la catégorisation de la cible (*i.e.*, le contexte biaise la catégorisation en facilitant la reconnaissance de la cible). Dans notre tâche, le contexte facilite ou dégrade la catégorisation, plutôt par un biais du mouvement de réponse vers la réponse congruente au contexte. Cependant, elle met en évidence de manière indirecte, une modulation de la collecte d'évidence par la présence d'un contexte congruent comme incongruent, biaisant celle-ci vers la vitesse alors qu'elle était biaisée vers l'exactitude lorsque qu'aucun contexte n'était présenté. Enfin, nous avons observé une dégradation globale de la prise de décision lorsque la cible était bruitée. Dans cette étude, nous n'avons pas directement mesuré la collecte d'information, nous avons inféré sa modulation via la mise en évidence de variations du CVE. En ce sens, nos résultats pointent vers l'idée que même lorsque que le processus de prise de décision consiste à identifier une cible suffisamment petite pour que l'explorer activement du regard ne soit pas nécessaire, la collecte d'évidence est adaptée à la quantité d'information présentée. Notre méthodologie ne nous permettait donc pas d'évaluer directement l'influence d'une collecte active d'évidence dans le processus de décision, et comment l'incertitude de l'information visuelle et les attentes générées par le contexte affectent l'accumulation d'évidence et la sélection de la réponse. C'est précisément l'objet de l'Étude 2.

2. Étude 2 : Influence de l'incertitude visuelle et de l'indiçage contextuel sur l'accumulation d'évidence via les mouvements des yeux et la prise de décision lors de la recherche visuelle

Pour mesurer directement la collecte active d'évidence, nous avons proposé, à l'instar de Kietzmann et collaborateurs (Kietzmann et al., 2011; Kietzmann & König, 2015) et Quinton et collaborateurs (Quinton et al., 2014), la mesure des mouvements des yeux. Son couplage avec un paradigme de *mouse-tracking* permet ainsi de mesurer simultanément la collecte d'évidence et des aspects de la prise de décision tels que la sélection de réponse, de façon continue. Les tâches employées dans l'Étude 1 n'étant pas très adaptées à la mesure des mouvements des yeux –les cibles de nécessitant pas une exploration visuelle approfondie pour être catégorisées– nous avons préféré opter pour une tâche de recherche visuelle employant des scènes visuelles naturelles. Comme présenté dans les sections précédentes, l'allocation du regard lors de ce type de tâches a été extensivement étudié, sans prendre en compte, néanmoins, le processus de prise de décision sous-jacent (ou seulement marginalement). L'Étude 2 propose cela au travers d'un article comprenant une expérience, publié dans la revue *Journal of Vision*. Les participants devaient rechercher dans des scènes d'intérieur de maison un objet cible que l'on peut usuellement trouver dans la pièce présentée. Ces objets pouvaient être placés à un endroit habituel ou inhabituel (indiçage contextuel), ou être absent de la scène. Les participants devaient indiquer, en cliquant sur le bouton correspondant, si la cible était présente ou non. De plus, la scène pouvait être bruitée ou non. Nous mesurions notamment les temps de détection et de vérification comme mesures globales de l'exploration visuelle, en plus des mesures des mouvements de la souris comme mesures du processus de prise de décision lui-même.

Dans le prolongement des résultats rapportés dans le Chapitre 2 Section 2.3.2 nous attendions des temps de détection plus faibles lorsque la cible était située à un endroit inhabituel. Lorsque la scène était bruitée, nous pouvions attendre une augmentation des temps de vérification et de la durée des fixations durant la période de détection, le bruit augmentant la durée des fixations (Harley et al., 2004; Röhrbein et al., 2015) ainsi qu'éventuellement, une augmentation des temps de détection dans le cas où le bruit dégraderait l'information en vision périphérique (Nuthmann, 2014). Concernant les mesures de prise de décision, nous attendions des temps de décision allongés et trajectoires déviées vers la réponse *absent* lorsque les scènes étaient bruitées et lorsque la cible était située à un endroit inhabituel. Celle-ci étant plus difficile à trouver dans les deux cas, la réponse *absent* reste plus longtemps une réponse valide, déviant les trajectoires dans sa direction. Cela sous-entend que des mécanismes d'exploration visuelle sont (au moins en partie) à l'origine de ces effets sur le processus de prise de décision.

Deux formes d'analyses statistiques ont été employées afin d'explorer la complémentarité des

Chapitre 3 – Études expérimentales

processus d'accumulation d'évidence et de prise de décision. Dans un premier temps, des analyses de médiation ont permis de rendre compte des rôles respectifs des processus de détection et de vérification pour expliquer les effets de l'indiçage contextuel et du bruit sur les mesures globales de la sélection de réponse (temps de réponse et aires-sous-la-courbe). Dans un second temps, nous avons mis en relation l'évolution temporelle de la décision avec le moment où la cible est détectée et le moment où la cible est vérifiée, permettant de mettre en avant le fait que les processus de détection puis de vérification de la cible se déroulent durant, et influencent en parallèle la sélection de la réponse. Pour cela, nous avons construit les profils de déviation et de vélocité en centrant l'axe temporel sur le moment de la détection de la cible, permettant de mettre en avant que la réponse et sa déviation débute avant la détection et se poursuit durant toute la période de vérification.

Article 2

Étude 2, expérience 1

Differential effects of visual uncertainty and contextual guidance on perceptual decisions: Evidence from eye and mouse tracking in visual search

Boris Quétard,

Jean-Charles Quinton,

Martial Mermillod,

Laura Barca,

Giovanni Pezzulo,

Michèle Colomb,

Marie Izaute









Journal: Journal of Vision

Statut: Publié

Détails de publication:

Quétard et al. (2016). *J Vis*, 16(11):28, 1-21. doi: 10.1167/16.11.28

Differential effects of visual uncertainty and contextual guidance on perceptual decisions: Evidence from eye and mouse tracking in visual search

Boris Quétard	Clermont University, Blaise Pascal University, LAPSCO, Clermont-Ferrand, France CNRS, UMR 6024, LAPSCO, Clermont-Ferrand, France	
Jean Charles Quinton	University Grenoble Alpes, LJK, Grenoble, France CNRS, UMR 5224, LJK, Grenoble, France	 
Martial Mermillod	University Grenoble Alpes, LPNC, F-38000 Grenoble, France CNRS, LPNC, F-38000 Grenoble, France	 
Laura Barca	Institute of Cognitive Sciences and Technologies, CNR, Rome, Italy	 
Giovanni Pezzulo	Institute of Cognitive Sciences and Technologies, CNR, Rome, Italy	 
Michèle Colomb	CEREMA, Département Laboratoire de Clermont-Ferrand, Clermont-Ferrand, France	
Marie Izaute	Clermont University, Blaise Pascal University, LAPSCO, Clermont-Ferrand, France CNRS, UMR 6024, LAPSCO, Clermont-Ferrand, France	 

Visual search can be seen as a decision-making process that aims to assess whether a target is present or absent from a scene. In this perspective, eye movements collect evidence related to target detection and verification to guide the decision. We investigated whether, in real-world scenes, target detection and verification are differentially recruited in the decision-making process in the presence of prior information (expectations about target location) and perceptual uncertainty (noise). We used a mouse-tracking methodology with which mouse trajectories unveil components of decision-making and eye-tracking measures reflect target detection and verification. Indoor scenes were presented, including a target in usual or unusual locations or no target, and were degraded with additive noise (or no noise). Participants had to respond to the target's presence or absence. Degrading the scene delayed the decision due

to increased verification times and reduced mouse velocity. Targets in unusual locations delayed the decision and deviated mouse trajectories toward the *target-absent* response. Detection times played a major role in these effects. Thus, target detection and verification processes influence decision-making by integrating the available sources of information differently and lead to an accumulation of evidence toward both the presence of a target and its absence.

Introduction

Visual scenes from everyday life can often be very complex, including elaborate details and textures as

Citation: Quétard, B., Quinton, J. C., Mermillod, M., Barca, L., Pezzulo, G., Colomb, M., & Izaute, M. (2016). Differential effects of visual uncertainty and contextual guidance on perceptual decisions: Evidence from eye and mouse tracking in visual search. *Journal of Vision*, 16(11):28, 1–21, doi:10.1167/16.11.28.

doi: 10.1167/16.11.28

Received December 22, 2015; published September 30, 2016

ISSN 1534-7362

This work is licensed under a Creative Commons Attribution 4.0 International License.
Downloaded From: <http://jov.arvojournals.org/pdfaccess.ashx?url=/data/Journals/JOV/935705/> on 10/04/2016



well as a large number of objects, and they cannot be fully processed in a single glance. Thus, visual information is sampled through discrete eye fixations. As actions can be made not only to progress toward a goal, but also to reduce the uncertainty about the states of the world (Bajcsy, Aloimonos, & Tsotsos, 2016; Kirsh & Maglio, 1994), active oculomotor behavior could be directed toward minimizing uncertainty about the composition of the visual scene (Friston, 2012; Friston, Adams, Perrinet, & Breakspear, 2012; Itti & Baldi, 2009; Renninger, Verghese, & Coughlan, 2007) or object identities (Kietzmann, Geuter, & König, 2011; Kietzmann & König, 2015; Quinton, Volpi, Barca, & Pezzulo, 2014). This behavior is usually driven by particular goals, such as understanding what is depicted, gathering action-related information, or searching for a particular object (Ballard & Hayhoe, 2009; Land & Hayhoe, 2001; Rothkopf, Ballard, & Hayhoe, 2007). Most real-life tasks require fast visual processing. Therefore, efficient visual exploration has to rely on rapidly available sources of information. This includes both the visual information that can be extracted directly from the current view of the environment and prior information (e.g., knowledge and expectations about this environment) built up as a result of past exposures to similar situations, the effects of which may start even before any visual stimulation occurs.

The former was previously formalized as bottom-up saliency maps (Borji, Tavakoli, Sihite, & Itti, 2013; Itti & Koch, 2000; Itti, Koch, & Niebur, 1998), with which highly conspicuous areas of the scene (the features of which differ from surroundings) are thought to attract attention and, ultimately, the gaze. This model reproduces some experimental human data especially in free viewing (Borji et al., 2013; Itti & Koch, 2000) and predicts human gaze fixation locations better than various random models in a scene-encoding task (Foulsham & Underwood, 2008; Underwood & Foulsham, 2006). However, it predicts neither the order of the fixations in the sequence (i.e., scanpaths; Foulsham & Underwood, 2008; Underwood, Foulsham, & Humphrey, 2009) nor the oculomotor behavior in visual search tasks (Henderson, Brockmole, Castelano, & Mack, 2007; Underwood & Foulsham, 2006; Underwood, Foulsham, van Loon, Humphreys, & Bloyce, 2006) or tasks in which the gaze supports navigation and goal-oriented behavior (Ballard & Hayhoe, 2009; Rothkopf et al., 2007). It has been proposed that bottom-up saliency could reflect surprise toward statistical outliers in space (Tatler, Hayhoe, Land, & Ballard, 2011) and therefore attract the gaze when relevant for the task. Considering that conspicuity-based models do not fully explain eye movements in natural behavior, other sources of information must be considered. For instance, throughout his or her life,

an observer memorizes knowledge about what objects look like and how they interact with each other and uses this knowledge to guide eye movements.

Searching for a specific object in a known scene can be made easier by memories of where this object is usually placed (Brockmole & Henderson, 2006; Brockmole & Le-Hoa Vo, 2010). Even in unknown scenes, knowledge about the laws of physics, cultural practices, or the object's function governing the object's location in the environment still apply and can be exploited. Thus, a target object located somewhere at which there is a high probability of finding it is detected and recognized more quickly than when it is placed somewhere at which there is a low probability of finding it (Malcolm & Henderson, 2010). It applies when the target is both consistent and inconsistent with the global layout of the scene (Castelano & Heaven, 2011). Hence, whether the scene is a kitchen or a bathroom, a mug would be found more easily on a sink or washbasin than on the floor. These expectations are built on a variety of sources of information, such as prior knowledge about the global layout of the scene (Castelano & Heaven, 2010; Castelano & Henderson, 2007; Pereira & Castelano, 2014), the laws of physics (Spotorno, Malcolm, & Tatler, 2014), or local information based on the relationships between the target and the other objects (Eckstein, Drescher, & Shimozaki, 2006; Pereira & Castelano, 2014; for reviews, see Vö & Wolfe, 2015; Wu, Wick, & Pomplun, 2014). Computational models that integrate contextual guidance predict eye fixations in natural scenes better than models implementing only bottom-up saliency (Ehinger, Hidalgo-Sotelo, Torralba, & Oliva, 2009; Kanan, Tong, Zhang, & Cottrell, 2009; Torralba, Oliva, Castelano, & Henderson, 2006).

How expectations stemming from the target's location influence eye movement guidance during visual search tasks has been extensively studied. However, the relationship between visual exploration and the higher-level decision-making process to which it contributes has received less attention. Although the purpose of the process is to decide whether or not the target is present in the environment, visual exploration with eye movements can be seen as a process of active evidence collection (hypothesis testing) in which each saccade reduces perceptual uncertainty and consequently enhances the quality of the choice (Friston et al., 2012). In the field of visual identification, this idea has recently been studied using ambiguous drawings representing two alternative objects or animals (Kietzmann et al., 2011) or ambiguous schematic figures interpolated between prototypical animals (Quinton et al., 2014). These two studies show how active eye behavior can disambiguate a figure with the former using an eye-tracking approach and the later a computational approach coupled with a mouse-track-

ing task. More specifically, Quinton et al. (2014) used a binary forced-choice task consisting of categorizing (un)ambiguous prototypical animals. They showed how the two alternative responses attracted mouse movement trajectories produced by participants in a way that is proportional to putative “votes” for the alternative responses at each time unit and used a computational model implementing predictive coding principles (Rao & Ballard, 1999) to reproduce analogous choice trajectories and next gaze location on the stimulus (see also Catenacci Volpi, Quinton, & Pezzulo, 2014).

This example illustrates how, by tracking mouse movements or hand movements during a choice, it is possible to measure distinct aspects of the underlying perceptual decision-making process in a binary forced-choice task (Dale, Kehoe, & Spivey, 2007; Freeman & Ambady, 2010; Song & Nakayama, 2009; Spivey, Grosjean, & Knoblich, 2005). In a typical mouse-tracking task, participants have to categorize a visual stimulus by clicking on the response buttons located in the top left and top right corners of the screen as mouse trajectories are recorded. Therefore, unlike with basic binary forced-choice tasks, the decision-making process is not only inferred through the response times (RTs) and error rates and also through various trajectory measurements whereby continuous aspects of the decision-making process can be captured (Freeman, Dale, & Farmer, 2011; Lepora & Pezzulo, 2015). Among the basic mouse-tracking measurements, the degree of deviation of each trajectory, estimated as the geometric area (area under the curve, AUC) between the actual trajectory and ideal response movement (line stretching from the start point to the accurate response), can be used as a global measure of the bias toward the alternative response in the decision-making process (Freeman & Ambady, 2010). Furthermore, trajectory deviation and mouse velocity can be analyzed as a function of time, providing useful information about how competition between alternatives unfolds over time (Farmer, Cargill, Hindy, Dale, & Spivey, 2007; Freeman & Ambady, 2011; Freeman, Ma, Han, & Ambady, 2013).

The mouse-tracking methodology has been used extensively with visual material, especially to study facial recognition of gender (Freeman, Ambady, Rule, & Johnson, 2008; Johnson, Freeman, & Pauker, 2012) and ethnicity (Freeman & Ambady, 2011; Freeman, Pauker, Apfelbaum, & Ambady, 2010) as well as the influence of stereotypes on face recognition (Freeman & Ambady, 2009; Freeman, Penner, Saperstein, Scheutz, & Ambady, 2011). It has also been used in the fields of numerical cognition (Faulkenberry, 2014; Marghetis, Núñez, & Bergen, 2014), psycholinguistics (Barca, Benedetti, & Pezzulo, 2016; Barca & Pezzulo, 2012, 2015; Farmer et al., 2007; Morett & MacWhin-

ney, 2013), and memory (Papesh & Goldinger, 2012). In visual categorization, the effectiveness of this methodology was assessed for discrimination of ambiguous categories of animals (Dale et al., 2007) and ambiguous visual stimuli representing animals (Quinton et al., 2014).

In the present study, we studied visual search by combining mouse-tracking and eye-tracking methods in order to measure aspects of, respectively, dynamical decision-making and collection of perceptual evidence. Participants had to indicate whether a target was present or absent from an indoor scene. We manipulated expectations of target location by placing it at usual or unusual locations. Noise was added to the images in order to degrade the scene and generate bottom-up visual uncertainty. To ensure that the noise would not disturb the perception of the scene context, we presented a scene preview of a smaller size (but without the target) before the visual scene itself. This allowed the participants to build a prior representation of the scene that was equivalent across noise levels (Castelhano & Henderson, 2007). We divided the visual search into two epochs defined through eye movements. The detection epoch stretches from the appearance of the scene to the first saccade landing on the target’s area of interest (AOI), and the verification epoch includes the first sequence of fixations within the target AOI. The total duration of the former (detection time or target latency) and later (verification time) can be interpreted as measurements of the time spent, respectively, finding and identifying the target (Castelhano, Pollatsek, & Cave, 2008; Malcolm & Henderson, 2009). It is to be noted that if there is no direct fixation of the target before the end of the detection epoch, the target can eventually be noticed peripherally around the end of this epoch. Despite this, detection time remains an indicator of the time spent finding the target because the next fixation after peripheral detection will be directed at the target most of the time, ending the detection epoch. Our objectives were to address the respective roles of target detection and verification in explaining the influence of both expectations about the target location and scene degradation on the decision-making process.

In a recent study, we examined how both expectation of a specific target and visual degradation of a road scene by fog contribute to the decision process, which leads to target identification by affecting detection time and verification (Quétard et al., 2015). We used mediation analysis (Preacher & Hayes, 2008) to assess whether the effects of expectations or fog on a mediating eye movement variable (detection or verification time) could account for the variations in the mouse-tracking variables. We observed that scene degradation caused by fog led to longer detection and verification times, and these effects delayed the

decision-making process and increased the deviation of the trajectory toward the incorrect response. Expecting the correct target led to shorter verification times, resulting in quicker decisions with a reduced bias toward the inaccurate response. However, we limited our analyses to two broad measurements of mouse responses (AUC and time when the largest mouse deviation was reached) whereas mouse-tracking tasks allow for a richer analysis of mouse responses. Moreover, this previous study presented a two-alternative, forced-choice task (identification of the stimulus within a visual scene). However, finding evidence that points toward one object or category rather than another does not involve the same cognitive processes as accumulating evidence of the *presence* or *absence* of a stimulus as necessary for visual search tasks. In terms of the evidence accumulation process, the former would require checking for distinctive features permitting the identification or categorization whereas the latter would require exploring the visual scene sufficiently until the target is found or enough confidence has been acquired that the target is not present.

At a methodological level, this new study specifically focuses on the effects of scene degradation caused by noise and expectations about the target location (rather than about target identity) on eye movements and mouse responses with a view to assessing active collection of evidence and perceptual decision-making in visual search. We analyzed RTs and AUC using mediation analysis and extended the analysis of the mouse responses to velocity and trajectory deviation across time (Castelhamo & Henderson, 2007). In line with previous studies, we predicted that unusual target locations would increase the target detection time and number and duration of fixations during the detection epoch (Castelhamo & Heaven, 2010, 2011; Malcolm & Henderson, 2010; Pereira & Castelhamo, 2014; Spontorno et al., 2014). These studies found mixed results for the influence of target location expectations on verification times. Visually degraded scenes should also increase both detection and verification times as it has been observed that noise increases fixation durations while reducing the number of fixations on the scene and impairs the spatial distribution of fixations on the scene with shorter saccades and a higher central bias (Röhrbein, Goddard, Schneider, James, & Guo, 2015). As the active oculomotor behavior aimed at exploring the visual scene is seen as an evidence-collection process (Friston et al., 2012; Kietzmann & König, 2015; Kietzmann et al., 2011; Quinton et al., 2014), impairment of target detection and verification by both noise and targets in unusual locations should impair the decision-making process as measured through mouse movements. Hence, mouse movements should be slowed and RTs longer, and there should be a bias toward the *absent* response. The basis for this

hypothesis lies in the intrinsic asymmetry between deciding whether a target is *present* or *absent*. This choice can be fit in with the popular evidence-accumulation framework for decision-making (Ratcliff & McKoon, 2008) if it is assumed that accumulating evidence for the *present* response rests on finding a target with a known identity whereas the *absent* response aims to build up confidence that the target is not in the scene. Consequently, there should be heightened competition between *present* and *absent* responses insofar as the *absent* alternative has more weight in the decision-making process over time because the target is harder to find. As target detection and verification permit this competition to be resolved, they should influence this competition. One important assumption in mouse-tracking paradigms is that competition between the two alternative responses, here target *present* or *absent*, unfolds parallel to the mouse movement (Freeman, Dale et al., 2011; Song & Nakayama, 2009). As this competition is updated as new evidence is collected through eye movements, studying the time course of measurements, such as movement velocity and deviation toward the *absent* response with regard to when the target is detected and verified, should highlight the effects of target location and noise. In other words, we should observe these effects even before target detection and during target verification.

Method

Participants

A total of 65 psychology students (including eight males) from Blaise Pascal University took part in the experiment. They received academic credits as compensation for taking part and gave their written consent. Their average age was 20.4 ($SD = 4.1$). All of them had normal or corrected-to-normal vision (glasses or contact lenses).

Stimuli

To build the stimuli set, we selected 60 indoor scene pictures (from the categories bathroom, living room, bedroom, child's bedroom, dining room, office, and kitchen) from the SUN database (Xiao, Hays, Ehinger, Oliva, & Torralba, 2010) and 60 object pictures from the POPORO database (Kovalenko, Chaumon, & Busch, 2012). The scene pictures were scaled at a resolution of 1024×768 pixels. Each target object was associated with a semantically congruent scene and placed either at a usual or an unusual location in the

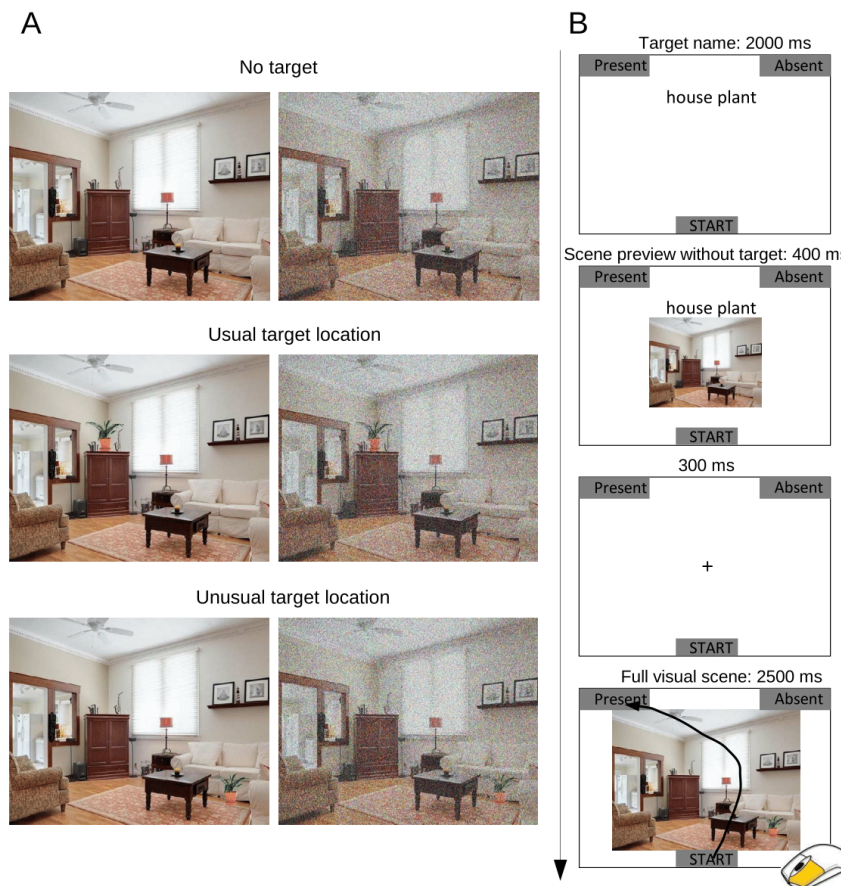


Figure 1. (A) Versions of a scene in each experimental condition. The scene either includes or does not include the target. The target is a house plant, displayed in a usual location and an unusual location. Left: without noise; right: with noise. (B) Time course of a trial.

scene or were absent from the scene (see Figure 1A). To determine the positions of each target, we controlled the laterality of the two target positions; for example, when the target's usual position was in the left part of the scene, its unusual position was in the right part and vice versa. We balanced the laterality of the targets across scenes by choosing usual and unusual locations on opposite sides of each scene. In approximately half of the scenes, the usual location was on the left side whereas the unusual location was on the right side and vice versa for the other half. Finally, we controlled for the central bias (Tatler, 2007) by choosing two approximately equidistant positions away from the center of the picture. For each location, the targets were the same size.

To assess the validity of the target positions, we pretested the scenes. We asked a total of 26 participants

to rate the likelihood of finding the target at this position in the scene (0 = not likely, 100 = very likely). We presented all of the scenes one by one in the same random order for half of the participants and in the opposite order for the other half. This method ensured that, for each scene, when half of the participants were presented with the usual location version first, followed by the unusual location, the other half saw the unusual location version first and the usual location version second. Targets positioned in a usual region ($M = 85.18$) were rated more likely to be found in this position than targets positioned in an unusual region ($M = 10.45$) as assessed using a repeated-measures ANOVA performed on the scenes as observations, $F(1, 59) = 1,314.65$, $p < 0.001$. Finally, we made sure that the rating of usual and unusual target positions differed within scenes by computing the confidence intervals for

each scene. The usual and unusual ratings were different for all the scenes with a confidence level of 99% except for one scene that had a 95% confidence level.

One scene had to be removed because of an error when building it, which we noticed after a quarter of the participants had performed the experiment. We used an algorithm to compute mean and maximal saliency within the target's AOI (Birem, Quinton, Berry, & Mezouar, 2014, adapted to the task specifics and derived from Frintrap, 2006; Itti et al., 1998) in order to check that the bottom-up saliency of the targets did not differ statistically between their two possible locations. There was no difference between usual and unusual locations for either mean, $F(1, 58) = 0.51$, $p = 0.477$, or maximal saliency, $F(1, 58) = 0.49$, $p = 0.485$.

To manipulate the visual uncertainty of the scenes, we degraded the pictures by adding a Gaussian noise to each pixel (independently on each RGB color component). The noise distributions were centered with a variance of 0.3 (for an intensity range of [0, 1]). Our final stimuli set totaled 354 pictures (59 scenes \times 3 target conditions \times 2 noise levels). We added 59 scene previews that did not include the target and were scaled at 600×450 pixels.

Apparatus

We used the MouseTracker software (Freeman & Ambady, 2010) to program and run the experiment. The stimuli were presented using a laptop with a 15.6-in. display set at a resolution of 1600×900 pixels and a refresh rate of 60 Hz. The participants were instructed to maintain their head at an approximately constant distance from the screen (61 cm). An ordinary mouse was used, and the movements were recorded by the MouseTracker software at a sampling rate of approximately 60 Hz.

Eye movements were recorded using a SMI RED-m head-free eye-tracker at a sampling rate of 60 Hz. We preprocessed the data using SMI's dispersion-based algorithm for fixation identification based on Salvucci and Goldberg (2000). We set the minimum fixation duration threshold at 80 ms and the minimum dispersion threshold at 1° (within the optimal range tested by Blignaut, 2009).

Procedure

The experiment began with a five-point eye-tracker calibration procedure and a four-point validation to establish the calibration error. Participants then completed a 16-trial training block, including natural scenes

and target animals rather than indoor scenes and target objects. Figure 1B shows the time course of a trial. To begin each trial, participants had to click with the mouse cursor on the *start* button at the bottom center of the screen with the mouse's cursor. The name of the target object appeared for a total of 2400 ms. After 2000 ms, a preview of one of the scenes appeared in the center of the screen below the target name for the remaining duration (400 ms). Then, after a blank screen with a 300-ms fixation cross, the same scene was presented for 2500 ms in one of the six conditions previously listed: degraded with noise or not, with or without a target presented in a usual or unusual location. The mouse cursor was locked in the middle of the *start* button before the scene appeared to ensure all the trajectories started from the same location. Participants were instructed to click on the *present* or *absent* response buttons in the top left and top right corners of the screen, according to whether the target was in the scene or not. If 1000 ms elapsed before they moved the mouse, a message was displayed after completion of the trial, urging them to begin their movement earlier in the upcoming trials. A response given after 2500 ms was invalid resulting in a "time out" message.

Each participant was presented with 59 trials, including 10 trials per experimental condition (3 target conditions \times 2 noise levels, nine trials for one condition due to the later removal of one scene) with only one version of each scene being presented. For instance, a participant being presented a visual scene depicting a specific kitchen with no target and no noise could not be presented the same kitchen in another experimental condition. Thus, the different versions of each scene were counterbalanced across participants using Latin squares, resulting in the formation of six sets of trials each presented to a subgroup of participants. The laterality of the response buttons was also counterbalanced with approximately half of the participants in each subgroup being presented with a leftward *present* button and a rightward *absent* button, and the other half were presented with the opposite button positions.

Results

Data processing

Eight participants were excluded from the analysis for the following reasons: invalid eye-tracking data due to glasses or lenses making the calibration impossible (five) and a high number of "time out" trials combined with weak calibration (three). We analyzed only the trials in which the target was present because our eye-

movement variables could not be calculated in target-absent trials. We excluded the trials with invalid eye-tracking data: In 4.1% of the trials, gaze was directed toward the target when the scene appeared, in 15.5% there was no fixation on the target, and in 1.4% the detection or verification time was less than 80 ms or more than 2500 ms. We then discarded trials in which the participant's response exceeded the time limit (5.6% of the trials with valid eye-tracking data) or in which he or she clicked on the wrong response (8.4%). Overall, the analyses were performed on 68% of the trials, including a target (1,491 trials). The low percentage of valid trials is not surprising because it was not possible to instruct the participants to fixate the target directly when responding because they were involved in a dynamic motor task in which they had to control their mouse movement.

We remapped the correct response buttons on the right side and the alternative on the left side of the screen, and we rescaled the x -axis between -1 and 1 and the y -axis between 0 and 1.5 in accordance with traditional mouse-tracking literature (Freeman & Ambady, 2010). Given the mouse-tracker time resolution, the time interval between each point on a trajectory was 16 ms. As eye-tracking measurements, we calculated the detection and verification times as well as the mean duration of fixations and number of fixations during the detection epoch. As temporal measurements of the dynamics of mouse responses, we calculated the RTs and velocity profiles of the trajectories. As measurements of the spatial deviation of the mouse toward the *absent* response, we calculated the AUC and mouse deviation profiles along the x -axis.

To assess the effects of the target location (usual vs. unusual location) and noise (no noise vs. noise) on the eye-tracking and mouse-tracking measurements, we used linear mixed modeling (LMM), which can take account of the variability of the parameters across participants and visual scenes simultaneously without having to perform separate *by-subject* and *by-item* analyses (Baayen, Davidson, & Bates, 2008). In our study, *by-item* analysis corresponds to *by-scene* analysis. This is interesting in our experiment because we used Latin squares to distribute the trials of each condition across both scenes and participants. Furthermore, these models are particularly well suited to our data because they enable efficient handling of repeated-measures data sets with imbalanced data and missing values. When required, we included the target location and noise as well as their interaction as fixed factors. We recoded the location variable as -0.5 for usual and $+0.5$ for unusual probability locations and the noise variable as -0.5 without noise and $+0.5$ with noise. As random parameters, we included the variance of the intercepts and slopes as well as the correlations between random parameters for both the subject and scene variables.

However, due to nonconvergence of the model testing the RTs, we took out the slope variances for the subject variable after ensuring they were nonsignificant. In that way, we used the maximum random structure allowing the models to converge (Barr, Levy, Scheepers, & Tily, 2013). To estimate the p values of the fixed effect parameters, we used the Kenward-Rogers approximation for the degrees of freedom. The “lme4” R package was used to compute the LMM and the “pbkrtest” R package to estimate the p values.

Eye movement measurements

Eye movements during the detection and verification epoch were analyzed. The detection epoch included all the fixations on the scene (including the latency of the first saccade) before the first saccade on the target AOI. Three measurements were analyzed for this epoch, namely total detection time, the number of locations fixated, and the mean duration of the fixations. The verification epoch spanned from the first fixation on the target's AOI to the first saccade ending outside the target AOI. Following the example of Malcolm and Henderson (2009, 2010), if the fixation following the first fixation on the target was not on the target but was followed by a fixation on the target, it was considered part of the verification epoch because this fixation could be part of a preprogrammed sequence of saccades. Verification time was computed as the sum of the durations of the fixations within the target AOI and included in the verification epoch.

Detection epoch

See Table 1 for mean detection times, number of fixations, and fixation durations by experimental condition. Targets positioned in an unusual location resulted in significantly longer detection times, $b = 204.59$, $t(62.8) = 8.44$, $p < 0.001$, and more fixations during this epoch, $b = 0.76$, $t(51.9) = 8.86$, $p < 0.001$, than when the target was in a usual location. No main effect of noise was found on detection times, $b = 13.03$, $t(27.5) = 1.12$, $p = 0.27$, or on the number of fixations, $b = -0.06$, $t(31.9) = -0.52$. There was no interaction between the two variables on detection times, $b = -20.77$, $t(1308) = -0.93$, but a marginal interaction was recorded for the number of fixations, $b = -0.22$, $t(1342.2) = -1.92$, $p = 0.056$. Therefore, positioning a target in an unusual location increased the number of fixations slightly less in a scene degraded with noise than in a scene with no noise, but the difference did not reach the significance threshold. Finally, mean fixation durations contrasted with these results: The only effect found with this measure was that fixation durations increased for a noisy scene, $b = 10.60$, $t(36.3) = 2.70$, $p =$

	Usual location		Unusual location	
	Without noise	With noise	Without noise	With noise
Detection time, ms	388.9 (9.9)	406.9 (10.8)	570.4 (15.2)	557.8 (15.4)
Fixations number	1.74 (0.04)	1.78 (0.05)	2.51 (0.07)	2.31 (0.06)
Fixations duration, ms	199.6 (3.1)	210.0 (3.7)	200.8 (3.4)	211.6 (3.7)
Verification time, ms	336.9 (6.7)	387.2 (9.4)	317.3 (6.8)	374.5 (9.6)

Table 1. Means (standard errors) for each eye movement measurement by target location and noise conditions.

0.011, whereas there was no main effect of target location, $b = 1.45$, $t(42.8) = 0.58$, and no interaction, $b = -0.82$, $t(1321.4) = 0.06$.

Verification epoch

Verification time was the only measurement used to characterize the verification epoch (see Table 1 for means). The only effect found was an effect of noise, $b = 63.84$, $t(34.2) = 7.87$, $p < 0.001$, with longer verification times when the scene was degraded by noise. No effect of target location, $b = -12.6$, $t(49.1) = -1.10$, $p = 0.277$, and no interaction between target location and noise, $b = 7.73$, $t(1320.3) = 0.56$, were found.

RTs and temporal dynamics of the mouse response

In order to analyze the dynamics of the responses, we first used mediation and moderation analyses to investigate how target location and noise affected the RTs and how the detection and verification times influenced location and noise effects on the RTs. Second, we further analyzed the dynamics of the responses, focusing on the time course of the velocity profiles of the mouse trajectories in relation to target detection and target verification.

Response times

The RT spanned from the appearance of the scene to the response click. By making the target harder to find, both scene degradation and unusual target locations should result in longer RTs. We used mediation and moderation analyses to assess whether the detection and verification times affected the effects of target location and noise on RT. The detection or verification times moderated the effects of the independent variable (IV), i.e., target location or noise, on the RT when there was an interaction between the IV and the moderator. Thus, depending on the moderator values, the effect of the IV on the RT will vary. On the other hand, with mediation analyses, it is possible to estimate whether

the effect of the IV on the mediator explained part (or all) of the effect of the IV on the RT. As there was an effect of target location on detection time and an effect of noise on verification time, the effect of target location on RT should be mediated or moderated by detection time whereas the effect of noise on RT should be mediated or moderated by verification time.

We assessed the mediative effects of detection and verification times using a bootstrapping procedure (Preacher & Hayes, 2008). It consists of estimating the indirect effect of the IV on the RT through the mediator. This indirect effect AB is calculated as the product of the IV's effect on the mediator (path A) multiplied by the mediator's effect on the RT while controlling for the effect of the IV (path B). The 95% confidence interval of this effect was assessed with a nonparametric bootstrapping procedure comprising 10,000 samples. To measure the size effect, we estimated the mediated proportion of the IV's total effect (path C). This can be done by calculating the proportion by which the IV's effect is reduced when controlling for the effect of the mediator (path C). Although this measurement is not perfect, it is a simple way of obtaining information about the effect size (for a short discussion, see Hayes, 2009). We estimated these parameters using LMMs with both IVs and the mediator (when needed) as well as all the possible interactions as fixed effects. We specified the same random structure for all models, i.e., the variance of the intercept and slopes of the IVs and the correlations between random parameters nested within the subject and scene random factors.

Figure 2A displays the mean RT for each condition, and Figure 2B displays in schematic form how target detection and verification times mediate the effects of target location and noise on RT. Our LMMs revealed a main effect of both noise, $(b = 71.27, t(50.9) = 4.7, p < 0.001)$, and target location, $(b = 148.69, t(55.5) = 6.07, p < 0.001)$, on RT. No interaction was found, $b = -28.58, t(1361.1) = -0.99$. Responses were faster when the scene was not degraded by noise as opposed to degraded and when the target was placed at a usual location as opposed to an unusual one. Detection times mediated the effect of target location on RT. When we controlled for the effect of detection time, the sign of the parameter of target location expectation (direct effect

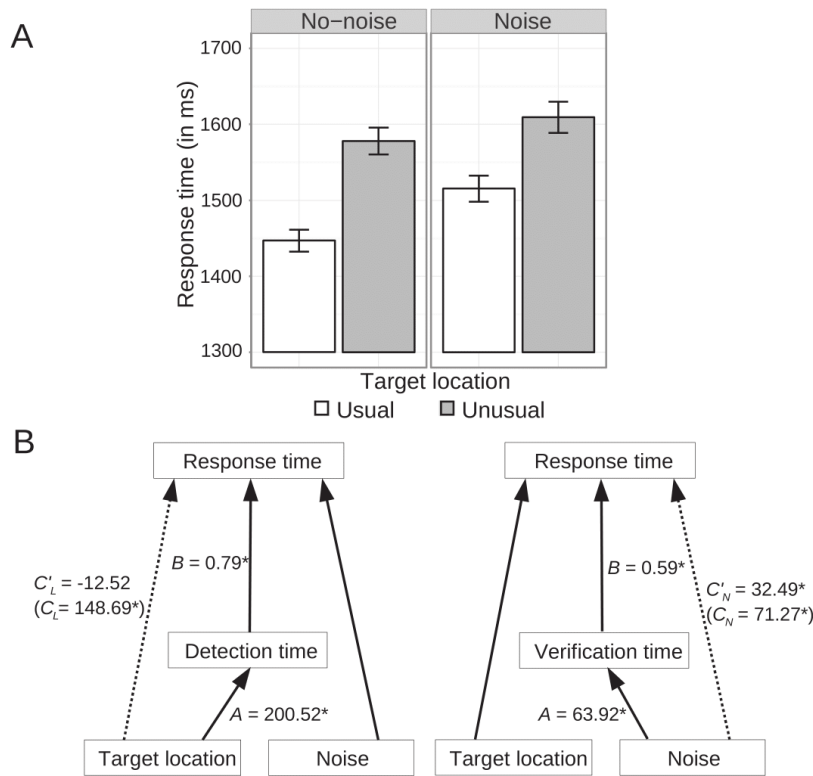


Figure 2. (A) Mean RTs as a function of target location and scene degradation. White bars: target in usual location; gray bars: target in unusual location. Error bars denote the standard errors. (B) Mediation of the effects of target location and noise by the eye movement measurements, i.e., target detection and verification times. Dashed arrows denote an effect mediated by detection or verification times. The model parameters are estimated with LMMs (see Results). When an effect is mediated, parameters A and B are displayed (indirect effect = $A \times B$) as well as the direct effects C' (effect of the independent variable when we controlled for the effects of the mediator) and total effect C (effect of the independent variable without including the mediator in the model); L indices: effect of target location; N indices: effect of the noise. *The parameter is significant ($p < 0.05$).

C') was reversed and lost significance, $b = -12.52$, $t(61.5) = -0.75$, and the indirect effect of target location expectation through detection time was significantly different from 0 ($AB = 159.40$, 95% bootstrap CI [135.15, 185.62]). When detection times are included in the model as a predictor, the target location effect disappears. Conversely, when we controlled for the effect of verification time, the direct effect of noise on RT was reduced by 54.4% and remained significant, $b = 32.49$, $t(55.6) = 2.19$, $p = 0.032$. The indirect effects of noise through verification time was significant ($AB = 38.14$, 95% CI [27.76, 52.17]). Finally, the indirect effect of target location through verification times and of noise through detection times were not significant (respectively, $AB = -7.60$, 95% CI [-17.52, 1.21] and

$AB = 10.66$, 95% CI [-9.77, 28.72]). Thus detection time mediated the effect of target location on RT whereas verification time mediated a large part of the effect of noise.

Velocity profiles

We further analyzed the dynamics of the competition between responses by means of mouse velocity as a function of time. Velocity profiles can estimate the beginning of the response movements as well as how the commitment toward the correct response builds up as time unfolds (Hehman, Stolier, & Freeman, 2015). Our underlying hypotheses are that greater difficulty finding and identifying the target due to unusual

location and/or noise should reduce the commitment toward the *present* response, hence reducing overall velocity. We analyzed the time course of velocity in relation to the moments when the target was detected (end of the detection epoch, corresponding to the detection time) and the end of target verification (end of the verification epoch), thus examining how it evolves during the detection and verification epochs. Finally, because accumulating evidence during the detection and verification epochs should affect the ongoing decision, mouse movements should begin during the detection epoch (and, thus, before the target is properly fixated). To perform this analysis, for each trajectory, we centered the time coordinates on the end of the detection epoch by subtracting the detection time from each time coordinate. Thus, the 0-ms time indicates approximately the moment along the response trajectory when the target was detected, referenced later on as target detection. The interval between each time stamp was 16 ms (~60 Hz time resolution). We measured the velocity for each time step of each trajectory. Then, we smoothed the velocity profiles of each trajectory using a convolution with a triangular kernel over a sliding window of 21 time steps. This smoothing step is required on account of the stochasticity of mouse events on most computer devices. This method allowed us to analyze the velocity prior and posterior to the moment when the target was detected, and it also allows for raw time analysis on a data set in which mouse movement onsets vary within a large time interval (2500 ms). More commonly used alternatives include time-normalizing the trajectories (Hehman et al., 2015) at the cost of losing part of the time information whereas our key objective was to study the raw timing of the responses compared to the target detection and target verification.

Figure 3 displays the mouse velocity as a function of time centered on detection of the target for each experimental condition. For each time step, the velocity was estimated by computing one intercept-only LMM (testing velocity against 0) *per* experimental condition. These LMMs included the intercept variations across both participants and scenes as random parameters. As inferential statistics, we computed the 95% Wald confidence intervals from the LMMs (shown in Figure 3). For targets in a usual location, velocity peaks occurred, respectively, for no noise and noise conditions, 528 ms and 560 ms after target detection and 191 ms and 173 ms after the end of target verification (estimated velocity peak minus mean verification time, shown as solid and dashed vertical lines in Figure 3A, B). For targets in unusual locations, they occurred 480 ms after target detection (for both no noise and noise) and, respectively, 163 ms and 106 ms after the end of target verification. The velocity peak of the profiles was higher when the scenes were not degraded by noise than

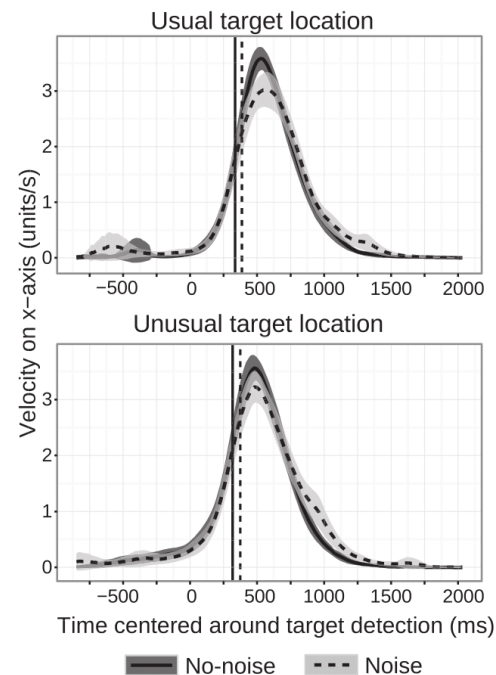


Figure 3. Mouse velocity as a function of time for scenes nondegraded by noise (solid lines, dark gray) and scenes degraded by noise (dashed lines, light gray). Top: target in a usual location; bottom: target in an unusual location. The x-axis represents time centered on detection of the target; 0 represents the moment the target was first fixated. Velocity and the error intervals (95% Wald confidence intervals) were estimated at each time frame using LMMs. On each plot, the black vertical lines denote the average moments when target verification ended (verification time). No noise: solid lines, noise: dashed lines.

when they were degraded irrespective of whether the target was in a usual (no noise: $b = 3.58$, 95% CI [3.38, 3.79]; noise: $b = 3.03$, 95% CI [2.71, 3.34]) or an unusual location (no noise: $b = 3.56$, 95% CI [3.34, 3.79]; noise: $b = 3.22$, 95% CI [2.94, 3.51]). The velocity was higher for nondegraded scenes from 416 ms to 592 ms (12 steps) after target detection for targets in a usual location and from 448 ms to 560 ms (eight steps) for targets in an unusual location. However, there was no effect of target location on the velocity around the velocity peak. A result of the reduced velocity for degraded scenes is that the end of the mouse movement exhibited a significantly higher velocity than for nondegraded scenes from 1104 ms to 1440 ms (22 steps) for targets in usual locations and from 816 ms to 1552 ms (47 steps) for targets in unusual locations. Finally, we

found evidence that mouse velocity began to increase even before target detection (0 on the x -axis of Figure 3). Overall, for targets in usual locations, velocity was significantly greater than 0 up to 272 ms before target detection and increasing from 0.056 units/s to 0.106 units/s once the target was detected. For targets in unusual locations, the velocity was significantly above 0 up to 640 ms before target detection and increased from 0.069 units/s to 0.338 units/s when the target was detected. Furthermore, when the scene was not degraded, the velocity for targets in unusual locations was significantly higher than for targets in usual locations from 288 ms before target detection to 416 ms after target detection (45 steps). When the scene was degraded, the velocity for targets in unusual locations was significantly higher than for targets in usual locations from 128 ms before target detection to 496 ms after target detection (40 steps). This indicates that, in conditions in which target detection is harder due to unusual target locations, thus leading to longer detection times, the response movement was initiated even before the first fixation on the target.

Mouse trajectory spatial deviation

Measuring the deviation of the mouse movements from the ideal response trajectory is a way of estimating the attraction toward the incorrect *absent* response in the decision-making process. As an overall deviation measurement, we calculated the AUC. In a similar way as we did for RT, we used mediation and moderation analysis to estimate the contribution of target detection and verification to the effect of target location expectations and noise on the AUC. Finally, we studied spatial deviation over time centered on target detection by analyzing the deviation on each non-normalized time step, using a similar method to the one used for the velocity profiles.

Area under the curve

The AUCs were calculated after time-normalizing the trajectories into 101 time steps (as usually done in mouse-tracking experiments; Freeman & Ambady, 2010). Making the target harder to find by means of both scene degradation and unusual target locations should increase the AUC. This implies that the difficulty finding a target should increase the competition between *present* and *absent* responses, and because target detection and verification should play parts in resolving this competition, they should mediate or moderate these effects. As with the RT analyses, the effects of target location should be mediated/moderated by detection times as only target location had an effect on detection times, and respectively, the effects of

noise should be mediated/moderated by verification times. Figure 4A displays the mean AUC for each condition and the averaged mouse trajectories for each target location condition. Only the main effect of target location on AUC was significant, $b = 0.22$, $t(49.8) = 3.35$, $p = 0.002$. Deviation from the ideal response trajectory was greater when the target was in an unusual location. No noise effect nor interaction was found, respectively, $b = 0.07$, $t(29.6) = 1.59$, $p = 0.12$ and $b = -0.04$, $t(1339.3) = -0.50$. Mediation and moderation analyses revealed that verification time had an effect on AUC, $b = 0.0005$, $t(1151.5) = 3.05$, $p = 0.002$, but did not moderate or mediate the effect of target location. On the other hand, when detection times were included in the model, the effect of target location was no longer significant, $b = -0.046$, $t(39.3) = -0.98$, $p = 0.333$. Detection time interacted significantly with target location, $b = 0.0007$, $t(907.7) = 4.13$, $p < 0.001$. Figure 4B displays this moderation model. Detection times increased the deviation toward the inaccurate response to a greater extent when the target was in an unusual location as opposed to a usual location.

Deviation on x -axis over time

To investigate the relationship between the trajectory deviation and the detection times further, we studied the deviation on the x -axis (x -deviation) prior and posterior to the moment the target was detected. Given that only target location affected the AUC, we focused on the x -deviation only on this variable. As the alternative responses are spatially distinguished on the x -axis, an analysis of x -deviations can index the ongoing competition across time. For example, greater deviations can indicate that the incorrect response is considered a valid alternative. As with velocity, we analyzed the time course of x -deviation in relation to when the target was detected (detection time) and the end of target verification (end of the verification epoch). We hypothesized that an unusual location should increase the deviation. As accumulating evidence toward an *absent* response is aimed at accumulating confidence that the target is not in the scene, the x -deviation should accordingly begin to increase during the detection epoch as it is during this epoch that the *target-absent* response competes with the *target-present* response. As we did for velocity profiles, for each trajectory, we centered the time coordinates on the detection time so that a 0-ms time coordinate indicates the moment on the response trajectory when the target was detected. The x -deviation was the difference between the coordinate on the x -axis of each point of the response trajectory and the x -coordinate of its perpendicular projection point on the ideal response trajectory (straight line from *start* to *present* button). A positive x -deviation value denotes a deviation toward

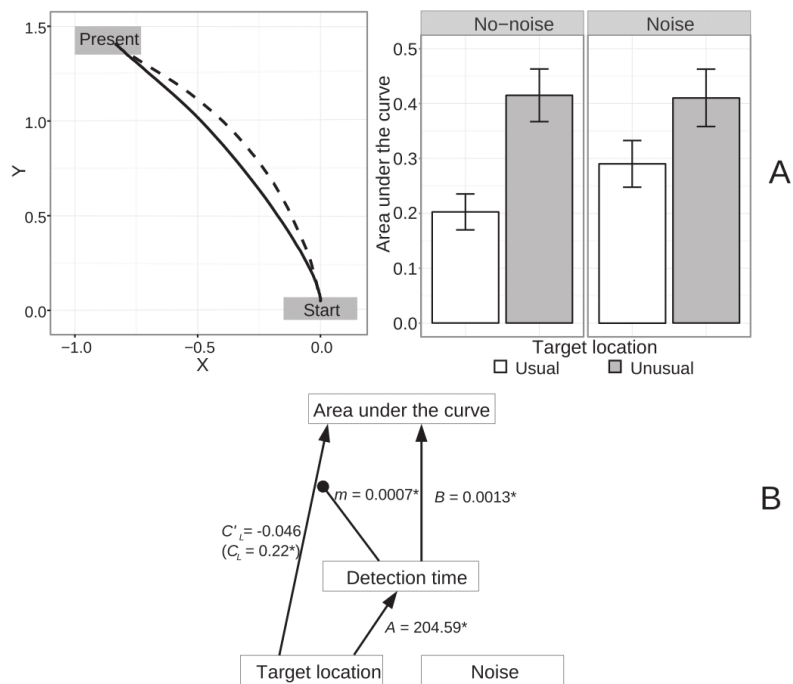


Figure 4. (A) Left: mean mouse trajectories as a function of usual (solid curve) and unusual (dashed curve) target location. Right: AUCs as a function of target location and scene degradation. White bars: target in usual location; gray bars: target in unusual location; error bars denote the standard errors. (B) Moderation of the effects of the target location and noise by the target detection time. The line ending with a black dot indicates a moderation effect with m being its parameter (moderation corresponds to an interaction between target location and detection time). Parameters A and B are displayed together with the direct effect C' (effect of the independent variable when the moderator is included in the model) and total effect C (effect of the independent variable without including the moderator in the model). L indices: effect of target location. *The parameter is significant ($p < 0.05$).

the *absent* response compared to the ideal trajectory. We smoothed the x -deviation profiles of each trajectory by using a convolution with a triangular kernel over a sliding window of 21 time steps. Figure 5A displays the x -deviation as a function of time centered on detection of the target for usual (solid line, dark gray error) and unusual target locations (dashed line, light gray error). The 95% Wald confidence intervals included in Figure 5A were estimated with intercept-only LMMs for each time step. As stated previously, these LMMs included the intercept's variation across both participants and scenes as random parameters. Two observations are relevant here. First, as with the results with the AUC, the overall deviation was greater for targets in an unusual location than in a usual location. This was the case from 304 ms before target detection to 640 ms after target detection (60 steps). For targets in an unusual location, the x -deviation attained its maximum value ($b = 0.115$, 95% CI [0.087, 0.143]) around 384 ms after

target detection, corresponding to the average end of target verification (343 ms, dashed vertical line). For targets in a usual location, the maximum x -deviation ($b = 0.063$, 95% CI [0.045, 0.08]) occurred 560 ms after the target detection, which corresponds to an approximate 200-ms delay after the end of target verification (360 ms, solid vertical line). This result may seem counter-intuitive, but given that the target was detected sooner when in a usual location and thus verified sooner, participants had much more time to finish their movement and validate their response before the time granted to respond elapsed. Second, we observed a significant deviation toward the *absent* response before target detection. The trajectory deviated up to 128 ms before target detection for targets in a usual location and 512 ms for targets in an unusual location. As reported previously, the deviation was significantly greater for targets in unusual locations starting 304 ms before target detection (Figure 5A).

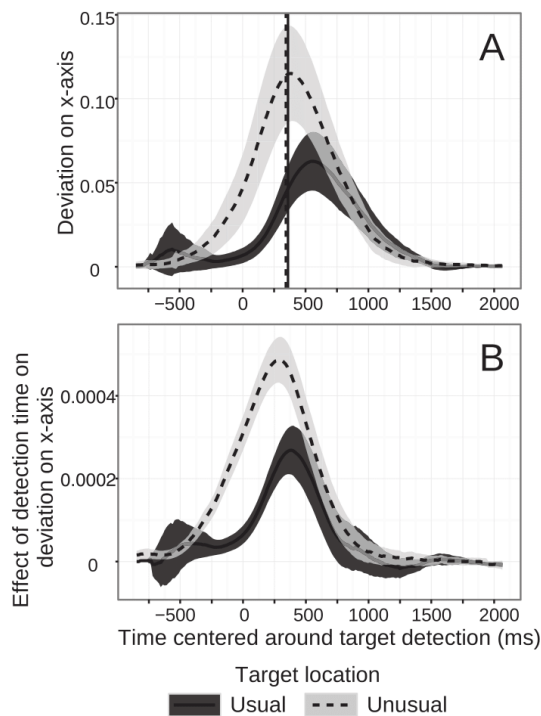


Figure 5. (A) Estimated deviation on the x -axis as a function of time and target location. (B) Estimated effects of target detection times on the x -deviation as a function of time and target location. Solid lines, dark gray: target in a usual location. Dashed lines, light gray: target in an unusual location. The x -axis represents the time centered on the detection of the target; 0 represents the moment the target was first fixated. X -deviations and the effect of detection at each step were estimated using LMMs, as well as the error intervals (95% Wald confidence intervals). The black vertical lines denote the average moments when target verification ended (verification time). Usual target location: solid, unusual target location: dashed.

With mediation and moderation analysis, we previously found an interaction between detection time and target location expectations on the AUC in which the effect of detection time was greater for targets in unusual locations than in usual locations. We examined the time course of the effect of the detection times on the x -deviation across these conditions more closely by plotting the parameters of this effect across time steps for usual and unusual target location (estimated using LMM with the same random parameters as in the last analysis). This type of analysis is similar to the estimation of correlations across time suggested by

Hehman et al. (2015), but we aimed to obtain less biased estimates of the relationship between detection times and deviation by taking account of the nested structure of our data (estimating the variance of x -deviation across participants and scenes). This is displayed in Figure 5B. Overall, detection times had a greater influence on the x -deviations when the target was in an unusual location (max effect: $b = 0.00049$, 95% CI [0.00043, 0.00054]) as opposed to a usual location (max effect: $b = 0.00027$, 95% CI [0.00021, 0.00033]). This is consistent with the interaction previously reported between detection times and target location expectations on the AUC. We also observed an effect of detection times on the x -deviation even before the end of the detection epoch, especially for targets in an unusual location. Hence, in the trials in which the target was detected later, the deviation toward the *absent* response was greater even in the time steps preceding the effective detection.

Discussion

We devised a novel study of visual search and perceptual decision-making, in which participants were required to report on whether a target object was present or absent from a visual scene and in which both the predictability of target location and noise levels were manipulated. To sum up, the main research questions underlying this study addressed (a) how expectations about target location and visual uncertainty due to noise drive the active visual evidence collection process through oculomotor behavior and (b) how these two factors differentially affected the decision about whether the target is present in the scene or absent. Our results (summarized in Box 1) show that targets in usual locations and nondegraded visual information facilitated, respectively, target detection and verification, and consequently, they highlight the differential roles of these two processes in the encompassing decision-making process.

Hence, both eye-movement and mouse-tracking measurements indicate that predictable target locations and absence of noise facilitate visual search. Increasing noise level in the scene impaired target verification and led to longer fixations during the detection epoch. By contrast, placing the target in an unusual location impaired its detection (compared to a usual location) and increased the number of fixations required to detect it but had no effect on target verification. The decision-making process was also impaired by noise and unusual target location. Both manipulations led to lower performance on the overall RTs. Scene degradation by noise reduced overall mouse velocity and increased it in the latest parts of the response

movement whereas targets in unusual locations increased the bias toward the incorrect *absent* response. Finally, our results indicate that detection and verification processes played a role in the decision-making process, respectively, by mediating the effects of target location and scene degradation on RT. Additionally, detection times moderated the effect of target location on response selection such that unusual target locations impaired the AUC to a greater extent for high target detection times. Both velocity and deviation were significantly different from zero for several time steps before the first fixation on the target, especially when it was in an unusual location.

Contribution of contextual knowledge and bottom-up visual information with respect to eye movement

The results for eye movement measurements are consistent with previous studies that investigated contextual guidance. Visual noise did not impair target detection but impaired the verification process as well as the duration of the fixations during the detection epoch in keeping with previous studies using degraded or low-contrast visual scenes (Harley, Dillon, & Loftus, 2004; Röhrbein et al., 2015). In visual search, variations in the verification time may reflect the identification of the target whereas variations in the duration of the fixation in the detection epoch may reflect that more time is required to reject the fixated location as a potential target location. Thus, in our experiment, degraded visual information made it harder to both reject the fixated location and identify the target.

Considering the effect of usual target locations, previous studies (Castelhano & Heaven, 2010, 2011; Malcolm & Henderson, 2010; Pereira & Castelhano, 2014; Spotorno et al., 2014) consistently showed a reduction in overall detection time and the number of fixations during the detection epoch, reflecting the number of locations visited before finding the target. Along with these studies, our results confirm that expectations about target location improve target detection (as is normally the case for targets in usual locations). As extensively discussed in the literature (Brockmole & Le-Hoa Vo, 2010; Ehinger et al., 2009; Torralba et al., 2006), this may be due to contextual knowledge guiding oculomotor behavior in visual search. However, contrasting results have been reported for the verification epoch. Some showed that contextual information facilitated the verification process (Castelhano & Heaven, 2011; Malcolm & Henderson, 2010) whereas others did not (Castelhano & Heaven, 2010; Pereira & Castelhano, 2014; Spotorno et al., 2014) as in the present study. It is to be noted that there were two main differences with these previous

studies due to our mouse-tracking paradigm. First, participants were subjected to considerable time pressure, and second, subsequent eye movements were needed on the response buttons for mouse movement control after the verification epoch. Thus, our study goes against the idea that identifying a target in an unusual location requires longer gaze allocation than a target in a usual location, especially in situations in which time is limited and subsequent eye movements are needed to perform a motor task efficiently. However, spending more time fixating the target in order to identify it or spending more time fixating a location in a scene before rejecting it as a potential target location could still be a valid strategy for reducing uncertainty about the decision in contexts in which the speed–accuracy trade-off is not biased toward quick responses (unlike in our task in which participants had to respond within a specified time limit and begin their response movement early).

Distinct contributions of detection and verification processes upon decision-making

Our results clearly distinguish between the contributions of target detection and identification to the decision-making process, respectively, influenced by target location and noise. Variations of RT due to target location were explained by the effect of target location on target detection time. Note however that we cannot conclude that detection times mediate the entirety of the effect because this claim relies on a null hypothesis (Rucker, Preacher, Tormala, & Petty, 2011), the effect of target location being nonsignificant when controlling for detection times. Conversely, a large proportion of RT variation (54.4%) due to degradation of the visual scene was explained by the effect of noise on target verification time. Thus, these results show that wrong expectations about the target's location and visual uncertainty due to noise both impair decision-making by requiring prolonged collection of evidence at different levels (target detection and verification), thus slowing the decision process. The results of the present study and the fact that we found no interaction between target location and scene degradation with our mouse-movement measurements could indicate that the influence of expectations and visual uncertainty on perceptual decision-making are processed additively and separately by influencing two distinct subprocesses: target detection and verification. In this respect, an analysis of the dynamics of the response movements indicated that the movements began toward the end of the detection epoch when the target had yet to be fixated. As the aim of the task is to find the target, this epoch is critical. Indeed, contrary to classic mouse-tracking experiments using two-alternative forced

choice, the target on which the decision has to be made is not directly and centrally available right from the beginning of the trial but has to be found first. On the other hand, the moment velocity was highest was situated around the end of the verification epoch and shortly after, thus coincidental with fixations made toward the response buttons after visual exploration of the scene. At this stage, degradation of the visual scene by noise was critical. Although we found no strong evidence of noise delaying the velocity peak, the maximum amplitude of the velocity profile was lower whereas velocity was higher in the latest parts of the response movement. This could indicate less commitment to the *present* response with the degraded target being harder to recognize and distinguish from the background, thus generating uncertainty. This would induce a more careful response with lower maximum velocity but faster toward the end to enable the response to be validated within the time constraints.

Relationship between expectations for target location and target detection upon response selection

Contrary to noise, unusual target locations did not reduce mouse velocity compared to usual target locations. Instead, they induced a spatial bias toward the *absent* response throughout a large portion of the response movement, starting before detection of the target (from approximately 304 ms before target detection to approximately 640 ms after target detection, and mean response times lasted around 1450 ms to 1600 ms). In addition, detection times moderated the effect of target location on the AUC. Increasing detection times inflated the mouse trajectory deviation toward the *absent* response to a larger extent for unusual target locations than for usual target locations. Consistent results were found for *x*-deviations across time as the effect of detection times was greater for unusual target locations during most of the response.

This indicates, first, that expectations about the target location influence the final decision about the presence of the target by affecting eye movements made to detect it and, second, that the inflation of the target location effect by detection time could be explained by a combination of two effects: the delay of detection as a result of unusual target location (wrong expectations) and a mechanism whereby certainty about the *target-absent* response builds up as time is spent unsuccessfully searching for the target through eye movements. In our results, we found some evidence of such a mechanism. We found that both mouse velocity and *x*-deviations began to increase before the end of the detection epoch and that the effect was larger and occurred sooner when targets were in unusual locations

(Figure 5A). This indicates growing competition between the *absent* and *present* alternatives beginning before the moment the target was detected. Furthermore, trials with longer detection times showed bigger *x*-deviations even before target detection, especially for targets in unusual locations (Figure 5B). Another way to express it would be to say that detection times correlated with increased deviation before the actual detection of the target. This may indicate that the competition increases with the difficulty finding the target but also while the target is actively searched with eye movements. The longer the visual search is unsuccessful, the more the *target-absent* alternative competes with the accurate *target-present* response and the attraction toward the *target-absent* alternative may reflect an indirect accumulation of evidence stacked in favor of this response owing to the lack of evidence in favor of the *present* response. The effect of detection time on *x*-deviation reached its maximum around the end of the verification epoch, indicating that the time spent fixating the target in order to verify its identity is needed to counteract the evidence that has accumulated toward the *target-absent* alternative and finally converge toward the accurate *target-present* response. The gradual increase in mouse velocity up to target verification consistently supports this interpretation. By extension, with this model, if the target is not found, the gradual commitment toward the *absent* response would finally lead to the decision to stop searching for the target. This could be plausible as a mechanism allowing for search termination (for review, see Wolfe, 2012) as a possible way an agent searching the environment for an object in an area decides to stop and to search in another area.

However, we need to remain careful when considering these conclusions given that, in our experiment, the examination of how target detection and verification relate to mouse movements is based on the incidental measurement of detection and verification times as invoked variables, and thus, only limited conclusions can be drawn about causation. Furthermore, the uncertainty about the visual input was directly manipulated with noise and top-down expectations using usual and unusual target locations, and we found that their effects on eye movements and mouse response movements could be linked. Our study tested how noise and target location influenced eye movements (active collection of evidence), in turn influencing decisions through mouse movements in a feed-forward fashion. What our study does not cover is how the accumulation of evidence toward each alternative itself influences how eye movements are directed to collect further information (see Kietzmann et al., 2011; Kietzmann & König, 2015). Nonetheless, it provides an interesting direction for further studies addressing both the causal relationship between active

vision and decision-making and the search termination problem in visual search tasks. For instance, an important step would be to manipulate detection times and verification times directly in similar mouse-tracking tasks, for instance, with designs inspired by the scene onset delay paradigm (Henderson & Smith, 2009; Nuthmann, Smith, Engbert, & Henderson, 2010). Comparing human data with simulated data using computational models implementing target detection and verification processes could also provide interesting insights. Additionally, because expectations about target location are linked to detection times, it would be important to control for their effects, for instance, by using nonmeaningful targets in real-world scenes, such as white dots as used by Brockmole and Henderson (2006) and Brockmole and Le-Hoa Vo (2010), or by using artificial grids instead of natural scenes.

Conclusion

On a theoretical level, the results of our study indicate that the two distinct visual subprocesses, i.e., target detection and target verification, may play different roles in decision-making, notably contributing in the influence of noise and expected target location on the decision process. Noise degrading bottom-up visual information slowed down decision-making by increasing target verification, consequently moderating the commitment to the decision (as reflected in the speed of the mouse movement). Target location influenced decision-making by affecting eye movements relating to target detection, consequently moderating the bias toward the *absent* response. Additionally, we suggested that selecting the *absent* response builds up from the lack of evidence in favor of the *present* response as a result of the time spent unsuccessfully searching for the target. Importantly, our results describe a close interaction between bottom-up and top-down processes with which both visual uncertainty and expectations drive the accumulation of evidence during visual search. On a behavioral level, this is an extension of existing models of top-down recognition that suggest fast and automatic interactions between bottom-up and top-down visual streams (Bar, 2004; Boffa et al., 2015; Kauffmann, Chauvin, Guyader, & Peyrin, 2015). This raises important questions about bottom-up and top-down integration within and across fixations during scene exploration. For example, one might consider nuanced integrative schemes in which the relative contributions of top-down and bottom-up information are weighted by their relative precision (or inverse variance) and address the issue of how the time course of integration relates to the precision of information and confidence levels (Ambrosini, Pezzulo, & Con-

stantini, 2015; Friston et al., 2012; Pezzulo, Rigoli, & Friston, 2015). Further psychophysical, neuroimaging, and modeling studies will need to establish how the current behavioral data could be understood within the theoretical frameworks provided in cognitive neurosciences.

At the methodological level, we think it is particularly interesting to study eye guidance in real-world scenes by considering the decision-making process underlying the visual search task to which the eye movements contribute. In that respect, one effective solution is to combine mouse tracking with eye tracking. This study provides an early experimental account of the potential of such a method. It is a potentially fruitful perspective for goal-oriented vision in that it focuses on a process common to a broad range of perceptual functions, including target identification (Trapp & Bar, 2015), face perception (Freeman & Ambady, 2011; Freeman, Penner et al., 2011), word reading (Barca et al., 2016; Barca & Pezzulo, 2012, 2015), and choosing between available actions for humans (Lemonnier, Brémond, & Baccino, 2014). If both experimental approaches were taken into account by means, for instance, of mouse-tracking tasks and/or signal detection theory and computational approaches implementing decision-making with eye movements (e.g., Friston et al., 2012; Quinton et al., 2014) as well as fixation durations (e.g., Nuthmann et al., 2010), it would be particularly beneficial for further understanding of how information is continuously and actively collected in order to reduce the uncertainty surrounding competing perceptual decisions in visual search.

Box 1: Summary of the main results

Eye movement measures: Targets placed in unusual locations led to detection time and more fixations during the detection epoch compared to targets in usual locations. In contrast to these results, noise increased verification time and fixation duration during the detection epoch.

Response time: Both targets in an unusual location and degradation with noise increased the RT. The effect of target location was mediated by detection time whereas the effect of noise was mediated by verification time, thus clearly distinguishing the influence of both on the RT.

Mouse velocity on x-axis: The *x*-velocity began increasing before the moment of target detection and was higher for unusual target locations than for usual from 288 ms to 128 ms before target detection to 416 ms to 496 ms after target detection. The peak of the *x*-velocity profiles occurred 106 ms to 191 ms after the end of target verification, and their amplitude was decreased by noise. Additionally,

noise increased x -velocity in the latest parts of the profile. Hence, target location influenced the early part of the response movement whereas noise influenced the late part.

Area under the curve (global attraction by absent response): Although noise had no significant influence, the AUC was higher for targets in unusual locations than for targets in usual locations. This effect was modulated by detection times: Higher detection times further increased the AUC for targets in unusual locations compared to usual locations.

Deviation on x -axis: The deviation from the ideal trajectory (straight movement) began increasing before the moment of target detection. It increased earlier and overall was larger for targets in unusual locations, indicating a competition between responses early on. Furthermore, detection time increased x -deviation before the moment of target detection and through most of the mouse trajectory, and the effect was greater for targets in unusual location.

Keywords: visual search, perceptual decision-making, eye movements, mouse tracking, context information, scene perception

Acknowledgments

This study was approved by the French Committee for the Protection of Persons “South-East VI” (Comité de Protection des Personnes “Sud-Est VI”; IRB00008526, CE33, 2013). The authors would like to thank Mélanie Marcotte and Jessica Guilhe for their contribution to the visual scene selection and data collection. This work was funded by grants from the French program “investissement d’avenir” managed by the National Research Agency (ANR) from the European Union (Auvergne European Regional Development Funds—ERDF—of Auvergne region) and from the “Région Auvergne” in the framework of the IMobS3 LabEx (ANR-10-LABX-16-01).

Commercial relationships: none.

Corresponding author: Marie Izaute.

Email: marie.izaute@univ-bpclermont.fr.

Address: Clermont University, Blaise Pascal University, LAPSCO, Clermont-Ferrand, France.

References

- Ambrosini, E., Pezzulo, G., & Constantini, M. (2015). The eye in hand: predicting others’ behavior by integrating multiple sources of information. *Journal of Neurophysiology*, 113(7), 2271–2279, doi:10.1152/jn.00464.2014.
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390–412, doi:10.1016/j.jml.2007.12.005.
- Bajcsy, R., Aloimonos, Y., & Tsotsos, J. K. (2016). Revisiting active perception. *Perception*, 1–22.
- Ballard, D. H., & Hayhoe, M. M. (2009). Modelling the role of task in the control of gaze. *Visual Cognition*, 17(6–7), 1185–1204, doi:10.1080/13506280902978477.
- Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience*, 5(8), 617–629, doi:10.1038/nrn1476.
- Barca, L., Benedetti, F., & Pezzulo, G. (2016). The effects of phonological similarity on the semantic categorisation of pictorial and lexical stimuli: Evidence from continuous behavioural measures. *Journal of Cognitive Psychology*, 28(2), 159–170, doi:10.1080/20445911.2015.1101117.
- Barca, L., & Pezzulo, G. (2012). Unfolding visual lexical decision in time. *PLoS One*, 7(4), e35932, doi:10.1371/journal.pone.0035932. [Article]
- Barca, L., & Pezzulo, G. (2015). Tracking second thoughts: Continuous and discrete revision processes during visual lexical decision. *PLoS One*, 10(2), e0116193, doi:10.1371/journal.pone.0116193. [Article]
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278, doi:10.1016/j.jml.2012.11.001.
- Beffara, B., Wicker, B., Vermeulen, N., Ouellet, M., Bret, A., Funes Molina, M. J., & Mermillod, M. (2015). Reduction of interference effect by low spatial frequency information priming in an emotional Stroop task. *Journal of Vision*, 15(6):16, 1–9, doi:10.1167/15.6.16. [PubMed] [Article]
- Birem, M., Quinton, J.-C., Berry, F., & Mezouar, Y. (2014). SAIL-MAP: Loop-closure detection using saliency-based features. In *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems* (pp. 4543–4548). IEEE, doi:10.1109/IROS.2014.6943206.
- Blignaut, P. (2009). Fixation identification: The optimum threshold for a dispersion algorithm. *Attention, Perception, & Psychophysics*, 71(4), 881–895, doi:10.3758/APP.71.4.881.
- Borji, A., Tavakoli, H. R., Sihite, D. N., & Itti, L.

- (2013). Analysis of scores, datasets, and models in visual saliency prediction. In *2013 IEEE International Conference on Computer Vision* (pp. 921–928). IEEE, doi:10.1109/ICCV.2013.118.
- Brockmole, J. R., & Henderson, J. M. (2006). Recognition and attention guidance during contextual cueing in real-world scenes: Evidence from eye movements. *The Quarterly Journal of Experimental Psychology*, *59*(7), 1177–1187, doi:10.1080/17470210600665996.
- Brockmole, J. R., & Le-Hoa Vo, M. (2010). Semantic memory for contextual regularities within and across scene categories: Evidence from eye movements. *Attention, Perception, & Psychophysics*, *72*(7), 1803–1813, doi:10.3758/APP.72.7.1803.
- Castelhano, M. S., & Heaven, C. (2010). The relative contribution of scene context and target features to visual search in scenes. *Attention, Perception, & Psychophysics*, *72*(5), 1283–1297, doi:10.3758/APP.72.5.1283.
- Castelhano, M. S., & Heaven, C. (2011). Scene context influences without scene gist: Eye movements guided by spatial associations in visual search. *Psychonomic Bulletin & Review*, *18*(5), 890–896, doi:10.3758/s13423-011-0107-8.
- Castelhano, M. S., & Henderson, J. M. (2007). Initial scene representations facilitate eye movement guidance in visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(4), 753–763, doi:10.1037/0096-1523.33.4.753.
- Castelhano, M. S., Pollatsek, A., & Cave, K. R. (2008). Typicality aids search for an unspecified target, but only in identification and not in attentional guidance. *Psychonomic Bulletin & Review*, *15*(4), 795–801, doi:10.3758/PBR.15.4.795.
- Catenacci Volpi, N., Quinton, J. C., & Pezzulo, G. (2014). How active perception and attractor dynamics shape perceptual categorization: A computational model. *Neural Networks*, *60*, 1–16, doi:10.1016/j.neunet.2014.06.008.
- Dale, R., Kehoe, C., & Spivey, M. J. (2007). Graded motor responses in the time course of categorizing atypical exemplars. *Memory & Cognition*, *35*(1), 15–28, doi:10.3758/BF03195938.
- Eckstein, M. P., Drescher, B. A., & Shimozaki, S. S. (2006). Attentional cues in real scenes, saccadic targeting, and Bayesian priors. *Psychological Science*, *17*(11), 973–980, doi:10.1111/j.1467-9280.2006.01815.x.
- Ehinger, K. A., Hidalgo-Sotelo, B., Torralba, A., & Oliva, A. (2009). Modelling search for people in 900 scenes: A combined source model of eye guidance. *Visual Cognition*, *17*(6–7), 945–978, doi:10.1080/13506280902834720.
- Farmer, T., Cargill, S., Hindy, N., Dale, R., & Spivey, M. (2007). Tracking the continuity of language comprehension: Computer mouse trajectories suggest parallel syntactic processing. *Cognitive Science: A Multidisciplinary Journal*, *31*(5), 889–909, doi:10.1080/03640210701530797.
- Faulkenberry, T. J. (2014). Hand movements reflect competitive processing in numerical cognition. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *68*(3), 147–151, doi:10.1037/cep0000021.
- Foulsham, T., & Underwood, G. (2008). What can saliency models predict about eye movements? Spatial and sequential aspects of fixations during encoding and recognition. *Journal of Vision*, *8*(2):6, 1–17, doi:10.1167/8.2.6. [PubMed] [Article]
- Freeman, J. B., & Ambady, N. (2009). Motions of the hand expose the partial and parallel activation of stereotypes. *Psychological Science*, *20*(10), 1183–1188, doi:10.1111/j.1467-9280.2009.02422.x.
- Freeman, J. B., & Ambady, N. (2010). MouseTracker: Software for studying real-time mental processing using a computer mouse-tracking method. *Behavior Research Methods*, *42*(1), 226–241, doi:10.3758/BRM.42.1.226.
- Freeman, J. B., & Ambady, N. (2011). Hand movements reveal the time-course of shape and pigmentation processing in face categorization. *Psychonomic Bulletin & Review*, *18*(4), 705–712, doi:10.3758/s13423-011-0097-6.
- Freeman, J. B., Ambady, N., Rule, N. O., & Johnson, K. L. (2008). Will a category cue attract you? Motor output reveals dynamic competition across person construal. *Journal of Experimental Psychology: General*, *137*(4), 673–690, doi:10.1037/a0013875.
- Freeman, J. B., Dale, R., & Farmer, T. A. (2011). Hand in motion reveals mind in motion. *Frontiers in Psychology*, *2*, 1–6, doi:10.3389/fpsyg.2011.00059. [Article]
- Freeman, J. B., Ma, Y., Han, S., & Ambady, N. (2013). Influences of culture and visual context on real-time social categorization. *Journal of Experimental Social Psychology*, *49*(2), 206–210, doi:10.1016/j.jesp.2012.10.015.
- Freeman, J. B., Pauker, K., Apfelbaum, E. P., & Ambady, N. (2010). Continuous dynamics in the real-time perception of race. *Journal of Experimental Social Psychology*, *46*(1), 179–185, doi:10.1016/j.jesp.2009.10.002.
- Freeman, J. B., Penner, A. M., Saperstein, A., Scheutz,

- M., & Ambady, N. (2011). Looking the part: Social status cues shape race perception. *PLoS One*, 6(9), e25107, doi:10.1371/journal.pone.0025107. [Article]
- Frintrop, S. (2006). *VOCUS: A visual attention system for object detection and goal-directed search*. In J. G. Carbonell & J. Siekmann (Eds.). Berlin Heidelberg, New York: Springer.
- Friston, K. (2012). Embodied inference and spatial cognition. *Cognitive Processing*, 13(Suppl. 1), S171–S177, doi:10.1007/s10339-012-0519-z.
- Friston, K., Adams, R. A., Perrinet, L., & Breakspear, M. (2012). Perceptions as hypotheses: Saccades as experiments. *Frontiers in Psychology*, 3, 1–20, doi:10.3389/fpsyg.2012.00151. [Article]
- Harley, E. M., Dillon, A. M., & Loftus, G. R. (2004). Why is it difficult to see in the fog? How stimulus contrast affects visual perception and visual memory. *Psychonomic Bulletin & Review*, 11(2), 197–231, doi:10.3758/BF03196564.
- Hayes, A. F. (2009). Beyond Baron and Kenny: Statistical mediation analysis in the new millennium. *Communication Monographs*, 76(4), 408–420, doi:10.1080/03637750903310360.
- Hehman, E., Stolier, R., & Freeman, J. (2015). Advanced mouse-tracking analytic techniques for enhancing psychological science. *Group Process & Intergroup Relations*, 8(3), 384–401, doi:10.1177/1368430214538325.
- Henderson, J. M., Brockmole, J. R., Castelano, M. S., & Mack, M. (2007). Visual saliency does not account for eye movements during visual search in real-world scenes. In R. P. G. Van Gompel, M. H. Fischer, W. S. Murray, & R. L. Hill (Eds.), *Eye movements: A window on mind and brain* (pp. 537–562). Amsterdam: Elsevier.
- Henderson, J. M., & Smith, T. J. (2009). How are eye fixation durations controlled during scene viewing? Further evidence from a scene onset delay paradigm. *Visual Cognition*, 17(6–7), 1055–1082, doi:10.1080/13506280802685552.
- Itti, L., & Baldi, P. (2009). Bayesian surprise attracts human attention. *Visual Research*, 49(10), 1295–1306, doi:10.1016/j.visres.2008.09.007.
- Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Visual Research*, 40(10–12), 1489–1506.
- Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions of Pattern Analysis and Machine Intelligence*, 20(11), 1254–1259, doi:10.1109/34.730558.
- Johnson, K. L., Freeman, J. B., & Pauker, K. (2012). Race is gendered: How covarying phenotypes and stereotypes bias sex categorization. *Journal of Personality and Social Psychology*, 102(1), 116–131, doi:10.1037/a0025335.
- Kanan, C., Tong, M. H., Zhang, L., & Cottrell, G. W. (2009). SUN: Top-down saliency using natural statistics. *Visual Cognition*, 17(6–7), 979–1003, doi:10.1080/13506280902771138.
- Kauffmann, L., Chauvin, A., Guyader, N., & Peyrin, C. (2015). Rapid scene categorization: Role of spatial frequency order, accumulation mode and luminance contrast. *Vision Research*, 107, 49–57, doi:10.1016/j.visres.2014.11.013.
- Kietzmann, T., Geuter, S., & König, P. (2011). Overt visual attention as a causal factor of perceptual awareness. *PLoS One*, 6(7), e22614, doi:10.1371/journal.pone.0022614. [Article]
- Kietzmann, T., & König, P. (2015). Effects of contextual information and stimulus ambiguity on overt visual sampling behavior. *Vision Research*, 110, 76–86, doi:10.1016/j.visres.2015.02.023.
- Kirsh, D., & Maglio, P. (1994). On distinguishing epistemic from pragmatic action. *Cognitive Science*, 18(4), 513–549, doi:10.1207/s15516709cog1804_1.
- Kovalenko, L. Y., Chaumon, M., & Busch, N. A. (2012). A pool of pairs of related objects (PO-PORO) for investigating visual semantic integration: Behavioral and electrophysiological validation. *Brain Topography*, 25(3), 272–284, doi:10.1007/s10548-011-0216-8.
- Land, M. F., & Hayhoe, M. (2001). In what ways do eye movements contribute to everyday activities? *Vision Research*, 41(25–26), 3559–3565, doi:10.1016/S0042-6989(01)00102-X.
- Lemonnier, S., Brémond, R., & Baccino, T. (2014). Discriminating cognitive processes with eye movements in a decision-making driving task. *Journal of Eye Movement Research*, 7(4), 1–14.
- Lepora, N. F., & Pezzulo, G. (2015). Embodied choice: How action influences perceptual decision making. *PLoS Computational Biology*, 11(4), e1004110, doi:10.1371/journal.pcbi.1004110.
- Malcolm, G. L., & Henderson, J. M. (2009). The effects of target template specificity on visual search in real-world scenes: Evidence from eye movements. *Journal of Vision*, 9(11):8, 1–13, doi:10.1167/9.11.8. [PubMed] [Article]
- Malcolm, G. L., & Henderson, J. M. (2010). Combining top-down processes to guide eye movements during real-world scene search. *Journal of Vision*, 10(2):4, 1–11, doi:10.1167/10.2.4. [PubMed] [Article]

- Marghetis, T., Núñez, R., & Bergen, B. K. (2014). Doing arithmetic by hand: Hand movements during exact arithmetic reveal systematic, dynamic spatial processing. *The Quarterly Journal of Experimental Psychology*, 67(8), 1579–1596, doi:10.1080/17470218.2014.897359.
- Morett, L. M., & MacWhinney, B. (2013). Syntactic transfer in English-speaking Spanish learners. *Bilingualism: Language and Cognition*, 16(1), 132–151, doi:10.1017/S1366728912000107.
- Nuthmann, A., Smith, T. J., Engbert, R., & Henderson, J. M. (2010). CRISP: A computational model of fixation durations in scene viewing. *Psychological Review*, 117(2), 382–405, doi:10.1037/a0018924.
- Papesh, M. H., & Goldinger, S. D. (2012). Memory in motion: Movement dynamics reveal memory strength. *Psychonomic Bulletin & Review*, 19(5), 906–913, doi:10.3758/s13423-012-0281-3.
- Pereira, E. J., & Castelano, M. S. (2014). Peripheral guidance in scenes: The interaction of scene context and object content. *Journal of Experimental Psychology: Human Perception and Performance*, 40(5), 2056–2072, doi:10.1037/a0037524.
- Pezzulo, G., Rigoli, F., & Friston, K. (2015). Active inference, homeostatic regulation and adaptive behavioural control. *Progress in Neurobiology*, 134, 17–35, doi:10.1016/j.pneurobio.2015.09.001.
- Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, 40(3), 879–891, doi:10.3758/BRM.40.3.879.
- Quétyard, B., Quinton, J.-C., Colomb, M., Pezzulo, G., Barca, L., Izaute, M., ... Mermillod, M. (2015). Combined effects of expectations and visual uncertainty upon detection and identification of a target in the fog. *Cognitive Processing*, 16(Suppl. 1), S343–S348, doi:10.1007/s10339-015-0673-1.
- Quinton, J.-C., Volpi, N. C., Barca, L., & Pezzulo, G. (2014). The cat is on the mat. Or is it a dog? Dynamic competition in perceptual decision making. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 44(5), 539–551, doi:10.1109/TSMC.2013.2279664.
- Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2(1), 79–87, doi:10.1038/4580.
- Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation*, 20(4), 873–922, doi:10.1162/neco.2008.12-06-420.
- Renninger, L. W., Verghese, P., & Coughlan, J. (2007). Where to look next? Eye movements reduce local uncertainty. *Journal of Vision*, 7(3):6, 1–17, doi:10.1167/7.3.6. [PubMed] [Article]
- Röhrbein, F., Goddard, P., Schneider, M., James, G., & Guo, K. (2015). How does image noise affect actual and predicted human gaze allocation in assessing image quality? *Vision Research*, 112, 11–25, doi:10.1016/j.visres.2015.03.029.
- Rothkopf, C. A., Ballard, D. H., & Hayhoe, M. M. (2007). Task and context determine where you look. *Journal of Vision*, 7(14):16, 1–20, doi:10.1167/7.14.16. [PubMed] [Article]
- Rucker, D. D., Preacher, K. J., Tormala, Z. L., & Petty, R. E. (2011). Mediation analysis in social psychology: Current practices and new recommendations. *Social and Personality Psychology Compass*, 6, 359–371.
- Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the symposium on eye tracking research & applications - ETRA '00* (pp. 71–78). New York: ACM Press, doi:10.1145/355017.355028.
- Song, J.-H., & Nakayama, K. (2009). Hidden cognitive states revealed in choice reaching tasks. *Trends in Cognitive Sciences*, 13(8), 360–366, doi:10.1016/j.tics.2009.04.009.
- Spivey, M. J., Grosjean, M., & Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences, USA*, 102(29), 10393–10398, doi:www.pnas.org/cgi/doi/10.1073/pnas.0503903102.
- Spotorno, S., Malcolm, G. L., & Tatler, B. W. (2014). How context information and target information guide the eyes from the first epoch of search in real-world scenes. *Journal of Vision*, 14(2):7, 1–21, doi:10.1167/14.2.7. [PubMed] [Article]
- Tatler, B. (2007). The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. *Journal of Vision*, 7(14):4, 1–17, doi:10.1167/7.14.4. [PubMed] [Article]
- Tatler, B., Hayhoe, M., Land, M., & Ballard, D. (2011). Eye guidance in natural vision: Reinterpreting saliency. *Journal of Vision*, 11(5):5, 1–23, doi:10.1167/11.5.5. [PubMed] [Article]
- 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(4), 766–786, doi:10.1037/0033-295X.113.4.766.
- Trapp, S., & Bar, M. (2015). Prediction, context, and

- competition in visual recognition. *Annals of the New York Academy of Sciences*, 1339, 190–198, doi:10.1111/nyas.12680.
- Underwood, G., & Foulsham, T. (2006). Visual saliency and semantic incongruency influence eye movements when inspecting pictures. *Quarterly Journal of Experimental Psychology*, 59(11), 1931–1949, doi:10.1080/17470210500416342.
- Underwood, G., Foulsham, T., & Humphrey, K. (2009). Saliency and scan patterns in the inspection of real-world scenes: Eye movements during encoding and recognition. *Visual Cognition*, 17(6–7), 812–834, doi:10.1080/13506280902771278.
- Underwood, G., Foulsham, T., van Loon, E., Humphreys, L., & Bloyce, J. (2006). Eye movements during scene inspection: A test of the saliency map hypothesis. *European Journal of Cognitive Psychology*, 18(3), 321–342, doi:10.1080/09541440500236661.
- Võ, M. L.-H., & Wolfe, J. M. (2015). The role of memory for visual search in scenes. *Annals of the New York Academy of Sciences*, 1339, 72–81, doi:10.1111/nyas.12667.
- Wolfe, J. M. (2012). When do I quit? The search termination problem in visual search. In *Nebraska Symposium on Motivation* (Vol. 59) (pp. 183–208). New York: Springer, doi:10.1007/978-1-4614-4794-8_8.
- Wu, C. C., Wick, F. A., & Pomplun, M. (2014). Guidance of visual attention by semantic information in real-world scenes. *Frontiers in Psychology*, 5, 1–13, doi:10.3389/fpsyg.2014.00054. [Article]
- Xiao, J., Hays, J., Ehinger, K. A., Oliva, A., & Torralba, A. (2010). SUN database: Large-scale scene recognition from abbey to zoo. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 3485–3492). IEEE, doi:10.1109/CVPR.2010.5539970.

2.1. Résumé des résultats de l'Étude 2

L'Étude 2 a permis de mettre en évidence un rôle différencié des processus de détection et de vérification dans le processus de prise de décision présidant à déterminer la présence ou l'absence d'une cible dans une scène visuelle. Ainsi, l'influence contextuelle affectait la prise de décision via le processus de détection de la cible, alors que la force de l'évidence sensorielle l'affectait via le processus de vérification de la cible. Par ailleurs, les résultats indiquaient une influence du temps de détection sur la sélection de la réponse (évaluée via la trajectoire de la souris) avant même la première fixation sur la cible. Cela suggère que rechercher du regard la cible sans la trouver dévie la trajectoire de la souris vers la réponse *absent*, semblant indiquer que de l'évidence en faveur de la réponse *absent* est accumulée à mesure que le temps passe sans que la cible ne soit détectée.

En plus de proposer un mécanisme permettant de terminer une recherche visuelle infructueuse, ce résultat met en avant une différence fondamentale entre tâches de catégorisation où l'évidence est accumulée en faveur d'une réponse ou de l'autre, et tâches de recherche visuelle. Dans ces dernières, l'évidence en faveur de la réponse *présent* est accumulée par la détection et vérification de la cible, alors que l'évidence en faveur de la réponse *absent* est accumulée lorsque l'exploration visuelle échoue à collecter de l'évidence en faveur de la présence de la cible durant la période de détection (*i.e.*, sa détection). L'Étude 3 vise à répliquer ces résultats, en employant un design portant des thématiques applicatives autour des thématiques centrales du LabEx IMobS³ et du laboratoire de Clermont-Ferrand du CEREMA, à savoir la détection d'obstacles dans un environnement routier en situation de brouillard. De part cette thématique, nous avons dû adapter le type d'attentes manipulées et avons donc préféré des attentes sur l'identité de la cible, plus facilement applicables au domaine routier.

3. Étude 3 : Influence du brouillard et des attentes sur l'identité de la cible sur la prise de décision et l'accumulation d'information via les mouvements des yeux

L'Étude 3 emploie une démarche similaire à l'Étude 2 tout en cherchant à explorer des thématiques applicatives autour de la perception visuelle en situation de conduite, en particulier en présence de brouillard. Elle comprend deux articles présentant chacun une expérience. L'Article 3 présente une tâche d'identification de cible (Expérience 1) et l'Article 4 (en version de travail) une tâche de recherche visuelle (Expérience 2). En collaboration avec le laboratoire de Clermont-Ferrand du CEREMA, nous avons eu accès à une salle équipée, permettant de produire du brouillard artificiel reproduisant les caractéristiques physiques du brouillard naturel. Nous avons donc construit un protocole permettant d'exploiter cet équipement mis à disposition, en particulier via la construction de scènes visuelles (statiques) où l'incertitude visuelle était manipulée de manière plus écologique que dans les études précédentes. Étant donné le matériel à disposition, nous avons été contraints de construire des scènes visuellement très simples, et manipuler un autre type d'attentes que dans les Études 1 et 2. L'Annexe 1 inclus le poster présenté à la conférence Transport Research Arena, présentant l'outil que nous avons employé pour construire des scènes standardisées à partir des photographies prises au CEREMA et l'Annexe 2 présente des exemples de scènes visuelles employées dans l'Expérience 2 (prises de vue réelles).

Les scènes visuelles représentaient une scène routière nocturne vue du cockpit d'une voiture et comprenaient une route, une voiture (servant à ajouter un élément contextuel) et un panneau signalant un danger (endroit fréquenté par des enfants, animaux sauvages, point d'exclamation). De plus, la scène pouvait comprendre un chevreuil/cerf (selon l'expérience) ou un piéton traversant la route (quatre positions possibles), ou aucun des deux. La présence éventuelle de ces stimuli et leur position, l'identité des panneaux de signalisation et la densité du brouillard variaient d'une scène à l'autre. La tâche des participants était soit d'indiquer l'identité du stimulus présent dans la scène pour l'Expérience 1 (piéton ou cerf, avec aucun essai où aucun des deux n'était présent), soit d'indiquer si un piéton était présent dans la scène pour l'Expérience 2. Pour la seconde tâche, les participants devaient donc répondre *absent* si un chevreuil était présent dans la scène, ou si ni piéton ni chevreuil n'étaient présents dans la scène. Le panneau de signalisation indiquait lequel de ces stimuli avait le plus de chances d'être dans la scène visuelle (panneau endroit fréquenté par des enfants : piéton ; panneau animaux sauvages : chevreuil/cerf ; danger générique : probabilité équivalente pour les deux stimuli). Comme pour l'Étude 2, nous avons mesuré à la fois les trajectoires de la souris et les mouvements des yeux (temps de détection et temps de vérification).

Au total, nous avons effectué cinq expériences dans le cadre de cette étude. Les trois premières

Chapitre 3 – Études expérimentales

expériences ont servi d'expériences pilotes pour l'Expérience 2. Elles ont été construites selon le schéma présenté ci-dessus, mais pour construire les scènes visuelles, nous avons utilisé un environnement 3D reproduisant les dimensions de la salle produisant du brouillard et le placement des différents éléments. Nous avons utilisé de la modélisation 3D pour ces expériences, d'une part pour des raisons de disponibilité de la salle produisant du brouillard et d'autre part, car elles visaient à pré-tester le protocole et déterminer définitivement la composition des scènes visuelles avant d'effectuer les prises de vue réelles. De plus, ces expériences nous ont permis de choisir certains paramètres tels que la durée et les modalités de présentation du panneau de signalisation avant apparition de la scène visuelle (le panneau était aussi présent dans la scène elle-même). Ces expériences ne seront pas présentées dans le présent document.

L'Article 3, publié dans la revue *Cognitive Processing*, présente l'Expérience 1 construite avec les scènes visuelles en 3D, où la tâche employée vise l'identification de la cible (piéton ou cerf). Cet article ayant un format très court, nous nous sommes limités à analyser, via des médiations, comment et dans quelle mesure les effets du brouillard et des attentes sur les mesures globales de prise de décision (aires-sous-la-courbe et temps de réponse) sont expliqués par les processus de détection et de vérification de la cible.

L'Article 4 est une version de travail présentant une tâche de recherche visuelle utilisant les scènes construites à partir des prises de vue réelles effectuées au CEREMA. Bien qu'en version de travail, l'article se suffit à lui-même. Les résultats présentés pour cette expérience sont des résultats finaux, se concentrant sur les analyses de médiation et modération des effets des attentes et du brouillard sur la déviation globale de la souris et les temps de réponse par les temps de détection et vérification. Les analyses du déroulement temporel des réponses n'étant pas terminées, elles ne font pas part de ce manuscrit (mais seront ajoutées à l'article avant sa soumission à une revue). Les analyses de médiation et modération seules permettent cependant de confirmer certains résultats obtenus dans l'Étude 2.

Similairement à l'Étude 2 nous attendions une dégradation de la sélection de réponse (mesurée via la déviation de la souris et les temps de réponse) par le brouillard, médiatisée par un effet du brouillard sur les temps de vérification et éventuellement sur les temps de détection (dans le cas où le brouillard gêne la vision périphérique). Par ailleurs, nous manipulons une forme d'attentes portant sur l'identité de la cible plutôt que sur sa localisation. Ces attentes peuvent être rapprochées des attentes sur l'apparence de la cible (ses caractéristiques visuelles) dans le cas des expériences de l'Étude 3 puisque la cible (piéton) et le distracteur (chevreuil) ont toujours la même apparence tout au long des expériences et cette apparence est donc très rapidement apprise par les participants (dès la phase d'apprentissage). Ainsi, dans une certaine mesure, s'attendre à un piéton revient à anticiper les caractéristiques du piéton dans la scène pour déterminer s'il est bien présent, et répondre en conséquence. Suivant cette logique, nous attendions que

Chapitre 3 – Études expérimentales

les attentes sur l'identité de la cible affectent le temps de vérification, et éventuellement le temps de détection dans le cas où les caractéristiques visuelles du piéton et du chevreuil/cerf sont suffisamment discriminables en vision périphérique ; des attentes correctes (*e.g.*, attentes d'un piéton lorsqu'un piéton fait partie de la scène) devraient réduire ces temps. Cette facilitation de la collecte d'évidence devrait entraîner une facilitation de la sélection de réponse (médiation ou modulation de l'effet des attentes sur la les aires-sous-la-courbe). Cependant, les attentes sur l'identité de la cible n'affectent pas la sélection de réponse uniquement via la facilitation de la vérification et/ou détection de la cible ; elles peuvent aussi provoquer un biais de réponse direct en faveur de la réponse correspondante aux attentes (indépendamment de l'information visuelle *bottom-up*). Les analyses de médiation et modulation permettent de distinguer ces deux effets.

Enfin, dans l'Étude 2, nous avons discuté du fait que dans une tâche de recherche visuelle, le temps passé à rechercher visuellement la cible sans la trouver pouvait permettre l'accumulation d'évidence en faveur de la réponse *absent*. L'Expérience 2 de l'Étude 3 nous a permis de formuler des hypothèses plus précises vérifiant cette idée. Cette expérience présente une tâche de recherche visuelle où le piéton est considéré comme une cible (réponse *présent*), le chevreuil comme un distracteur (réponse *absent*), et comprenant des essais où ni l'un, ni l'autre ne font partie de la scène. Si passer plus de temps à détecter l'objet accumule de l'évidence en faveur de la réponse *absent*, des temps de détection élevés devraient augmenter la déviation vers la réponse incorrecte (*absent*) lorsqu'un piéton est dans la scène et diminuer la déviation vers la réponse incorrecte (*présent*) lorsqu'un chevreuil est présent dans la scène. Dans ce dernier cas, la difficulté à détecter le chevreuil devrait biaiser la trajectoire de la réponse en direction de la réponse *absent*, qui s'avère être la réponse correcte. Si nous pouvions formuler cette hypothèse simple, il faut noter que les attentes sur l'identité de la cible, dont nous faisons l'hypothèse qu'elles affectent à la fois la détection et la sélection de réponse, risquaient de rendre les résultats plus complexes à interpréter.

Article 3

Étude 3, expérience 1

Combined effects of expectations and visual uncertainty upon detection and identification of a target in the fog

Boris Quétard,

Jean-Charles Quinton,

Michèle Colomb,

Giovanni Pezzulo,

Laura Barca,

Marie Izaute,

Owen Kevin Appadoo,

Martial Mermillod

Journal: Cognitive Processing

Statut: Publié

Détails de publication:

Quétard et al. (2015). *Cogn Process*, 16(1), 343-348. doi: 10.1007/s10339-0.150673-1



Combined effects of expectations and visual uncertainty upon detection and identification of a target in the fog

Boris Quétard^{1,2} · Jean-Charles Quinton^{3,4} · Michèle Colomb⁵ · Giovanni Pezzulo⁶ · Laura Barca⁶ · Marie Izaute^{1,2} · Owen Kevin Appadoo⁷ · Martial Mermillod^{8,9,10}

© Marta Olivetti Belardinelli and Springer-Verlag Berlin Heidelberg 2015

Abstract Detecting a pedestrian while driving in the fog is one situation where the prior expectation about the target presence is integrated with the noisy visual input. We focus on how these sources of information influence the oculomotor behavior and are integrated within an underlying decision-making process. The participants had to judge whether high-/low-density fog scenes displayed on a computer screen contained a pedestrian or a deer by executing a mouse movement toward the response button (mouse-tracking). A variable road sign was added on the scene to manipulate expectations about target identity. We then analyzed the timing and amplitude of the deviation of mouse trajectories toward the incorrect response and, using an eye tracker, the detection time (before fixating the

target) and the identification time (fixations on the target). Results revealed that expectation of the correct target results in earlier decisions with less deviation toward the alternative response, this effect being partially explained by the facilitation of target identification.

Keywords Visual search · Perceptual decision-making · Mouse-tracking · Perceptual uncertainty · Expectations

Introduction

Detecting and identifying objects, like pedestrians while driving, is the result of a continuous decision-making process, including the progressive collection of visual information through oculomotor behavior in order to reduce the uncertainty of the visual input (Drugowitsch et al. 2012; Friston et al. 2012; Kietzmann et al. 2011). Prior knowledge, such as various expectations about the identity and features of the target, is also useful to constrain the recognition process. The effects of these information sources on the oculomotor behavior when searching for a target in a scene have been extensively studied (Castelhano and Heaven 2010; Castelano et al. 2008; Malcolm and Henderson 2009), leading to a distinction between a target detection epoch and a target verification epoch. The first corresponds to the time elapsed from the presentation of the scene to the first saccade on the target and the latter to the duration of the first sequence of fixations directed on the target's area. However, the relationships between these epochs and the encompassing decision-making process leading to the categorization of the target have not received sufficient attention in the field. A recent methodology proposes to measure the continuous aspects of the decision-making process by tracking the computer mouse's response

✉ Boris Quétard
boris.quetard@univ-bpclermont.fr

¹ LAPSCO, Clermont University, Blaise Pascal University, 63037 Clermont-Ferrand, France

² LAPSCO, CNRS, UMR 6024, 63037 Clermont-Ferrand, France

³ Pascal Institute, Clermont University, Blaise Pascal University, 63000 Clermont-Ferrand, France

⁴ Pascal Institute, CNRS, UMR 6602, 63177 Aubière, France

⁵ Département Laboratoire de Clermont-Ferrand, CEREMA, 63017 Clermont-Ferrand, France

⁶ Institute of Cognitive Sciences and Technologies, CNR, Rome, Italy

⁷ Pascal Institute, Clermont University, Auvergne University, 63000 Clermont-Ferrand, France

⁸ LPNC, Université Grenoble Alpes, 38040 Grenoble, France

⁹ LPNC, CNRS, UMR 5105, 38040 Grenoble, France

¹⁰ Institut Universitaire de France, Paris, France

Published online: 25 July 2015

Springer

movements in a binary forced-choice categorization task (Dale et al. 2007). Participants have to categorize a visual stimulus by clicking on the response buttons located on the top corners of the screen. The degree of deviation of each trajectory from an ideal response (line from the start point to the accurate response), computed as an area under the curve (AUC), can be used to evaluate the bias toward the alternative response in the decision-making process. The time at which the maximal deviation (MD-time) from the ideal trajectory is reached can be interpreted as the moment when the decision is made (Freeman and Ambady 2010).

Here we studied how expectations about target identity are integrated using a target detection task in which a road sign predicted the presence of a specific target (pedestrian or deer) within a foggy driving scene. We used the mouse-tracking methodology to evaluate the global decision-making components of the responses, whereas eye movements were measured to evaluate processes related to the collection of evidence, i.e., target detection and verification. We hypothesized that both a high level of fog and no expectations or expectations for the incorrect target (sign incongruent with target) would (1) increase the need for collecting additional information through eye movements, thus increasing detection and verification times, and (2) make decision-making harder, increasing AUC and MD-times. However, our design implies that these latter measures are not only determined by perceptual processes, but also by strategies such as a motor bias toward the expected response buttons. Thus, we used mediation analysis (Hayes 2009) to evaluate whether the effect of the fog and target-sign congruency on the mouse-tracking measures is due to perceptual processes (i.e., detection and verification). We consequently investigated whether the combined effect of the perceptual uncertainty (fog) and target's expectations interacted.

Method

Participants

Twenty psychology students (3 males, mean age = 19.3) took part in the experiment. All of them had a normal or corrected-to-normal vision (glasses or contact lenses).

Stimuli

Figure 1 shows instances of the visual stimuli used in the experiment. They depicted a driver's point of view, including a straight one-way road and a car cockpit, generated with a 3D modeler. Three objects were placed in the scene: a car on the road, a road sign and either a pedestrian or a deer on one of the four positions (Fig. 1 top-left). The

road signs were an "undefined danger sign," a "wild animal warning sign" or a "children warning sign" (Fig. 1 bottom-left). Additionally, half of the pictures contained a low fog level and the other half contained a high fog level, integrated using the blender tool. A total of 48 pictures were built with a resolution of 1200 × 900 pixels.

Apparatus

Eye movements were recorded using an SMI RED-m eye tracker (sampling rate of 120 Hz). For fixation identification, the SMI's dispersion-based algorithm was set with a minimum duration threshold at 100 ms and a minimum dispersion threshold at 1°. We used the MouseTracker software (Freeman and Ambady 2010) to program and run the experiment. The stimuli were presented using a laptop with a 15.6-inch display set at a resolution of 1350 × 900 pixels and a refresh rate of 60 Hz. A chin and forehead rest stabilized the eyes approximately 62 cm away from the display. A regular mouse was used, and the movements were recorded at a sampling rate of 70 Hz.

Procedure

After clicking on the start button (bottom-center of the screen), a fixation cross (300 ms) followed by a road sign prime (400 ms) and a second fixation cross (500 ms) were presented. The mouse cursor was locked on the start button for this total duration. Then, the visual scene with the same road sign and a low or high fog level was presented for 2000 ms. The participants had to judge whether the target was a deer or a pedestrian by clicking on the corresponding response button before 2000 ms elapsed (Fig. 1, top-right shows an instance of the display). The position of the response buttons was counterbalanced across subjects.

The animal sign predicted a deer target and the children sign a pedestrian target: 80 % of the trials with an animal or a children sign included the congruent target. Among the trials with a danger sign, 50 % included a deer and 50 % a pedestrian (neutral condition). See Fig. 1 bottom-left.

Results

Data processing

We computed AUC and MD-times (mouse-tracker measures) and the detection and verification times (eye-tracking measures). We excluded invalid trials based on responses (exceed time limit and errors 2.48 %) and on eye-tracking measures (invalid detection or verification times 6.86 %; no target fixation 7.48 %).

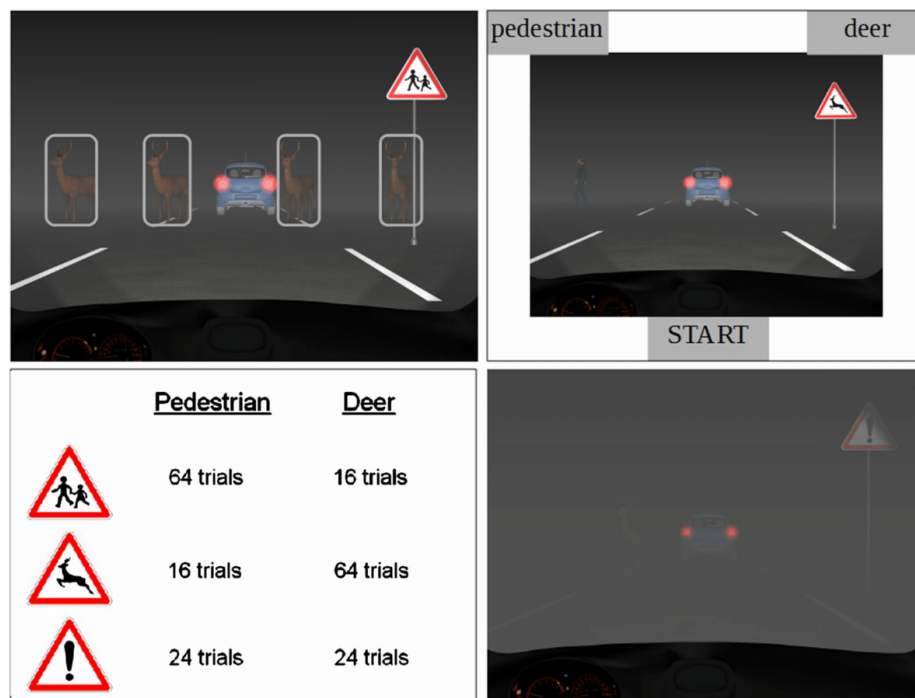


Fig. 1 *Top-left* Visual scene displaying the four alternative positions where the target (here a deer) can be found in the experiment. *Top-right* Example of a scene presented during the experiment including an incongruent visual scene and displaying the start and response buttons. *Bottom-left* Repartition of trials across targets (*columns*) and

road signs (*rows*) (total 208 trials); the road signs were presented before the scene presentation and were also included in the scene. *Bottom-right* Visual scene with the neutral road sign and a high fog level

We tested the congruency and fog effects on these measures using linear mixed models (LMM) with the participants as a random factor (Baayen et al. 2008). We recoded the congruency variable as two contrasts opposing, respectively, the neutral (−0.5) versus congruent (+0.5) modalities (C1) and neutral (−0.5) versus incongruent (+0.5) modalities (C2) with the nontested modality recoded 0. Finally, we evaluated the contribution of the detection and verification times as mediators of the independent variables' effects on the mouse-tracking measures. We estimated the reliability of the indirect effects by computing bootstrap confidence intervals (Hayes 2009). See Table 2 and Fig. 2c for additional details.

Fog effects on mouse movements and their mediation by eye movements

Table 1 displays the estimated parameters of the fog and congruency effects and Table 2 the results of the mediation analyses. Fog level induced both longer detection times

(difference = 56.6 ms) and verification times (difference = 58.9 ms). The effect of the fog level on all the mouse-tracking measures was significant, a high level inducing higher AUC and MD-time. Mediation analysis revealed that the effect of fog level on verification time explained part of these effects with both indirect effects significantly different from 0. When adding the verification time as a covariant, the total effect *c* was reduced by 46.96 % for AUC (*c* = 0.148, *c'* = 0.079) and 38.92 % for MD-time (*c* = 76.19, *c'* = 46.54). Detection time did not mediate the fog effect on AUC, but it decreased significantly its effect on MD-time by 24.7 % (*c* = 74.95, *c'* = 59.68) when included as covariant.

Congruency effects on mouse movements and their mediation by eye movements

As can be seen in Table 1, verification times were shorter for a congruent sign than for a neutral sign (difference = 11.2 ms). Incongruent signs (*m* = 0.80) elicited

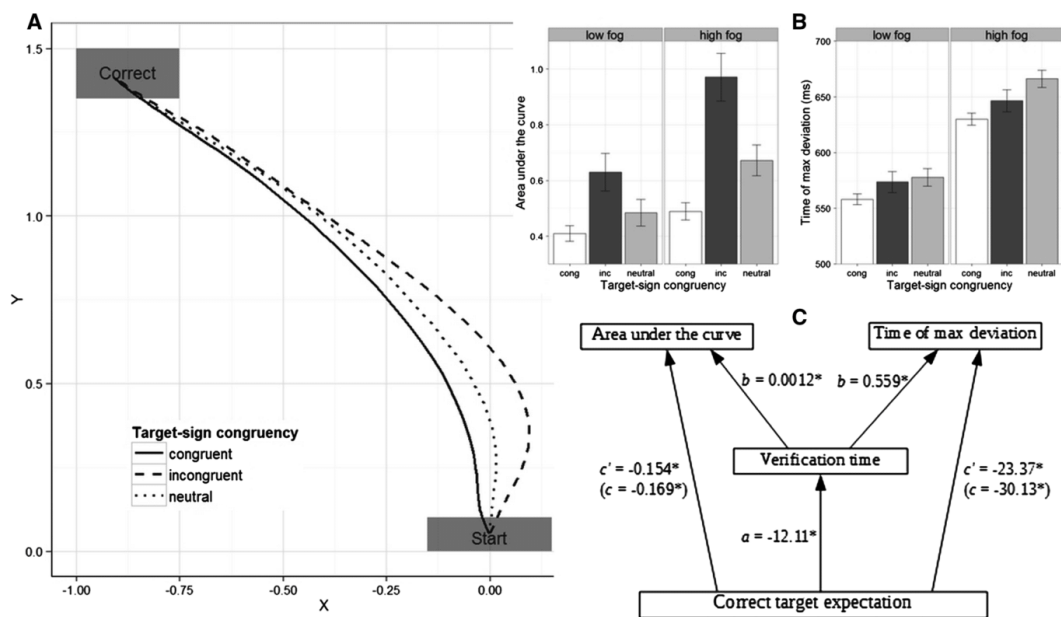


Fig. 2 Overview of the main results. **a** Mean mouse trajectories toward the correct response when the road sign is neutral, congruent or incongruent with the target. **b** Areas under the curve and times of maximal deviation as a function of fog level and sign-target congruency. Error bars represent the standard errors of the mean. **c** Mediation of the effect of the target expectations on the AUC and MD-times by the verification times. Correct target expectation

corresponds to the comparison between neutral and congruent road sign (contrast 1). The numbers indicate the parameters of the tested models; parameters b and c' are estimated in a model including both parameters, whereas the total effect c is estimated without the mediator as covariate; *asterisk* significant parameter (95 % bootstrap CI)

Table 1 Parameters (b) and t values for the effects of the causal variables, as a function of the different mouse movements and eye movement measures

Parameters	AUC		MD-time		Detection time		Verification time	
	b	t	b	t	b	t	b	t
Fog	0.200	4.294***	8.688	13.236***	3.597	19.807***	2.787	11.602***
Congruency contrast 1	-0.320	-3.395**	-32.343	-8.882***	-0.055	-0.420	-0.531	-3.359***
Congruency contrast 2	0.390	3.852***	5.147	0.503	0.212	1.192	0.120	0.558
Fog*Congruency C1	-0.231	-2.530*	-12.982	-0.941	-0.289	1.102	-0.460	-1.454
Fog*Congruency C2	0.237	1.902†	-4.143	-0.221	-0.496	-1.392	0.456	1.061

Congruency contrast 1 opposes congruent and neutral signs; congruency contrast 2 opposes incongruent and neutral signs. The parameters were estimated using linear mixed models with the subject factor as random factor, including the maximal random structure allowing for the model to converge (if no convergence, nonsignificant random slopes or correlations were removed one by one until convergence). For reasons relative to the assumptions of the LME, analysis of the detection and verification times were done on log-transformed measures. The p values were computed using the Kenward-Rogers approximation

*** $p < .001$; ** $p < .01$; * $p < .05$; † $p < .1$

trajectories which were more attracted to the inaccurate response button than neutral signs ($m = 0.58$), whereas the inverse relation was found for congruent signs ($m = 0.45$) (see Fig. 2a and b). Table 2 and Fig. 2c show that 8.9 % of

the difference between congruent and neutral signs on the AUC was mediated by verification time ($c = -0.169$, $c' = -0.154$). The difference between incongruent and neutral signs on AUC was not mediated by verification

Cogn Process

Table 2 Indirect effects estimated by the mediation analyses of the areas under the curve and times of maximum deviation for each mediator

Measures	Causal variables	Mediator			
		Detection time		Verification time	
		Indirect effect ab	CI 95 %	Indirect effect ab	CI 95 %
AUC	Fog	0.021	[-0.003, 0.041]	0.07	[0.042, 0.096] ^a
	Congruency C1	0.001	[-0.004, 0.006]	-0.015	[-0.026, -0.002] ^a
	Congruency C2	0.002	[-0.006, 0.007]	-0.009	[-0.023, 0.005]
MD-time	Fog	15.25	[8.35, 20.59] ^a	29.57	[23.64, 34.89] ^a
	Congruency C1	0.66	[-2.18, 3.31]	-6.77	[-11.23, -2.09] ^a
	Congruency C2	1.01	[-2.96, 4.67]	-4.06	[-10.30, 2.01]

Congruency C1 and C2 oppose, respectively, the neutral sign to the congruent and incongruent signs. The indirect effect ab is computed by multiplying the effect of the causal variable on the mediator (*a*) and the effect of the mediator on the dependent measure when controlling for the effect of the causal variable (*b*). These effects were estimated using linear mixed models with the causal variable and the mediator (when needed) as fixed factors and the subjects as random factor, including random intercept and slope of the causal variable. The 95 % confidence intervals are estimated through 10,000 bootstrap samples

^a Indicates that the indirect effect is significant

time. We also observed increased MD-times for congruent signs ($m = 595.21$ ms) compared to neutral signs ($m = 623.09$ ms), with verification time significantly mediating 22.4 % of this effect ($c = -30.13$, $c' = -23.37$). No difference was found between incongruent signs and neutral signs. None of these effects were mediated by detection time.

Finally, the fog level interacted with the expectations. The fog level significantly decreased AUC to a larger extent for congruent than neutral signs and marginally increased AUC to a larger extent for incongruent than neutral signs. These interactions were not mediated by the perceptual measure.

Discussion

We found a consistent effect of fog on all eye and mouse movement measures, with mouse movement measures mediated by target detection (only for MD-time) and target verification; these results show that a dense fog induces enough uncertainty to influence the decision-making process. The road signs predicting a congruent target decreased AUC and MD-time, indicating more direct trajectories and earlier decisions toward the correct response, whereas incongruent road signs generated a stronger bias toward the incorrect response. We obtained interactions between expectations and fog on the AUC indicating that congruent signs generated a stronger bias toward the correct response for a high fog level compared to neutral signs, and conversely, incongruent signs generated a stronger bias toward the incorrect response (although this interaction is marginally significant, which could be due to

the reduced number of trials). These findings are in accordance with a supra-additive integration of both sources of information within the decision-making process, where top-down expectations could be used to compensate for the bottom-up uncertainty.

Moreover, a road sign predicting a congruent target decreased the verification times, indicating that identifying an expected target was easier, in accordance with other studies (Castelhano and Heaven 2010; Malcolm and Henderson 2009). However, there was no effect of the congruent or the incongruent road sign on detection time. This may be explained by the use of simple visual scenes, where target and road sign were the only changing objects, with no distractor, whereas the previous studies used complex scenes.

Finally, our most important result was that a significant part of the expectancy effects was mediated by verification times. Even though the deviation of the mouse trajectory should be massively influenced by motor anticipations toward the target indicated by the road sign, we found a small part of its effect (around 8.9 %) influenced by the verification of the target. The influence of the target verification on the congruency effect was somewhat larger when considering the timing of the decision (around 22.4 % of the effect). These results suggest that the facilitation of the target's identification by a congruent road sign induced a (small) bias toward the congruent response within the decision-making process and allowed faster decisions to be made. Although we did not focus on determinants involved in the decision-making process other than the evidence collection, they still explain the vast majority of the influence of the expectancies on the response selection process. For instance, the expectancies could induce a motor bias toward the congruent response (Bruhn 2013; Hollingworth and Henderson 1998)

and a late inhibition if the expectation is proven wrong (Flowers 1990). Overall, our results suggest that expectations about target identity exert an influence at the perceptual level over the decision-making process. This top-down influence from conceptual to perceptual processes could be mediated through a situated cognition approach (Riou et al. 2014). Following this view, the decision-making process investigated here could be related to a continuous online adjustment between the embodied representation and the motor behavior of the participant (Lepora and Pezzulo 2015). This line of research has to be carefully addressed in further experiments.

Furthermore, the effect of the perceptual uncertainty (fog) was strongly mediated by the collection of visual evidence (both target detection and verification). Taken together, these results clearly show that the collection of visual evidence—as determined by both the perceptual uncertainty and the prior expectations—contributes significantly to the response selection.

Acknowledgements This study was approved by the French Comité de Protection des Personnes “Sud-Est VI” (IRB00008526, CE65, 2014). The authors would like to thank Annike Smeding for her comments on the submitted version of this article.

Funding This work was funded by grants from the French program “investissement d’avenir” managed by the National Research Agency (ANR), from the European Union (Auvergne European Regional Development Funds -ERDF- of Auvergne region) and from the “Région Auvergne” in the framework of the IMobS3 LabEx (ANR-10-LABX-16-01).

References

- Baayen RH, Davidson DJ, Bates DM (2008) Mixed-effects modeling with crossed random effects for subjects and items. *J Mem Lang* 59(4):390–412
- Bruhn P (2013) Emergence of spontaneous anticipatory hand movements in a probabilistic environment. *Adv Cogn Psychol* 9(2):62–73
- Castelhano MS, Heaven C (2010) The relative contribution of scene context and target features to visual search in scenes. *Atten Percept Psychophys* 72(5):1283–1297
- Castelhano MS, Pollatsek A, Cave KR (2008) Typicality aids search for an unspecified target, but only in identification and not in attentional guidance. *Psychon Bull Rev* 15(4):795–801
- Dale R, Kehoe C, Spivey MJ (2007) Graded motor responses in the time course of categorizing atypical exemplars. *Mem Cognit* 35(1):15–28
- Drugowitsch J, Moreno-Bote R, Churchland AK, Shadlen MN, Pouget A (2012) The cost of accumulating evidence in perceptual decision making. *J Neurosci* 32(11):3612–3628
- Flowers JH (1990) Priming effects in perceptual classification. *Percept Psychophys* 47(2):135–148
- Freeman JB, Ambady N (2010) Mouse tracker: software for studying real-time mental processing using a computer mouse-tracking method. *Behav Res Methods* 42(1):226–241
- Friston K, Adams RA, Perrinet L, Breakspear M (2012) Perceptions as hypotheses: saccades as experiments. *Front Psychol* 3:1–20
- Hayes AF (2009) Beyond Baron and Kenny: statistical mediation analysis in the New millennium. *Commun Monogr* 76(4):408–420
- Hollingworth A, Henderson JM (1998) Does consistent scene context facilitate object perception? *J Exp Psychol Gen* 127(4):398–415
- Kietzmann T, Geuter S, König P (2011) Overt visual attention as a causal factor of perceptual awareness. *PLoS ONE* 6(7):e22614
- Lepora N, Pezzulo G (2015) Embodied choice: how action influences perceptual decision making. *PLoS Comput Biol* 11(4):e1004110
- Malcolm GL, Henderson JM (2009) The effects of target template specificity on visual search in real-world scenes: evidence from eye movements. *J Vis* 9(11):1–13
- Riou B, Rey AE, Vallet GT, Cuny C, Versace R (2014) Perceptual processing affects the reactivation of a sensory dimension during a categorization task. *Q J Exp Psychol* 68(6):1–8

Article 4

Étude 3, expérience 2

A shape in the fog: Expectations, visual uncertainty and the accumulation of evidence in visual search using a combination of mouse and eye tracking methods

Boris Quétard,

Michèle Colomb,

Jean-Charles Quinton,

Giovanni Pezzulo,

Laura Barca,

Marie Izaute,

Martial Mermillod

Journal: Aucun

Statut: Non soumis

A shape in the fog: Expectations, visual uncertainty and the accumulation of evidence in visual search using a combination of mouse and eye tracking methods.

Boris Quétard ^{a *},

Michèle Colomb ^b,

Jean-Charles Quinton ^c,

Giovanni Pezzulo ^d,

Laura Barca ^d,

Marie Izaute ^a,

Martial Mermillod ^e,

^a Université Clermont Auvergne, CNRS, LAPSCO, F-63037 Clermont-Ferrand, France

^b CEREMA, Département Laboratoire de Clermont-Ferrand, F-63017 Clermont-Ferrand, France

^c Université Grenoble Alpes, CNRS, LJK, F-38000 Grenoble, France

^d Institute of Cognitive Science and Technologies, CNR, Rome, Italy

^e Université Grenoble Alpes, CNRS, LPNC, F-38040, Grenoble, France

*Corresponding author. LAPSCO 34 avenue Carnot 63037 Clermont Ferrand Cedex1. E-mail:

boris.quetard@uca.fr.

Secondary contacts: G. Pezzulo: giovanni.pezzulo@isct.cnr.it; M. Izaute: marie.izaute@uca.fr.

Abstract

While driving in foggy conditions and encountering an obstacle, identifying it can be crucial to anticipate what can happen in the next few seconds to avoid it. A human pedestrian may not react the same way a roe deer would. This can be seen as a visual search task for which visual uncertainty and expectations about target identity influence a decision-making process. Then, eye movements allow for actively accumulating evidence by detecting the target and verifying its identity, thus resolving the decision. We conducted a visual search experiment with driving-themed visual scenes, constructed with a realistic fog in which the participants had to determine whether a pedestrian was *present* or *absent*. A road sign indicated whether an object, a pedestrian (target) or a roe deer (distractor), was more likely to be in the scene. The scene could also include none of them. We used eye-tracking to measure object detection and verification processes and mouse-tracking to measure the decision-making process in a richer and more continuous way. Fog level and incongruent road signs (incorrect expectations) increased decision times and biased response selection towards the incorrect response. Target and distractor detection mediated the effects of fog and target verification the effects of expectations. Furthermore, this study shows that spending time searching for the object accumulates evidence towards the *target absent* response. Hence, in visual search, noisy visual evidence is accumulated and integrated with expectations through eye movements related to detecting and verifying target and distractor. Thus, accumulation of evidence is an active process in which eye movements play a crucial role.

Keywords: fog, driving, visual search, perceptual decision-making, eye movements, mouse-tracking.

Introduction

An agent must achieve many goals and tasks in his everyday life that require searching for objects or for other agents in the environment and identifying them. For instance, assessing the dangerousness of a situation when driving in the fog requires to quickly detect the blurry shape on the side of the road and to accurately identify it as either a pedestrian who could try to cross the road, or as a non-moving object such as a bush. For a driver, even quickly detecting and discriminating a pedestrian from an animal can be of critical importance as they may not behave the same way.

In these ecological situations where the visual input can be very complex and noisy, it is often necessary to direct the gaze at different locations of the environment to detect or search for an object or an obstacle. In this perspective, eye movements are goal-directed movements allowing to reduce the uncertainty about the elements of the scene which are pertinent in achieving the goals of the task (Donnarumma, Costantini, Ambrosini, Friston, & Pezzulo, 2017; Friston, Adams, Perrinet, & Breakspear, 2012; Kietzmann, Geuter, & König, 2011; Land & Hayhoe, 2001; Renninger, Verghese, & Coughlan, 2007; Sprague & Ballard, 2003). In order to rapidly direct the gaze at the relevant areas, candidate locations for fixations are selected by combining both early low-level visual information (Borji, Tavakoli, Sihite, & Itti, 2013; Itti & Baldi, 2009; Tatler, Hayhoe, Land, & Ballard, 2011) and prior knowledge, for instance about context (Bar, 2004, 2009) or target location and appearance (Castelhana & Heaven, 2010; Malcolm & Henderson, 2010; Malcolm & Shomstein, 2015; Quétard et al., 2016; Spotorno, Malcolm, & Tatler, 2014; Wu, Wick, & Pomplun, 2014). Models integrating these sources of information have been implemented for gaze allocation in visual search (Ehinger, Hidalgo-Sotelo, Torralba, & Oliva, 2009; Kanan, Tong, Zhang, & Cottrell, 2009; Torralba, Oliva, Castelhana, & Henderson, 2006).

This active visual exploration can be seen as an active sampling with the accumulation of evidence when engaged in perceptual decision-making tasks performed towards determining the identity of the object being seen, or determining whether the target is present or absent. For instance, it has been shown that eye movements directed at different parts of ambiguous figures lead to decide which object it represents while being themselves directed by early hypotheses about the object identity (Kietzmann et al., 2011). See also Quinton, Volpi, Barca and Pezzulo (2014) and Volpi, Quinton and Pezzulo (2014) for related computational modelling studies. These studies highlight how eye movements, by directing the high-resolution foveal area of the retina towards informative parts of the object, could be used to test hypotheses about its identity, ultimately reducing the uncertainty of the noisy sensory representation (Friston et al., 2012). The competing perceptual hypotheses are built and updated on the basis of prior knowledge as well as visual information sampled and integrated progressively. The resulting collection of

information allows for the convergence of the perceptual choice (Bar, 2009; Gold & Shadlen, 2007; Hegdé, 2008). Conversely, the availability of top-down priors and the quantity and strength of sensory information about the object identity influences the choice of an adequate oculomotor strategy. Exploratory behaviour (shorter fixations, higher saccade amplitude) is adopted in situations of uncertainty that require accumulating further evidence, whereas an exploitation behaviour (longer fixations, lower saccade amplitude) is adopted when a high quantity of information is available (Kietzmann & König, 2015). This further supports the active nature of the collection of evidence and shows that, while guiding the choice process, eye movements behaviour is also modulated by the demands and progresses of decision-making.

Most decision-making studies use a standard binary forced-choice task aiming for the identification of a target or discrimination between two categories and it is common to study it as a decision-making process. The processes related to the decision-making aspects of the categorization can be measured using a mouse-tracking methodology. This methodology allows us to measure different aspects of the decision-making process in a richer way than reaction times tasks by tracking the computer mouse trajectories (Freeman & Ambady, 2010; Freeman, Dale, & Farmer, 2011; Spivey, Grosjean, & Knoblich, 2005); See also Song and Nakayama (2009) for the hand reaching version of this method.

Prototypical mouse-tracking tasks consist of the binary forced-choice categorization of a stimulus while the response movement begins at the bottom-centre of the display (*start* button) and the response buttons are located at its top-left and top-right corners. For instance, beside calculating the response time (RT) corresponding to the duration of the global process, the geometric area between the mouse trajectory and the ideal trajectory, computed as Area Under the Curve (AUC) allows to estimate the attraction exerted by the alternative incorrect response in the response selection process. Thus, more graded and continuous response selection bias can be captured compared to classic accuracy and reaction time studies (Freeman, Dale, et al., 2011; Lepora & Pezzulo, 2015).

This task has been successfully used to investigate perceptual decision-making for categorization of visual stimuli such as animals typical of broad categories such as *fishes* or *mammals* (e.g. cat, horse) or sharing features with the incorrect category (e.g., dolphin, whale) (Dale, Kehoe, & Spivey, 2007), discrimination of ambiguous visual stimuli representing animals through eye movements (Quinton et al., 2014) as well as integration of ambiguous visual information with stereotypes (Freeman & Ambady, 2009; Freeman, Penner, Saperstein, Scheutz, & Ambady, 2011; Johnson, Freeman, & Pauker, 2012) or high level category cues (Freeman & Ambady, 2011; Freeman, Ambady, Rule, & Johnson, 2008) as top-down information in face perception. Using this kind of tasks, we recently examined the influence of eye movements as collection of evidence, on the categorization of a target (pedestrian or deer) appearing at different locations in a simple driving-themed 3D scene while manipulating both uncertainty of visual

information and prior expectations (Quétard et al., 2015).

However, categorization tasks differ from visual search tasks, where the aim is to determine whether a specific target is *present* or *absent*. In this kind of tasks, one of the two responses is made on the basis of the absence of a stimulus. The standard view of decision as an accumulation of evidence for each competing hypotheses can also account for these tasks (*present vs. absent*) as proposed in a computational study (Dufau, Grainger, & Ziegler, 2012) where discriminating *non-words* from *words* is based on the lack of evidence for *words*. Studying visual search tasks in realistic scenes is rarely studied as a decision-making task and eye movements are usually examined separately from the outcome of the task, without taking into account their involvement and effectiveness in achieving its goal.

The present experiment investigates whether, in a driving-themed scene including a pedestrian (target), a deer (distractor), or none of them, visual exploration explains the effects of visual uncertainty and prior expectations on the decision-making process determining whether a pedestrian is present in the scene. In particular, we photographed a visual scenery in real fog and defined low or high fog levels to manipulate visual uncertainty and prior expectations were manipulated with road signs indicating whether a pedestrian (children sign) or a deer (wild animal sign) was more likely to appear. A danger sign indicated that both could appear with the same probability. The road signs were not informative about the probability that no object would be part of the scene (for each road sign, the probability that the scene included no object was 25%). Hence, the participants had to determine whether a pedestrian was part of the visual scene by clicking on *present* or *absent* response buttons, while the scene could contain one of two objects, a pedestrian (target) or a roe deer (distractor) at one of four possible locations, or none of these objects.

We divided the eye movements related to visual search into two epochs well defined in the literature (Castelhano, Pollatsek, & Cave, 2008; Malcolm & Henderson, 2009). The detection epoch elapses from the appearance of the scene to the first saccade ending within the target area of interest (AOI). Its duration (detection time or target latency) is a measure estimating the time spent before finding the target. The verification epoch includes the first sequence of fixations in the target AOI. Its duration (verification time) is a measure estimating the time spent identifying the target. We used mediation analysis to analyse the relationship these eye movement variables have with fog and expectations in explaining decision-making, measured by RT and AUC, in this visual search task. Beside emphasizing how eye movements are integrated in the decision-making as an accumulation of evidence mechanism, this methodology allows to examine the respective contributions of visual detection and identification processes. As the visual scene was rather simple, with the pedestrian and deer being the only changing elements in the scene across trials, the detection time could mainly be determined by peripheral detection of the target (few fixations during the detection epoch) whereas the verification time would still reflect the central recognition of the target.

Article 4 (Étude 3) – Recherche visuelle dans le brouillard : attentes et accumulation d'évidence

Nuthmann (2014) previously highlighted this distinction by testing directly the effects of degraded peripheral and central visual fields on these epochs of visual exploration.

We predicted that the fog would impair both target detection and verification because it would add uncertainty over the bottom-up visual information. Additionally, expecting for the target to be present should facilitate its identification and impair the distractor's identification, as measured by verification times and *vice-versa* when expecting for the distractor. In fact, since the appearance of the target and distractor are known, this prior knowledge can be used to recognize the target faster (Castelhano & Heaven, 2010; Malcolm & Henderson, 2009). If the knowledge of the object's features are used to detect it peripherally, the same predictions can be made for the detection times. Furthermore, the effects of fog and expectations on detection and verification times described above should induce corresponding effects on the mouse-tracking time measures. Thus, high fog should increase the RT whereas the road signs should decrease the RT for the congruent object and increase it for the incongruent object. Mediation and moderation analysis allowed to estimate whether these effects could be explained by facilitation or impairment of detection and verification processes by fog and expectations, highlighting the contribution of these two processes in the decision-making.

Considering the response selection, a sign congruent or incongruent with the object induces expectancies towards the *correct* or *incorrect* response respectively. Thus, we expected corresponding deviations of the mouse trajectories, *i.e.* lower AUC for congruent and higher AUC for incongruent signs. A large part of this effect is expected to be driven by motor anticipation since the road sign informs participants about which response button will probably be the correct one. Here, we focus on the part of the effect driven by perceptual processes as measured through eye movements. Part of the overall effect of road signs on AUC should be mediated by verification times, as expecting for a target whose appearance is known should facilitate its recognition by providing a template (Castelhano et al., 2008; Hout & Goldinger, 2015; Malcolm & Henderson, 2009). If this template also affects peripheral detection, a corresponding mediation of detection times should be found.

Finally, this is not clear whether high fog level would increase AUC. In one hand, if bottom-up uncertainty motivates the adoption of a careful response strategy where the participants do not make a clear choice until further evidence is collected, this would translate in increased AUC. Indeed, the participants would execute mouse trajectories deviating towards the centre of the screen. On the other hand, the fact that the object is harder to find could bias response selection towards the *absent* response button. As it is not immediately found, the *absent* response stays a valid alternative, thus attracting the mouse movement and increasing the AUC but only when the pedestrian is part of the scene (the *incorrect* response being the *absent* response in this case). This effect should be partly driven by impaired perceptual

collection of evidence and thus be mediated by detection and/or verification times. By contrast, if the only effect of bottom-up uncertainty is to motivate the adoption of a careful strategy in the execution of the response, then there should be no mediation of the effect of fog by detection/verification times on AUC.

This hypothesis implies that in decision-making oriented towards determining whether a target is *present* or *absent*, exploring the scene without finding the target would accumulate evidence towards the *absent* response. Hence, as proposed by Dufau et al. (2012), the accumulation of evidence for the *absent* response is a function of the lack of evidence for the *present* response. We previously observed another evidence of such a mechanism in the form of an effect of the detection times on the AUC when searching for target objects in indoor visual scenes, so that longer search for the target leads to trajectories deviated towards the *absent* response (Quétard et al., 2016). We also searched for such effect in the present experiment, as this enlighten directly how evidence is accumulated towards a no-stimulus alternative in realistic visual scenes. Especially, this experiment is particularly interesting for testing directly this hypothesis as we should not expect the same effect of detection on the AUC for a pedestrian target and a roe deer distractor. When a pedestrian is in the scene, the incorrect response is *absent*, so longer detection times should increase the AUC (deviation towards the incorrect *absent* response). On the contrary, when a roe deer is in the scene, longer detection times should either reduce the AUC (less deviation towards the incorrect *present* response) or have no effect.

Method

Participants

Forty-nine psychology students (10 males, mean age = 20.07, sd = 1.98) were part of the experiment, which took place at Université Clermont Auvergne (France). They completed written consent form for their participation in the experiment. All of them had a normal or corrected-to-normal vision (glasses or contact lenses).

Stimuli

The visual scenes containing the target were pictures depicting a driving situation at night, where a driver (point of view) encounters either a pedestrian child, a roe deer or none and is correctly or incorrectly advertised by a road sign: children crossing the road, wild animal, or generic danger (Figure 1). They were photographed at the CEREMA's laboratory of Clermont Ferrand, in a room able to produce fog with the physical properties of a natural fog (Colomb et al., 2008). They were taken from an adjacent room simulating the cockpit of a car and thus, depicted a driver's point of view. The fog room included a one-way road and three elements were placed in order to build the scene: A car on the road (14 m from the observer), a road sign rightward of the road (7.5 m), and a dressed child mannequin or a stuffed roe deer on one of 4 positions (all positions at 12 m). Since the mannequin was taller than the roe deer, we dressed it with a dark coat and a lighter trouser in order increase the salience of its lower features, which moves the pedestrian's salient features at the same height than the roe deer's. The 4 possible positions of the child and roe deer were rightward of the road at approximately 6.2° and 2.7° from the centre of the picture on X-axis and leftward of the road at approximately 3.5° and 7°. Additionally, a transmissiometer measured the fog density (in meters of visibility).

We took different series of pictures, each of the series including either an object (pedestrian or roe deer) placed at one of four positions (total of 8 series) or a road sign with no object (3 series). Each series was made of thirty nine pictures at different fog densities (from 6 to 45 meters of visibility). From each series, we selected one picture with a low fog and one picture with a high fog. For the series including an object, we asked 10 participants to rate each picture. We presented the pictures series by series, from the highest fog density picture to the lowest. The series were presented in a random order. The participants used a Likert scale ranging from 7 (invisible object) to 1 (object not degraded) while the 5th point indicated that the object could be identified when *a priori* knowing its appearance and location, and the 4th point indicated an identifiable but strongly degraded object. For each participant and each series, we selected the highest fog density among those rated 4 as identification threshold and tested the inter-rater reliability

Article 4 (Étude 3) – Recherche visuelle dans le brouillard : attentes et accumulation d'évidence

using intra-class correlation ($ICC = .89$, $CI\ 95\% = [.79; .96]$). Within each series, the high fog density picture was chosen as the picture whose fog density corresponded to the rounded mean of the identification thresholds. Since the fog density was only measured at one spatial location (where the transmissiometer was), the degradation of the object varied a lot at low fog densities because of the formation of eddies. Thus, we chose the low fog density picture among the pictures rated 2 (object slightly degraded) and with at least a difference in fog density of ten meters of visibility with the high fog picture, to ensure that our low and high fog pictures differed enough. Finally, we calculated the mean low and high fog density of the previously selected pictures and took these values for the selection of the low and high fog pictures which included a road sign but no object. A total of 11 pictures were selected for each of the two fog conditions (4 pedestrian, 4 roe deer and 3 road sign pictures).

After selecting these pictures, we relied on photo editing techniques implemented in Matlab for the purpose of this experiment to generate various combinations of road signs, targets (pedestrian or roe deer) and fog levels (Qué tard et al., 2014). We were thus able to control for image variability (especially due to the fog), while also eliminating possibly salient yet non-ecological elements present in the scene. Such elements included the laser beam emitted by the transmissiometer (required for fog density estimation) or the ceiling pipes producing the fog droplets. Applying the same method, we built three different no-object pictures for each of the two fog levels and for each road signs. This allowed displaying varying configurations of fog for the no-object scenes for each road sign, as this was the case for the scenes containing a pedestrian or a roe deer. In these scenes, there was a different fog configuration by object position. In total, 24 scenes were built for each object, pedestrian or roe deer (4 position, 2 fog levels, 3 road signs) and 18 scenes with no-object (3 fog configurations, 2 fog levels and 3 road signs). All had a resolution of 1350 x 900 pixels. We completed the stimuli set with three 400 x 400 pixels pictures depicting the three triangular road signs mentioned above.

Apparatus

To record the eye movements, we used a SMI RED-m eye-tracker at a sampling rate of 120 Hz. The SMI's dispersion-based algorithm (based on Salvucci & Goldberg, 2000) ensured the identification of the eye fixations. The minimum duration threshold was set at 100 ms and the minimum dispersion threshold at 1° (within the optimal range tested by Blignaut, 2009). We used the MouseTracker software (Freeman & Ambady, 2010) to program and run the experiment. The stimuli were presented using a laptop with a 15.6-inch display set at a resolution of 1600 x 900 pixels and a refresh rate of 60 Hz. A chin and forehead rest stabilized the eyes approximately 62 cm away from the display. A regular mouse was used and the movements were recorded by the MouseTracker software at a sampling rate of 70 Hz.



Figure 1: Examples of visual scenes under various levels of fog. 1st row: Roe deer (distractor) under low fog primed with incongruent expectations. 2nd row/left: No target with low fog and no expectations. 2nd row/right: Roe deer under high fog with congruent expectations. 3rd row/left: Pedestrian (target) under low fog with congruent expectations. 3rd row/right: Pedestrian under high fog with no expectations.

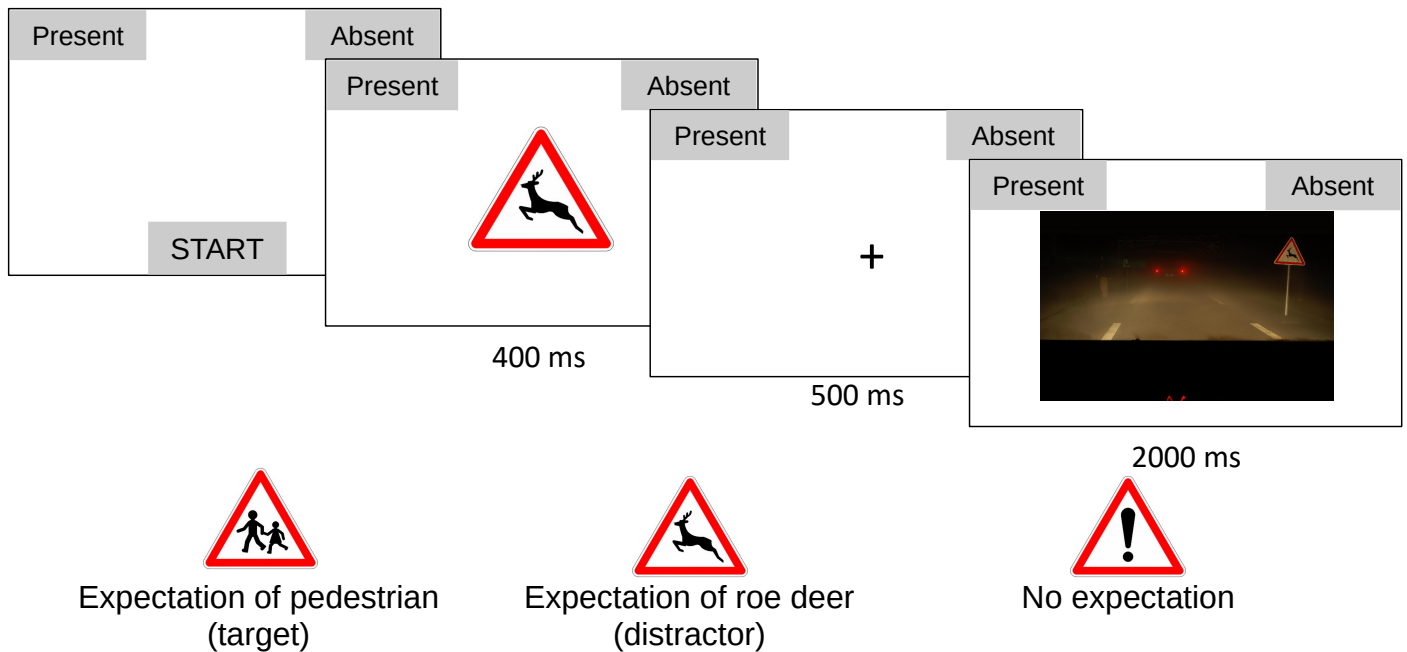


Figure 2: Time course of a trial. The road sign primed was always the same as the road sign included within the scene. For the predictive road signs (children and wild animals), there were more congruent trials than incongruent. For the danger road sign, there were as much trials with a pedestrian than with a roe deer or with nothing.

Procedure

Upon beginning the session, the instructions were given to the participants and they performed a training block of 15 trials. Then, we set the eye-tracking through a five-points calibration and a four-points validation procedure, ensuring the mean error on the X and Y axes was less than 0.5°.

Figure 2 illustrates the time course of a standard trial. The display consisted of two response buttons, *present* and *absent*, located at the top corners of the screen and one start button located at the bottom of the screen. The participants had to click on the start button to begin each trials. A fixation cross appeared on the centre of the screen for 300 ms, followed by a road sign prime for 400 ms and a second fixation cross for 500 ms. The mouse cursor was locked on the start button for this total duration. Finally, the visual scene including the same road sign than the prime and a low or high fog level, was presented for 2000 ms. The participants had to judge whether a pedestrian (the target) was *present* and click on the corresponding response button before 2000 ms elapsed. Thus, when a roe deer (the distractor) or none of the two was in the scene, they had to respond *absent*. When their mouse movements began after 600 ms, they were asked to begin their response quicker for the next trials.

The participants were also instructed that a children or animal road sign indicated that there was a

Article 4 (Étude 3) – Recherche visuelle dans le brouillard : attentes et accumulation d'évidence

higher probability that, respectively, a pedestrian or a roe deer would appear in the scene whereas a danger sign indicated an equal probability. When a danger sign was presented, there was 16 trials including a pedestrian, a roe deer or none of them (neutral object-sign relationship). When a predictive road sign (children or animal) was presented, the number of trials was higher for congruent object-sign relationships (pedestrian-children sign and roe deer-animal sign) than for incongruent relationships (pedestrian-animal sign and roe deer-children sign). We added a manipulation of the strength of the expectations generated by the animal and children signs. The trials were dispatched between the congruency conditions following two configurations where these road signs were either moderately or highly predictive of the appearance of the target or distractor, with separate participants samples (respectively 24 and 25 participants). For each predictive road sign (children and animal), there was respectively 32 (moderate) or 40 (high) congruent trials and 16 (moderate) or 8 (high) incongruent trials. When neither a pedestrian nor a roe deer was included in the scene, there was 16 trials by road sign.

Finally, half of the participants were displayed leftward *present* and rightward *absent* response buttons and the opposite for the other half.

Results

Data analysis

Data processing. We only analysed the conditions where either the target or the distractor was present and included both the participants with moderately and highly predictive road signs. Six participants were discarded due to unreliable mouse-tracking (e.g. abnormally high number of time-out trials) or eye-tracking data (e.g. weak calibration). For the remaining participants, the trials with invalid eye-tracking data were also excluded for the following reasons: The gaze being directed on the object when the scene appeared (9.6 %), there was no fixation on the object in the trial (10.8 %) or the detection or verification time was below 80 ms (0.7 %). Then, we excluded the erroneous responses (0.9 %) and the time out responses (2.4 %).

In order to analyse the mouse responses, we pre-processed the mouse trajectories as described in Freeman and Ambady (2010) using the Mouse-tracker software's Analyzer. We computed the response times (RT) and, after time-normalizing the trajectories, we computed the areas under the curve of the trajectories (AUC). As eye-tracking variables of interest, we computed the detection and verification times. The detection time represented the time elapsed from the appearance of the scene to the first saccade on

the target. The verification time was calculated as the duration of the first sequence of fixations directed on the target before the gaze was directed elsewhere. As in Malcolm & Henderson (2009) if the second fixation of this sequence was not directed on the target but was immediately followed by a fixation on the target, it did not terminate the verification epoch since it could be part of a pre-programmed sequence of saccades but it was not included in the calculation of the verification time.

Inferential statistical models. We used linear mixed models (LMM) to assess the effects of congruency and fog on these measures, these models being able to handle heavily unbalanced and within-subjects designs such as ours (Baayen, Davidson, & Bates, 2008). We analysed separately the trials including a pedestrian and those including a roe deer as we expected different results, especially with the mouse deviation (AUC). The congruency variable was recoded using orthogonal Helmert contrasts. The congruency contrast C1 tests the difference between congruent (-0.5) and incongruent (+0.5) road signs and C2 the difference between the absence of expectations (neutral sign, -0.66) and their presence (congruent and incongruent signs, +0.33). The fog variable was recoded such as low fog = -0.5 and high fog = +0.5. As random parameters, we included the random intercept and slopes for the fog and both congruency contrasts as well as the correlations between random parameters for the subjects variable, which was the maximal random structure leading to convergence of the models (Barr, Levy, Scheepers, & Tily, 2013). Finally, we included the prediction strength of the road signs and its interactions with other variables as fixed parameters. While it is out of the focus of this study, this between-subject variable is important in the structure of the data, hence the importance of taking it into account in these inferential statistical models. In order to estimate the p -values of the fixed effect parameters, we used the Kenward-Rogers approximation of the degree of freedom (df). In order to clarify the interaction effect, we estimated the conditional effects and their 95% confidence intervals using the "lsmeans" R package with the Kenward-Rogers approximation of the df . For linear models, this package estimates the slope of a variable of interest for each specified values of the interacting variable.

Mediation and moderation. We evaluated whether the detection and verification times mediated or moderated the effects of the fog and congruency contrasts (causal variables, CV) on the mouse movements measures (outcomes). As these analyses involve adding mediator variables and their interactions to the previous and already complex model, we excluded the expectation strength variable from these analyses to keep more parsimonious statistical models and reduce the risk of non-convergence. A moderation of these effects by the detection or verification times was assessed by testing the interaction between the moderator and the CV (Judd, Yzerbyt, & Muller, 2014). If no moderation was found, we assessed the

mediating effects of the detection and verification times using the joint significance method which consists in testing the significance of the CV's effect on the mediator (path a) and of the mediator's effect on the outcome while controlling for the CV (path b). If both of these conditions are met, then the total effect of the CV on the outcome (path c) is at least partially mediated (Judd et al., 2014).

As we tested these effects within models including all of the CVs, one mediator and all interactions (as suggested by Fairchild & MacKinnon, 2009), mediations could be moderated by another CV (moderator). Moderated mediations can be assessed with joint significance if either, both path a and the interaction between path b and the moderator (noted path b_m) are significant, or both path b and the interaction between path a and the moderator (path a_m) are significant (Muller, Judd, & Yzerbyt, 2005). In these cases, the part of the total effect which is mediated varies across the modalities of the moderator. A weakness of this method for assessing mediations is that it does not provide any confidence interval for the part of the effect that is mediated, *i.e.* the indirect effect ab (calculated as path a x path b). In order to provide such information, we used the standard bootstrapping procedure (Hayes, 2009; Preacher & Hayes, 2008). Hence, in the case of moderated mediations, this allowed to estimate in a more comprehensive way, the variations of the indirect effect across the moderator's modalities.

We estimated the 95 % confidence intervals (percentile method) of the indirect effects through 5000 bootstrap samples. It is to be noted that the bootstrap estimation of the indirect effect is one of the most usual ways for assessing the presence of a mediation, but inflated *Type I* error rates have been reported under certain conditions (Judd et al., 2014), which is not the case for the joint significance method. This makes reporting the results using both methods strongly relevant. To estimate the magnitude of mediation (for each modalities of the moderator in the case of moderated mediations), we used the proportion by which the total effect (path c) is reduced when controlling for the effect of the mediator (path c'). Though this method is not perfect, it is a simple way to provide this information (for a short discussion, see Hayes, 2009). To estimate all the parameters (a , b , a_m , b_m , c and c'), we used LMMs including both CVs and the mediator (when needed) with all the interactions as fixed effects. The dependant variables were either the mouse-tracking variables (outcomes) or the mediator when required. We specified the same random structure for all models, *i.e.* the intercept, slopes of the CVs and correlations between random parameters nested within the subject random factor.

Eye movement measures

Pedestrian in the scene (target). See Table 1 for the mean detection and verification times by experimental conditions and their standard error, for visual scenes including a pedestrian. High fog level induced higher target detection times ($b = 28.95$, $t(63.6) = 6.19$, $p < .001$) whereas neither the difference between congruent and incongruent road sign (C1: $b = 0.9$, $t(49.1) = 0.16$) and the difference between the presence (animal and children signs) and absence (danger sign) of expectations were significant (C2: $b = -0.16$, $t(42.2) = -0.04$). The interactions between fog and each congruency C1 was not significant but the effect of fog interacted with the presence of expectations (C2) ($b = 19.35$, $t(2063.4) = 2.26$, $p = 0.024$). To clarify this interaction, we estimated of the conditional effects of fog for each values of congruency C2., Fog increased detection times in a bigger extent when some expectations were induced ($b_{F,C2=0.33} = 32.60$; $CI = [22.90; 42.30]$, $df = 68.6$) than when no expectations were induced ($b_{F,C2=-0.66} = 13.47$; $CI = [-1.38; 28.33]$, $df = 369.4$).

Table 1:

Mean (standard error) by mouse-tracking measure and eye-tracking measure as a function of fog and prime-target congruency when a pedestrian is presented (target)

	low fog			high fog		
	congruent	incongruent	neutral	congruent	incongruent	neutral
response time	1163.7 (9.8)	1246.7 (15.2)	1185.6 (13.4)	1187.4 (10.2)	1263.7 (16.4)	1225.4 (13.2)
area under the curve	0.30 (0.03)	0.65 (0.08)	0.44 (0.05)	0.42 (0.04)	0.74 (0.08)	0.63 (0.07)
detection time	301.7 (3.5)	296.5 (4.7)	307.2 (5.5)	328.3 (3.4)	335.2 (6.3)	323.4 (5.3)
verification time	236.7 (5)	272.6 (11.4)	246.9 (7.8)	249.8 (5.2)	255.4 (8.1)	267.5 (8.7)

For verification times, there was no effect of fog ($b = 2.13$, $t(72.2) = 0.33$). Incongruent sign increased the target verification times compared to congruent sign ($b = 27.05$, $t(48.7) = 3.18$, $p = .003$), and there was a significant interaction with fog ($b = -42.73$, $t(2068.6) = -3.20$, $p = .001$). As pointed out by the estimated conditional effects, the increase in target verification times due to incongruent compared to congruent sign, was lower (and non significant) with high fog ($b_{C1,F=0.5} = 5.65$; $CI = [-16.16; 27.47]$, $df = 115.5$) than with low fog ($b_{C1,F=-0.5} = 48.92$; $CI = [27.72; 70.12]$, $df = 106.5$). No significant difference was found between the presence and absence of expectations ($b = -0.61$, $t(53.3) = -0.08$) but this contrast (C2) interacted with fog ($b = -23.26$, $t(2058.2) = -2.00$, $p = .045$). Thus, fog increased verification times when no expectations were induced ($b_{F,C2=-0.66} = 26.20$; $CI = [5.85; 46.55]$, $df = 345.8$) and this effect was reduced near zero when expectations were induced ($b_{F,C2=0.33} = 2.37$; $CI = [-11.06; 15.81]$, $df = 66.76$). This interaction was itself bigger for a high strength of expectations (bigger difference in fog effects between absence and

Article 4 (Étude 3) – Recherche visuelle dans le brouillard : attentes et accumulation d'évidence

presence of expectations) than with a medium strength, as indicated by the three ways interaction we observed between fog, congruency C2 and the strength of the expectations ($b = -64.97$, $t(2058.2) = -2.80$, $p = .005$). We do not add further details of this three-way interaction in order to simplify our results.

For both eye movement measures, we found effects of the strength of expectations, and no interaction of this variable with the other fixed parameters (all $|t| < 1.7$, $p > .05$), besides the aforementioned three-ways interaction with fog and congruency C2.

Deer in the scene (distractor). See Table 2 for the mean detection and verification times by experimental conditions and their standard error, for visual scenes including a deer. Similarly to scenes containing a pedestrian, a high fog level induced longer distractor detection times ($b = 62.93$, $t(67.3) = 13.33$, $p < .001$) but there was no effect of fog on distractor verification times ($b = 5.00$, $t(70) = 0.83$). Conversely, incongruent road sign did not significantly affect detection times compared to congruent road signs ($b = 3.84$, $t(49) = 0.80$) but they increased verification times ($b = 18.77$, $t(50.3) = 2.10$, $p = .041$). The difference between presence and absence of expectations was neither significant on detection times ($b = -3.64$, $t(44.8) = -0.80$) nor on verification times ($b = -9.44$, $t(54.4) = -1.39$, $p = 0.17$). No interaction was significant between fog and both congruency contrasts on these measures, and we found no effect of expectation strength nor any significant interaction with fog or congruency contrasts (all $|t| < 1$, $p > .1$ for detection times and all $|t| < 1.5$, $p > .1$ for verification times).

Table 2:

Mean (standard error) by mouse-tracking measure and eye-tracking measure as a function of fog and prime-target congruency when a deer is presented (distractor)

	low fog			high fog		
	congruent	incongruent	neutral	congruent	incongruent	neutral
response time	1156.5 (11.1)	1208 (15.5)	1179.9 (14.3)	1188.9 (11.6)	1225.8 (15.1)	1215.6 (15.7)
area under the curve	0.28 (0.03)	0.52 (0.08)	0.30 (0.05)	0.23 (0.03)	0.45 (0.07)	0.34 (0.05)
detection time	272 (2.8)	275.1 (3.8)	279.5 (5.5)	335.3 (4.2)	341.4 (5.6)	338 (4.6)
verification time	240.6 (5.5)	248.7 (10)	260.4 (10.2)	249.3 (5.6)	262.5 (9.5)	266.5 (9)

Response times

Pedestrian in the scene (target). Figure 3A (top) displays the mean RT by experimental conditions and their standard error when a pedestrian is in the scene. High fog level significantly increased the RT compared to low fog level ($b = 22.20$, $t(66.9) = 2.38$, $p = .020$). We observed an effect of the strength of

Article 4 (Étude 3) – Recherche visuelle dans le brouillard : attentes et accumulation d'évidence

expectations ($b = -142.16$, $t(43.6) = -3.72$, $p < .001$), with expectations with high strength diminishing the RT compared to medium strength. Incongruent road sign led to longer RT compared to congruent road sign (C1: $b = 58.58$, $t(52.4) = 3.65$, $p < .001$) whereas the presence of expectations did not significantly influence RT compared with their absence (C2: $b = 7.25$, $t(46.1) = 0.70$). However, we observed an interaction of expectation strength with both congruency C1 ($b = 76.76$, $t(52.4) = 2.39$, $p = .020$) and congruency C2 ($b = 54.46$, $t(46.1) = 2.62$, $p = .012$). Hence, in the presence of expectations (children and animal signs) response times were longer compared with the absence of expectations (danger), but only when the strength of expectations was high (Medium: $b_{C2,S=0.5} = -19.93$; $CI = [-48.99; 9.13]$, $df = 35.56$; High: $b_{C2,S=0.5} = 34.66$; $CI = [4.22; 65.10]$, $df = 46.37$). This is explainable by the incongruent road signs strongly increasing the RT for high expectation strength, as highlighted by the interaction of expectation strength with congruency C1 where incongruent signs increased the RT compared to congruent signs in a larger extent for high expectation strength ($b_{C1,S=0.5} = 97.09$; $CI = [49.51; 144.66]$, $df = 51.22$) than for medium expectation strength ($b_{C1,S=0.5} = 20.20$; $CI = [-23.93; 64.33]$, $df = 33.45$). No other interactions between congruency contrasts and fog variables were found (all $|t| < 1.6$, $p > .1$).

We expected the effects of fog and congruency on RT described above to be mediated by eye movements measures such as detection and verification time. For the following mediation analyses (and for every mediation/moderation analyses in this article), we used statistical models excluding the expectation strength variable. Hence, the effect of the fog on RT was mediated by the target detection times, as assessed by the joint significance of the effect of fog on detection time (path *a*) and the effect of detection time on RT (while controlling for the effect of fog, path *b*). See Figure 3B (top) for a schematics of this mediation for RT. Consistently with the analyses presented in the precedent section, fog significantly increased detection times ($b = 27.56$, $t(60.1) = 6.05$, $p < .001$) and longer detection times increased RT ($b = 0.61$, $t(2034.8) = 11.02$, $p < .001$). Furthermore, this mediation was moderated by congruency C2, since both the interaction with fog on detection times ($b = 17.40$, $t(2069) = 2.06$, $p = .040$) and with detection times on RT ($b = 0.24$, $t(1539.6) = 2.18$, $p = .029$) were significant.

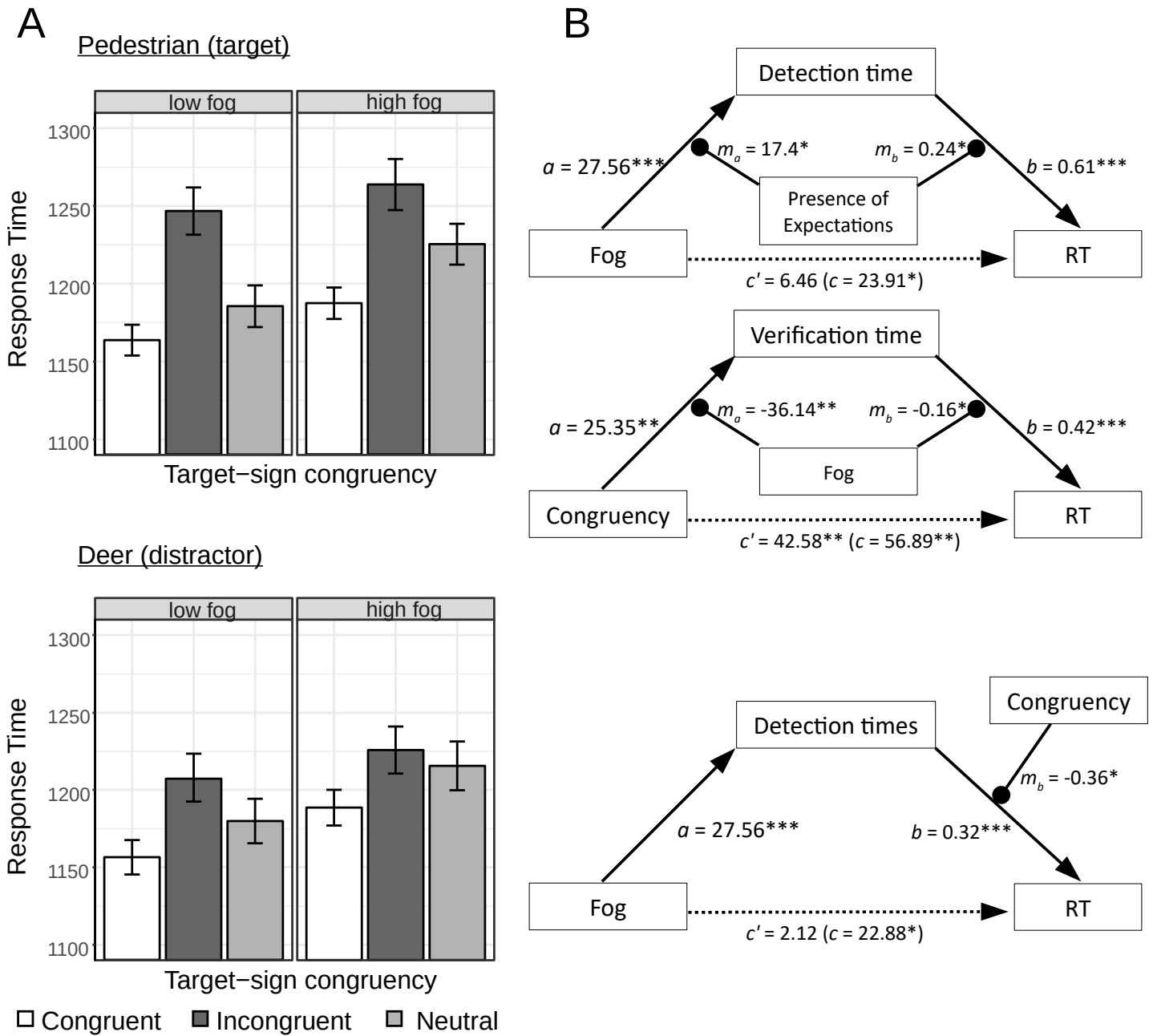


Figure 3: Overview of the results for Response Times (RT) when: 1st row, the object included in the scene is a pedestrian (target); 2nd row, the object included in the scene is a deer (distractor). **A.** Mean RT as a function of road sign congruency and fog level. Error bars display the standard errors. **B.** Mediation of the effects of fog and object-road sign congruency (noted Congruency) on RT by the eye movement measures, respectively detection times and verification times. Congruency corresponds to the congruency contrast C1 and Presence of expectations corresponds to the congruency contrast C2. Dashed arrows represent an effect mediated by detection or verification times. Arrows ending with a black dot represent moderation (interaction) effects. Parameters a and b are displayed as well as the direct effects c' and total effect c for the mediated effects. Parameters m_a and m_b correspond to the moderation of, respectively, paths a and b by a moderator. * $p < .05$, ** $p < .01$, *** $p < .001$.

For the absence of expectations, the total effect of fog on RT was reduced by 17.8% when controlling for the detection time ($c = 41.15$, $c' = 33.82$) with an indirect effect significantly different from zero ($ab = 7.18$, $CI = [0.52, 17.47]$), as assessed with bootstrap 95% confidence intervals. By contrast, for the presence of expectations, the total effect of fog on RT was reduced by 148.3% when controlling for the detection time ($c = 15.29$, $c' = -7.39$), with a bigger and significant indirect effect ($ab = 22.81$, $CI = [15.30, 31.33]$). This means that the effect of fog on RT is mediated by detection time when expectations are induced while it is less the case when no expectations are induced. Consistently with these results, we observed an interaction between fog and congruency C2 on RT when controlling for the effect of detection times (and thus, subtracting the part of the effect of fog mediated by detection time), in which fog increases RT in a larger extent without expectations than with expectations ($b = -41.63$, $t(2076.3) = -2.23$, $p = .026$). Thus, when considering the direct effect of fog on RT, we find that it is larger without expectations, which is to be expected since the total effect of fog on RT is explained by the indirect effect (through detection time) when expectations are induced.

Figure 3B (top) displays the mediation of the congruency effect (C1) on RT by verification times. Hence, incongruent road signs significantly increased verification times compared with congruent road signs ($b = 25.35$, $t(45.2) = 2.99$, $p = .004$) and longer verification times increased RT ($b = 0.42$, $t(1868.4) = 11.20$, $p < 0.001$). Furthermore, this mediation was moderated by the fog level as both the interactions of fog with congruency C1 on verification times ($b = -36.14$, $t(2082.2) = -2.80$, $p = .005$) and of fog with verification times on RT were significant ($b = -0.16$, $t(709.6) = -2.31$, $p = .021$). For low fog, the total effect of congruency C1 on RT was reduced by 40.7% when controlling for the verification time ($c = 60.82$, $c' = 36.04$), with a significant indirect effect ($ab = 21.68$, $CI = [9.58, 35.29]$). However, for high fog, the total effect of congruency C1 on RT was reduced by 7.3% when controlling for the verification time ($c = 52.97$, $c' = 49.12$), and the indirect effect was not significant ($ab = 2.49$, $CI = [-2.85, 9.80]$).

Deer in the scene (distractor). Figure 3A (bottom) displays the mean RT by experimental conditions and their standard error when a roe deer is in the scene. When the visual scenes included a deer, high fog level induced longer RT compared to low fog level ($b = 21.55$, $t(72.3) = 2.12$, $p = .038$). However, in contrast with scenes including a pedestrian, no effect of congruent vs. incongruent road sign nor of presence vs. absence of expectations were found (C1: $b = 17.42$, $t(57.7) = 1.03$, $p = .306$; C2: $b = -11.67$, $t(44.7) = -1.07$, $p = .289$). High expectation strength significantly decreased the RT compared with medium expectation strength ($b = -118.94$, $t(43.2) = -2.60$, $p = .013$). None of the interactions were significant (all $|t| < 1.2$, $p > .1$).

About the mediation analyses, both longer detection times ($b = 0.32$, $t(1733.7) = 4.41$, $p < .001$) and

Article 4 (Étude 3) – Recherche visuelle dans le brouillard : attentes et accumulation d'évidence

longer verification times ($b = 0.39$, $t(1710.5) = 9.96$, $p < .001$) increased the RT while controlling for the causal variables. Besides, the effect of detection times interacted with congruency C1 ($b = -0.36$, $t(1317.9) = -1.97$, $p = .05$). Hence, the time spent searching for the target increased the RT in a lesser extent for incongruent than congruent road sign. The effect of fog on RT was significantly mediated by detection times, as high fog increased the detection times ($b = 62.67$, $t(63) = 13.58$, $p < .001$). See Figure 3B (bottom) for a schematics of this mediation model. Congruency C1 moderating this mediation, the total fog effect was reduced by 94.7% when controlling for the detection times for congruent road sign ($c = 30.16$, $c' = 1.60$) with a significant indirect effect ($ab = 28.08$, $CI = [17.73, 41.65]$) and by 80.0% for incongruent road sign ($c = 16.9$, $c' = 3.37$) with a non-significant indirect effect ($ab = 13.12$, $CI = [-1.31, 32.58]$). There was no effect of congruency C1 or C2 neither on RT nor on detection times or verification times, hence no mediation of the congruency effects by the mediators.

Areas under the curve

Pedestrian in the scene (target). Figure 4 (left) displays the mean response trajectories by congruency conditions and Figure 5A (top) the mean AUC by experimental conditions and their standard error. There was an effect of the fog on the AUC, where high fog level induced trajectories more deviated towards the *absent* response than low fog ($b = 0.13$, $t(80.1) = 3.03$, $p = .003$). Strength of expectations interacted with this effect ($b = -0.18$, $t(80.1) = -2.05$, $p = .044$) with the increase of AUC by high fog being smaller for highly ($b_{F,S=0.5} = 0.06$; $CI = [-0.06; 0.17]$, $df = 42.2$) than for moderately reliable road signs ($b_{F,S=0.5} = 0.22$; $CI = [0.11; 0.34]$, $df = 41.9$). Additionally, deviations towards the *absent* response were higher for incongruent compared to congruent road signs (C1: $b = 0.37$, $t(53.8) = -3.18$, $p = .002$) but we found no significant difference between presence and absence of expectations (C2: $b = 0.02$, $t(46.6) = 0.42$). The effect of expectation strength ($b = 0.07$, $t(42) = 0.56$) as well as the interaction between congruency C1 and expectation strength ($b = 0.40$, $t(53.8) = 1.74$, $p = .088$) were non significant. All of the interactions other than the ones previously reported were not significant either ($|t| < 1$, $p > .1$).

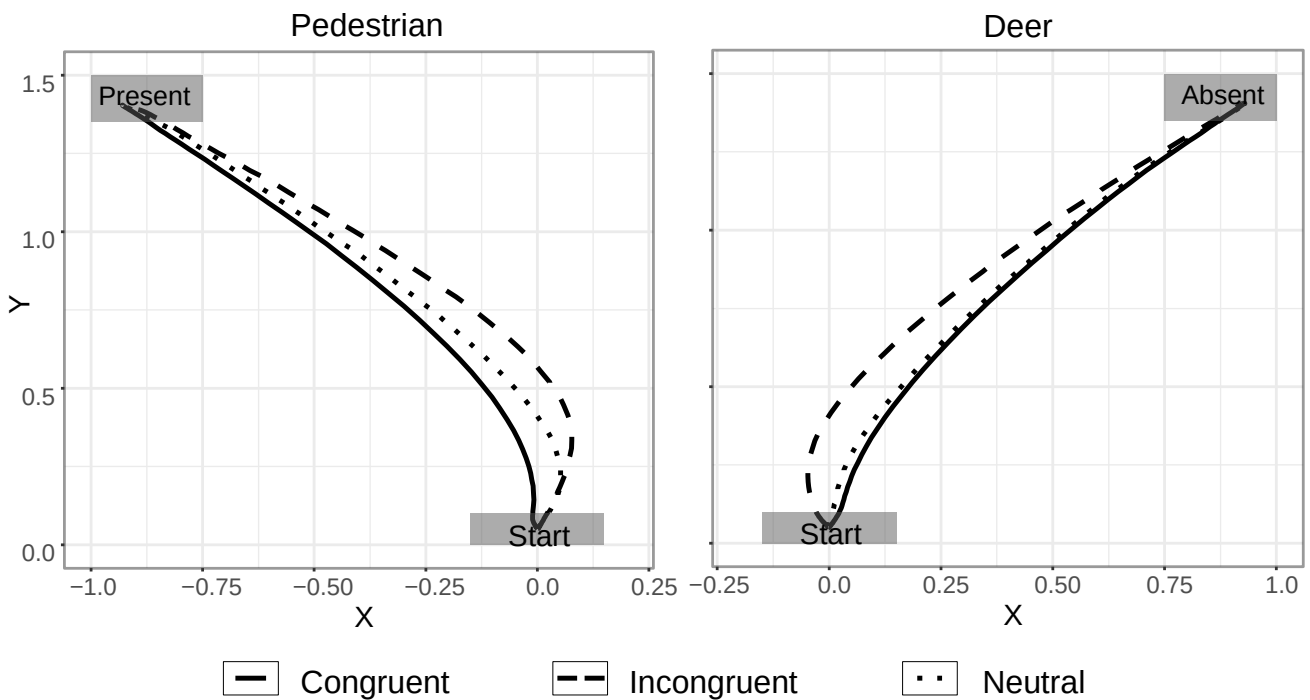


Figure 4: Mean mouse trajectories by object included in the visual scene and by congruency between the road sign and the object.

Figure 5B (top) displays the mediations of the effects of fog and expectations. The effect of the fog level was significantly mediated by detection times. Hence, as previously reported, high fog increased detection times ($b = 27.56$, $t(60.1) = 6.05$, $p < .001$) and longer detection times increased the AUC (when controlling for the effect of fog) ($b = 0.001$, $t(2150.0) = 3.66$, $p < .001$). Interestingly, this aforementioned effect of detection time on AUC indicates that the longer time elapses while searching for the target, the larger the deviation of the mouse movement towards the incorrect *absent* response. This mediation was moderated by congruency C2 as it interacted with both fog on detection time ($b = 17.40$, $t(2069.0) = 2.06$, $p = .04$) and detection time on AUC ($b = 0.001$, $t(1533.0) = 2.23$, $p = .026$). In the absence of expectations, 2.1% of the total effect was reduced when controlling for detection times ($c = 0.19$, $c' = 0.19$), the indirect effect being non significant ($ab = 0.003$, $CI = [-0.02, 0.02]$) whereas in the presence of expectations, the total effect was reduced by 41.4% ($c = 0.11$, $c' = 0.07$) and the indirect effect was significantly different from 0 ($ab = 0.045$, $CI = [0.01, 0.08]$).

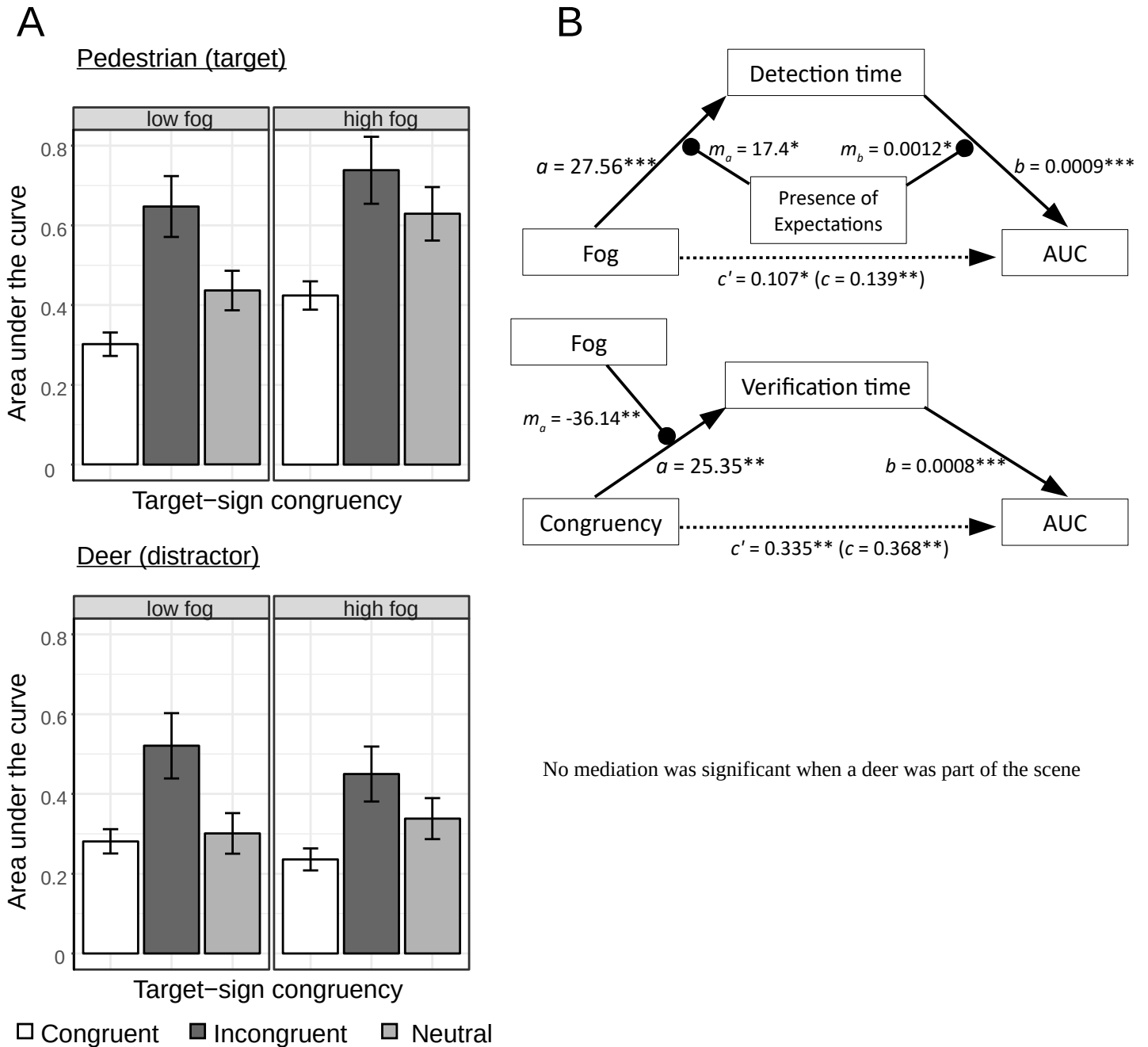


Figure 5: Overview of the results for the Areas Under the Curve (AUC) when: 1st row, the object included in the scene is a pedestrian (target); 2nd row, the object included in the scene is a deer (distractor). **A.** Mean AUC as a function of road sign congruency and fog level. Error bars display the standard errors. **B.** Mediation of the effects of fog and object-road sign congruency (noted Congruency) on AUC by the eye movement measures, respectively detection times and verification times. Congruency corresponds to the congruency contrast C1 and Presence of expectations correspond to the congruency contrast C2. Dashed arrows represent an effect mediated by detection or verification times. Arrows ending with a black dot represent moderation (interaction) effects. Parameters a and b are displayed as well as the direct effects c' and total effect c for the mediated effects. Parameters m_a and m_b correspond to the moderation of, respectively, paths a and b by a moderator. * $p < .05$, ** $p < .01$, *** $p < .001$.

Additionally, the effect of congruency C1 on AUC was mediated by verification time. Incongruent road sign increased the verification times compared to congruent road sign ($b = 25.35$, $t(45.2) = 2.99$, $p = .004$) and verification times increased the AUC ($b = 0.008$, $t(2075) = 4.69$, $p < .001$). Furthermore, this mediation was moderated by fog, the interaction between fog and congruency C1 on verification times being significant ($b = -36.14$, $t(2082.2) = -2.80$, $p = .005$). The interaction between verification time and fog on AUC was not significant ($b = -0.0001$, $t(745.4) = -0.39$). For a low fog level, the total effect of congruency C1 on AUC was reduced by 14.7% when including the verification time in the model ($c = 0.39$, $c' = 0.33$) and the indirect effect was significantly different from 0 ($ab = 0.04$, $CI = [0.01, 0.08]$). For a high fog level, only 2.72% of this total effect was mediated ($c = 0.35$, $c' = 0.34$), the indirect effect being non significant ($ab = 0.006$, $CI = [-0.007, 0.03]$).

Deer in the scene (distractor). Figure 4 (right) displays the mean trajectories by congruency condition and Figure 5A (bottom) the mean AUC by experimental conditions and their standard error. Similarly than with scenes including a pedestrian, for scenes including a deer, an incongruent road sign induced bigger deviations towards the incorrect *present* response than a congruent road sign ($b = 0.24$, $t(50.5) = 2.66$, $p = .011$). However, there was no significant difference between the presence and the absence of expectations ($b = 0.08$, $t(51) = 1.59$, $p = .119$), between low and high fog ($b = -0.02$, $t(73.9) = -0.63$), and between medium and high strength of expectations ($b = 0.03$, $t(43) = 0.377$). None of the interactions were significant (all $|t| < 1.5$, $p > .1$).

While controlling for the effects of the causal variables, we found that spending time fixating the distractor (verification) increased the deviation towards the incorrect, *present* response ($b = 0.001$, $t(1513) = 6.55$, $p < .001$) whereas, as we hypothesized, spending time for the detection of the distractor reduced the deviation of the response trajectories towards the incorrect *present* response ($b = -0.0006$, $t(1900) = -2.02$, $p = .043$). This latter effect interacted with congruency C1 ($b = -0.002$, $t(1818) = -2.61$, $p = .009$). The analysis of the conditional effects pointed out that for congruent road sign (expectations towards *deer/absent* response), the detection time had no effect on the deviation towards the *present* response ($b_{DT,C1=0.5} = 0.0003$; $CI = [-0.0002; 0.0009]$, $df = 1844.9$) whereas for incongruent road sign (expectations towards *pedestrian/present* response) it decreased this deviation ($b_{DT,C1=+0.5} = -0.002$; $CI = [-0.003; -0.0004]$, $df = 41.9$). In other words, when detection times are long, the deviation towards the incorrect *present* response (AUC) is decreased when induced by incongruent road signs compared to congruent road signs. Thus, the influence of the predictive road signs on the mouse trajectory deviations was reduced with the time spent searching for the distractor. Finally, no effects of the causal variables on AUC were mediated when a deer was part of the scene.

Discussion

We designed an experiment to test the hypothesis that the uncertainty of visual information (degradation by fog) and the expectations about the identity of the object included in the visual scene would affect the decision-making process by affecting the underlying active accumulation of evidence operated through eye movements. Consistently with our hypothesis, visual search and decision-making process were impaired when visual information was degraded by fog. Additionally, both were facilitated when the pedestrian (target) presence or deer (distractor) presence were predictable (congruent road sign) compared to when the prediction about the target's presence was incorrect (incongruent road sign). Our study focused more particularly on the integration of visual uncertainty and top-down expectations within the perceptual decision-making process through visual exploration by estimating whether a significant part of the effects on mouse-tracking measures was attributable to perceptual processes, *i.e.* object detection and verification.

Visual uncertainty increased the temporal decision-making measure (RT) and biased the response selection towards an *absent* response when a target was in the scene and these effects were partly explained by impaired object detection. Conversely, object expectations influenced both the temporal decision-making measure and the response selection, these effects being partly explained by object verification. Furthermore, they distinguished the contributions of both sources of information linking each of them to one of these two distinct perceptual processes, and underline that, within the decision-making process, expectation and visual uncertainty may combine at the level of evidence collection. We observed that the effect of fog was only mediated by the detection time when expectations were induced (whether they predicted the correct or incorrect response), and that the increased fog (uncertainty) level mitigated how expecting for a pedestrian or a deer affected decision-making through the verification time. Furthermore, our results highlight how eye movements related to object detection and verification contribute to the decision-making process by collecting evidence towards *present* and *absent* responses, with evidence accumulating for the *pedestrian absent* response as time is spent searching for the object (whether a target or a distractor is present in the scene).

Influence of expectations and visual uncertainty on object detection and verification

In our experiment, high visual uncertainty increased object detection whereas expectations about the object identity influenced object verification. Literature about degraded visual conditions in driving-themed visual scenes, showed for instance the overestimation of distance under fog conditions (Caro, Cavallo, Marendaz, Boer, & Vienne, 2007; Cavallo, Caro, Dore, Colomb, & Dumont, 2007) and impaired

Article 4 (Étude 3) – Recherche visuelle dans le brouillard : attentes et accumulation d'évidence

stimulus detection and word reading under rainfall conditions (Bernardin et al., 2014). To our knowledge, the effects of fog on perceptual decision-making and visual exploration through eye movements have not been investigated yet, and our study constitute one among few research investigating these topics using ecological fog.

However, some studies were interested in whether the availability of visual information impacted the saccades and fixations in real-world scenes. When a lower amount of information is available, saccades tend to have shorter amplitude and less efficient scan-paths towards the target as well as fixations with bigger duration, as observed using gaze-contingent moving windows (Nuthmann, 2013; Pomplun, Reingold, & Shen, 2001), gaze-contingent multiresolutional display (Loschky & McConkie, 2002) and scene degradation using diverse filters and noises (Röhrbein, Goddard, Schneider, James, & Guo, 2015). One study examined more precisely the different epochs of visual search with a gaze contingent moving window paradigm with low peripheral or central resolution (Nuthmann, 2014). It shows increased initiation times (latencies of first saccade on the scene) and scanning times (from first saccade on the scene to first saccade on the target) for peripheral degradation and increased verification times for central degradation. In our experiment, detection times regrouped initiation and scanning times, since only a few fixations were made before looking at the target due to the very simple scene composition.

In accordance with this study, we observed longer detection times for high fog level scenes. However, we did not observe longer verification times with high fog, indicating that fog impaired only the peripheral detection of the object but not its central identification. This could be due to the fact that in the experiment, the same objects were repeated in very similar visual scenes whose variations came mainly from the object's location and the fog configuration. Thus, the objects' visual features were learnt early, allowing for the discrimination, in the central visual field, between target and distractor based on salient details not much affected by the fog (whereas overall, the object is highly degraded).

Contrasting with the effects of visual uncertainty, expectations for either a target or a distractor influenced the objects verification rather than its peripheral detection. Hence, expecting for the correct object decreased the verification times compared with expecting for the incorrect object, therefore facilitating object recognition. Bravo and Farid (2009, 2012) showed that previously learned visual representations of the target improves visual search. Besides, previous results showed that knowledge of the visual features of the target decreases both the detection and verification times in visual search tasks, facilitating target recognition (Castelhano & Heaven, 2010; Castelhano et al., 2008; Malcolm & Henderson, 2009, 2010; Spotorno et al., 2014). In our experiment, as the objects appearances were well known (no variation across trials), the road signs might have activated the visual representation of their linked object. As a result, for congruent objects, it might have reduced the necessity to integrate more visual information,

Article 4 (Étude 3) – Recherche visuelle dans le brouillard : attentes et accumulation d'évidence

and thus reduced the time needed to spend looking at the object. Conversely, a road sign incongruent with the object might have a cost for pointing towards the incorrect representation and might necessitate the integration of more information by looking at the object longer.

This effect was affected by the fog conditions when the object was a pedestrian. In these conditions, the incongruent road signs increased verification time (compared with congruent) only for a low fog level, indicating that when the visual uncertainty was high, expectations for an object had reduced effects. This can be explained by the fact that, as time to respond is limited and high fog induces longer detection time, the time spent fixating the target is reduced. This would indicate that accumulating more visual evidence for correcting wrong expectations is not mandatory and could reflect a strategic resource allocation in order to optimize a trade-off between reaching a maximal level of certainty (and accuracy) while limiting the time spent (Heitz, 2014; Heitz & Schall, 2012; Lewis, Shvartsman, & Singh, 2013).

Influence of perceptual effects of expectations on decision-making

The expectations about the object's identity had an influence on global response times and biased the response selection towards *absent* or *present* responses, when a pedestrian target was in the scene. When a deer distractor was in the scene, it only biased the response selection. Mediation analysis allowed us to explain part of these effects by the influence exerted by the expectations on object verification. Expectations inducing anticipation for the visual features of the target (which can be taken as evidence towards the *present* response) deviated the response movement towards this response (lower AUC) and resulted in faster decision compared with expectations inducing anticipation for the distractor. Conversely, incorrect expectations generated predictions for the distractor, impairing target's recognition and therefore, deviated the mouse trajectories in the wrong direction and had a temporal cost. This indicates that part of the influence of expectations on response selection and RT is due to facilitated or impaired target identification which is consistent with a previous study in which visual scenes were build using 3D modelling (Quétard et al., 2015).

This mediation was moderated by fog, as expectations affected verification times only under low fog conditions. Accordingly, verification times significantly mediated the effect of expectations in this condition. As expectations did not affect the time spent verifying the target through eye fixation under high fog, the increased response times and deviations for incorrect expectations (expecting for a distractor) were only due to response selection biases whereas under low fog, part of these effects (25.2% for RT, 14.7% for AUC), were due to the influence of expectations on the collection of visual evidence on the target. The relatively small proportion of the effect mediated through object verification is not surprising since the design of our task allowed for the road sign to exert a motor anticipation of the most probable response, leading the

Article 4 (Étude 3) – Recherche visuelle dans le brouillard : attentes et accumulation d'évidence

mouse movement to be deviated accordingly (Bruhn, 2013; Bruhn, Huette, & Spivey, 2014). Thus, this motor response bias is probably the main mechanism explaining the influence of object expectations on mouse trajectory deviations. However, the mediation analyses permitted to confirm that a significant part of this effect could be explained by perceptual processes, more particularly object verification, rather than by a response bias.

However, this mediation was not found when a distractor (deer) was included in the scene since incongruent expectations did not significantly increased verification times in our mediation analyses. An explanation could be that roe deers were easier to identify than pedestrians but our data did not support it, as there was no significant difference between deer and pedestrian on verification times. Alternatively, a lack of statistical power could also explain this inconsistency. Participants had to press the *absent* response, which was also the correct response when no object was in the scene. This imbalance between *absent* and *present* responses could have induced a general bias of the trajectories towards the *absent* response (independent from visual exploration), leading to an overall reduction of the size of the effects when the deer is present in the scene and therefore reducing the statistical power. Such bias can be verified in our data, as indicated by the deviation of the trajectories towards the incorrect response (AUC) being larger when a pedestrian was in the scene than when a deer was in the scene ($b = 0.18$, $t(47) = 4.53$, $p < .001$, as tested in a LMM including the object identity, fog and congruency as fixed and random parameters along with all fixed interactions).

Visual uncertainty and accumulation of evidence upon response selection

Contrasting with the influence of expectations being mediated by object verification, visual uncertainty exerted its influence on RT through object detection-related eye movements. The fog delayed the overall response process –thus inducing longer RT– by delaying the moment when the object is detected. This was the case both when a pedestrian target or a deer distractor was present in the scene. There was also an effect of visual uncertainty on AUC, but only when the object included in the scene was a pedestrian target. In this case, a high fog level induced a larger mouse deviation towards the incorrect *absent* response and this effect was partly mediated through target detection. As the target is harder to find due to fog, it is detected later and spending time searching for it could accumulate evidence towards the *target absent* alternative. We previously found this kind of results in a task where the participants had to search for object targets in real-world indoor scenes, while it could be placed in usual or unusual locations, inducing expectations about the spatial location rather than the object identity (Quétard et al., 2016).

The advantage of the present study is that it can account for this result in a more precise and complete way due to the inclusion of either a target or a distractor in the scene. Hence this explanation

imply divergent hypotheses about the effect of detection times on the mouse trajectory deviation depending on the object included in the scene (target or distractor). First, if a pedestrian target is included, higher detection times should induce the mouse trajectories to deviate towards *target absent* response, thus increasing the AUC and this is the case in our results. Second, if a deer distractor is included, higher detection times should still induce a mouse trajectory deviation towards the *target absent* response but then, this would result in decreasing the AUC since in this case, the *target absent* response is the correct response. We also found such effect.

This also explains why no effect of fog could be found on AUC when a distractor was presented. The role of visual uncertainty in this scheme is central since in both cases, high fog level reduces the visibility of the pedestrian and the deer, thus increasing their detection times and consequently biasing the decision towards the absence of the target. Hence, decision towards the absence of the target could be achieved through a mechanism gradually accumulating evidence based on the time elapsed while the target is not found.

Notably, this mechanism would allow for the accumulation of evidence towards a *no-object* response. As long as the scene is visually explored with no evidence collected in favour of the alternative *target present* response, evidence for the hypothesis that there is no object is accumulated. This is done until an object is detected or else, a sufficient level of certainty for the absence of any target object is reached and the visual search is interrupted. As mentioned earlier, Dufau et al. (2012) proposed a model based on the evidence accumulation scheme, implementing such decision rule for determining whether a letter sequence is a non-word (which would be an equivalent of an *absent* response in psycholinguistics). This could be a plausible way to implement a search termination mechanism (see for a review Wolfe, 2012) as exploring visual scenes in real-world situations often require to decide pursuing the search for specific objects at a particular location or interrupting it and switching location.

Furthermore, this mechanism interacted with the integration of prior object expectations since the presence of such expectations (children or animal predictive road signs) increased the share of the effect of fog (on RT and AUC) which was mediated by the detection time of target pedestrians. Expecting for any object induced a bigger effect of fog on the detection process, and thus a larger bias of decision-making towards the incorrect *absent* response (AUC) with corresponding increased RT compensating this bias. This indicates a higher reliance on target search when inducing expectations.

This moderation was not found for deer distractor since for this object, detection times had the opposite effect on AUC. However, when this object was included in the scene, higher detection times induced a bigger reduction of the deviation towards the *present* response (AUC) for incongruent expectations compared to congruent (no effect of detection time for congruent expectations), and they

Article 4 (Étude 3) – Recherche visuelle dans le brouillard : attentes et accumulation d'évidence

increased RT in a lesser extent. For congruent expectations, as responses are already biased towards the *absent* response, detection time can only decrease the mouse trajectories in by a small (or null) margin, whereas incongruent expectations bias the trajectories towards the *present* response and thus, increase this margin. As a consequence, with incongruent expectations, as detection times allowed to select the correct response quicker, their effect on the decision time was lower. This highlights how competing evidence sources, one built on prior expectations and the other built on failing to find the target (facilitated by visual uncertainty), influence the response selection in a dynamic manner, until the object is recognized, leading to the convergence towards the correct response.

Ultimately, these results enlighten mechanisms of evidence accumulation through eye movements in visual search and its interactions with prior expectations, which could be implemented in models including both eye movements generation mechanisms and evidence accumulation processes such as, for instance, Purcell et al. (Purcell et al., 2010; Purcell, Schall, Logan, & Palmeri, 2012; Schall, Purcell, Heitz, Logan, & Palmeri, 2011) or Lewis et al. (2013) approaches. However, further experimental or computational studies should address the whether this mechanism is caused by the deployment of an active visual search behaviour being unsuccessful as no target is found –for instance implementing costs when engaged in an unsuccessful behaviour (Drugowitsch, Moreno-Bote, Churchland, Shadlen, & Pouget, 2012)– or by time passively elapsing while no significant evidence for the target is accumulated.

Conclusion

This study examined the effects of visual uncertainty and prior object expectations on eye movements related to detection and identification of the object, and how these processes contribute to the decision-making leading to determine whether the target is *present* or *absent*. Based on this experiment, visual uncertainty and prior object expectations both exert an influence on decision-making through eye movements related to collection of evidence. However, their contribution may rest on distinct perceptual processes. Not only expectations bias the response selection through motor anticipation, but they also influence the verification of an object's identity by activating congruent (or incongruent) object representations, which in turn bias the response selection and speed up (or slow down) the decision process. Visual uncertainty, manipulated with fog, would reduce the object's visibility and increase the time to detect it, hence impairing decision-making in two ways: By delaying the object's recognition (and thus the final decision), and by gradually biasing the response selection towards the *target absent* response while the object is not yet detected. This is of particular importance when driving, in which visual search is often made under time pressure for instance, when checking if a pedestrian approaches a crossroad. In these conditions, if target detection takes a significant proportion of the available time, the resulting

accumulation of evidence towards the absence of the target (the pedestrian) can lead to not noticing it.

It is important to note that our analyses only account for a category of models where visual uncertainty and expectations influence the perceptive processes dedicated to collect evidence, which in turn influence decision and response selection in a feed-forward manner. However, commitment to a choice, which can be reflected through mouse trajectories, can also have an influence on eye movement, for instance by defining the location of the information able to confirm this decision (Friston et al., 2012; Kietzmann et al., 2011; Kietzmann & König, 2015). This points towards a category of models where collection of evidence processes and commitment for a response interact in a dynamic manner. In order to investigate this question, using mouse-tracking techniques as a continuous measure of the decision-making process, together with eye-tracking measurements can provide a fruitful experimental framework. This can be combined with a modelling approach, for instance based on dynamic decision-making such as sequential sampling models (Friedman, Brown, & Finkbeiner, 2013; Lepora & Pezzulo, 2015; Ratcliff & McKoon, 2008) or attractor dynamics (Quinton et al., 2014; Volpi et al., 2014) as they are convenient to model mouse deviation. How eye movements are integrated in the architecture of these models and how they influence the various decision parameters of these models (*e.g.*, drift rates, decision criteria) constitute an exciting challenge for future studies in the field of visual search in real world scenes. This presents particularly interesting perspectives towards computational modelling in the field of driving, as degraded vision and expecting risks are central issues and affect both visual exploration and decision-making in driving environment.

Acknowledgements and funding:

The authors want to acknowledge the CEREMA engineers Philippe Morange and Jean-Luc Bicard, for their help during the tests at the “Fog platform” in Clermont-Ferrand.

The authors gratefully thank Mr. and Mrs. Mialon for lending stuffed animals (including the roe deer) for the visual scenes' construction.

This study was approved by the french Comité de Protection des Personnes “Sud-Est VI” (IRB00008526, CE65, 2014).

This work was funded by grants from the French program « investissement d'avenir » managed by the National Research Agency (ANR), from the European Union (Auvergne European Regional Development Funds -ERDF- of Auvergne region) and from the « Région Auvergne » in the framework of the IMobS3 LabEx (ANR-10-LABX-16-01).

Bibliography

- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. <http://doi.org/10.1016/j.jml.2007.12.005>
- Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience*, *5*(8), 617–629. <http://doi.org/10.1038/nrn1476>
- Bar, M. (2009). The proactive brain: memory for predictions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1235–1243. <http://doi.org/10.1098/rstb.2008.0310>
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. <http://doi.org/10.1016/j.jml.2012.11.001>
- Bernardin, F., Bremond, R., Ledoux, V., Pinto, M., Lemonnier, S., Cavallo, V., & Colomb, M. (2014). Measuring the effect of the rainfall on the windshield in terms of visual performance. *Accident Analysis & Prevention*, *63*, 83–88. <http://doi.org/10.1016/j.aap.2013.10.008>
- Blignaut, P. (2009). Fixation identification: The optimum threshold for a dispersion algorithm. *Attention, Perception, & Psychophysics*, *71*(4), 881–895. <http://doi.org/10.3758/APP.71.4.881>
- Borji, A., Tavakoli, H. R., Sihite, D. N., & Itti, L. (2013). Analysis of Scores, Datasets, and Models in Visual Saliency Prediction. In *2013 IEEE International Conference on Computer Vision* (pp. 921–928). IEEE. <http://doi.org/10.1109/ICCV.2013.118>
- Bravo, M. J., & Farid, H. (2009). The specificity of the search template. *Journal of Vision*, *9*(1):34, 1–9. <http://doi.org/10.1167/9.1.34>
- Bravo, M. J., & Farid, H. (2012). Task demands determine the specificity of the search template, *Attention, Perception, & Psychophysics*, *74*(1), 124–131. <http://doi.org/10.3758/s13414-011-0224-5>
- Bruhn, P. (2013). Emergence of spontaneous anticipatory hand movements in a probabilistic environment. *Advances in Cognitive Psychology*, *9*(2), 62–73. <http://doi.org/10.2478/v10053-008-0132-y>
- Bruhn, P., Huette, S., & Spivey, M. (2014). Degree of certainty modulates anticipatory processes in real time. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(2), 525–538. <http://doi.org/10.1037/a0034365>
- Caro, S., Cavallo, V., Marendaz, C., Boer, E., & Vienne, F. (2007). The influence of fog on motion discrimination thresholds in car following. In *Proceedings of the 4th International Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design* (pp. 446–451).
- Castelhano, M. S., & Heaven, C. (2010). The relative contribution of scene context and target features to visual search in scenes. *Attention, Perception & Psychophysics*, *72*(5), 1283–1297. <http://doi.org/10.3758/APP.72.5.1283>

Article 4 (Étude 3) – Recherche visuelle dans le brouillard : attentes et accumulation d'évidence

- Castelhano, M. S., Pollatsek, A., & Cave, K. R. (2008). Typicality aids search for an unspecified target, but only in identification and not in attentional guidance. *Psychonomic Bulletin & Review*, *15*(4), 795–801. <http://doi.org/10.3758/PBR.15.4.795>
- Cavallo, V., Caro, S., Dore, J., Colomb, M., & Dumont, E. (2007). Risky driving in fog: Psychological explanations. In *Proceedings of the Road Safety on Four Continents Conference* (Vol. 14, pp. 12–12).
- Colomb, M., Hirech, K., André, P., Boreux, J. J., Lacôte, P., & Dufour, J. (2008). An innovative artificial fog production device improved in the European project “FOG.” *Atmospheric Research*, *87*(3–4), 242–251. <http://doi.org/10.1016/j.atmosres.2007.11.021>
- Dale, R., Kehoe, C., & Spivey, M. J. (2007). Graded motor responses in the time course of categorizing atypical exemplars. *Memory & Cognition*, *35*(1), 15–28. <http://doi.org/10.3758/BF03195938>
- Donnarumma, F., Costantini, M., Ambrosini, E., Friston, K. J., & Pezzulo, G. (2017). Action perception as hypothesis testing. *Cortex*, *89*, 45–60. <http://doi.org/10.1016/j.cortex.2017.01.016>
- Drugowitsch, J., Moreno-Bote, R., Churchland, A. K., Shadlen, M. N., & Pouget, A. (2012). The Cost of Accumulating Evidence in Perceptual Decision Making. *Journal of Neuroscience*, *32*(11), 3612–3628. <http://doi.org/10.1523/JNEUROSCI.4010-11.2012>
- Dufau, S., Grainger, J., & Ziegler, J. C. (2012). How to say “no” to a nonword: A leaky competing accumulator model of lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(4), 1117–1128. <http://doi.org/10.1037/a0026948>
- Ehinger, K. A., Hidalgo-Sotelo, B., Torralba, A., & Oliva, A. (2009). Modelling search for people in 900 scenes: A combined source model of eye guidance. *Visual Cognition*, *17*(6–7), 945–978. <http://doi.org/10.1080/13506280902834720>
- Fairchild, A. J., & MacKinnon, D. P. (2009). A General Model for Testing Mediation and Moderation Effects. *Prevention Science*, *10*(2), 87–99. <http://doi.org/10.1007/s11121-008-0109-6>
- Freeman, J. B., & Ambady, N. (2009). Motions of the hand expose the partial and parallel activation of stereotypes. *Psychological Science*, *20*(10), 1183–1188. <http://doi.org/10.1111/j.1467-9280.2009.02422.x>
- Freeman, J. B., & Ambady, N. (2010). MouseTracker: Software for studying real-time mental processing using a computer mouse-tracking method. *Behavior Research Methods*, *42*(1), 226–241. <http://doi.org/10.3758/BRM.42.1.226>
- Freeman, J. B., & Ambady, N. (2011). Hand movements reveal the time-course of shape and pigmentation processing in face categorization. *Psychonomic Bulletin & Review*, *18*(4), 705–712. <http://doi.org/10.3758/s13423-011-0097-6>
- Freeman, J. B., Ambady, N., Rule, N. O., & Johnson, K. L. (2008). Will a category cue attract you? Motor output reveals dynamic competition across person construal. *Journal of Experimental Psychology: General*, *137*(4), 673–690. <http://doi.org/10.1037/a0013875>
- Freeman, J. B., Dale, R., & Farmer, T. A. (2011). Hand in Motion Reveals Mind in Motion. *Frontiers in*

Article 4 (Étude 3) – Recherche visuelle dans le brouillard : attentes et accumulation d'évidence

Psychology, 2:59. <http://doi.org/10.3389/fpsyg.2011.00059>

Freeman, J. B., Penner, A. M., Saperstein, A., Scheutz, M., & Ambady, N. (2011). Looking the part: social status cues shape race perception. *PLoS One*, 6(9), e25107. <http://doi.org/10.1371/journal.pone.0025107>

Friedman, J., Brown, S., & Finkbeiner, M. (2013). Linking cognitive and reaching trajectories via intermittent movement control. *Journal of Mathematical Psychology*, 57(3–4), 140–151. <http://doi.org/10.1016/j.jmp.2013.06.005>

Friston, K., Adams, R. A., Perrinet, L., & Breakspear, M. (2012). Perceptions as Hypotheses: Saccades as Experiments. *Frontiers in Psychology*, 3(MAY), 151. <http://doi.org/10.3389/fpsyg.2012.00151>

Gold, J. I., & Shadlen, M. N. (2007). The Neural Basis of Decision Making. *Annual Review of Neuroscience*, 30(1), 535–574. <http://doi.org/10.1146/annurev.neuro.29.051605.113038>

Hayes, A. F. (2009). Beyond Baron and Kenny: Statistical mediation analysis in the new millennium. *Communication Monographs*, 76(4), 408–420. <http://doi.org/10.1080/03637750903310360>

Hegd , J. (2008). Time course of visual perception: coarse-to-fine processing and beyond. *Progress in Neurobiology*, 84(4), 405–439. <http://doi.org/10.1016/j.pneurobio.2007.09.001>

Heitz, R. P. (2014). The speed-accuracy tradeoff: history, physiology, methodology, and behavior. *Frontiers in Neuroscience*, 8:150. <http://doi.org/10.3389/fnins.2014.00150>

Heitz, R. P., & Schall, J. D. (2012). Neural Mechanisms of Speed-Accuracy Tradeoff. *Neuron*, 76(3), 616–628. <http://doi.org/10.1016/j.neuron.2012.08.030>

Hout, M. C., & Goldinger, S. D. (2015). Target templates: the precision of mental representations affects attentional guidance and decision-making in visual search. *Attention, Perception, & Psychophysics*, 77(1), 128–149. <http://doi.org/10.3758/s13414-014-0764-6>

Itti, L., & Baldi, P. (2009). Bayesian surprise attracts human attention. *Vision Research*, 49(10), 1295–1306. <http://doi.org/10.1016/j.visres.2008.09.007>

Johnson, K. L., Freeman, J. B., & Pauker, K. (2012). Race is gendered: how covarying phenotypes and stereotypes bias sex categorization. *Journal of Personality and Social Psychology*, 102(1), 116–131. <http://doi.org/10.1037/a0025335>

Judd, C. M., Yzerbyt, V. Y., & Muller, D. (2014). Mediation and Moderation. In H. Reis & C. M. Judd (Eds.), *Handbook of research methods in social psychology and personality psychology* (2nd ed., pp. 653–676). Cambridge: Cambridge University Press.

Kanan, C., Tong, M. H., Zhang, L., & Cottrell, G. W. (2009). SUN: Top-down saliency using natural statistics. *Visual Cognition*, 17(6–7), 979–1003. <http://doi.org/10.1080/13506280902771138>

Kietzmann, T. C., Geuter, S., & K nig, P. (2011). Overt visual attention as a causal factor of perceptual awareness. *PLoS One*, 6(7), e22614. <http://doi.org/10.1371/journal.pone.0022614>

Kietzmann, T. C., & K nig, P. (2015). Effects of contextual information and stimulus ambiguity on overt visual

Article 4 (Étude 3) – Recherche visuelle dans le brouillard : attentes et accumulation d'évidence

- sampling behavior. *Vision Research*, 110, 76–86. <http://doi.org/10.1016/j.visres.2015.02.023>
- Land, M. F., & Hayhoe, M. (2001). In what ways do eye movements contribute to everyday activities? *Vision Research*, 41(25–26), 3559–3565. [http://doi.org/10.1016/S0042-6989\(01\)00102-X](http://doi.org/10.1016/S0042-6989(01)00102-X)
- Lepora, N. F., & Pezzulo, G. (2015). Embodied choice: How action influences perceptual decision making. *PLOS Computational Biology*, 11(4), e1004110. <http://doi.org/10.1371/journal.pcbi.1004110>
- Lewis, R. L., Shvartsman, M., & Singh, S. (2013). The Adaptive Nature of Eye Movements in Linguistic Tasks: How Payoff and Architecture Shape Speed-Accuracy Trade-Offs. *Topics in Cognitive Science*, 5(3), 581–610. <http://doi.org/10.1111/tops.12032>
- Loschky, L. C., & McConkie, G. W. (2002). Investigating spatial vision and dynamic attentional selection using a gaze-contingent multiresolutional display. *Journal of Experimental Psychology: Applied*, 8(2), 99–117. <http://doi.org/10.1037//1076-898X.8.2.99>
- Malcolm, G. L., & Henderson, J. M. (2009). The effects of target template specificity on visual search in real-world scenes: evidence from eye movements. *Journal of Vision*, 9(11):8, 1–13. <http://doi.org/10.1167/9.11.8>
- Malcolm, G. L., & Henderson, J. M. (2010). Combining top-down processes to guide eye movements during real-world scene search. *Journal of Vision*, 10(2):4, 1–11. <http://doi.org/10.1167/10.2.4>
- Malcolm, G. L., & Shomstein, S. (2015). Object-Based Attention in Real-World Scenes. *Journal of Experimental Psychology: General*, 144(2), 257–263.
- Muller, D., Judd, C. M., & Yzerbyt, V. Y. (2005). When moderation is mediated and mediation is moderated. *Journal of Personality and Social Psychology*, 89(6), 852–863. <http://doi.org/10.1037/0022-3514.89.6.852>
- Nuthmann, A. (2013). On the visual span during object search in real-world scenes. *Visual Cognition*, 21(7), 803–837. <http://doi.org/10.1080/13506285.2013.832449>
- Nuthmann, A. (2014). How do the regions of the visual field contribute to object search in real-world scenes? Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance*, 40(1), 342–360. <http://doi.org/10.1037/a0033854>
- Pomplun, M., Reingold, E. M., & Shen, J. (2001). Peripheral and parafoveal cueing and masking effects on saccadic selectivity in a gaze-contingent window paradigm. *Vision Research*, 41(21), 2757–2769. [http://doi.org/10.1016/S0042-6989\(01\)00145-6](http://doi.org/10.1016/S0042-6989(01)00145-6)
- Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, 40(3), 879–891. <http://doi.org/10.3758/BRM.40.3.879>
- Purcell, B. A., Heitz, R. P., Cohen, J. Y., Schall, J. D., Logan, G. D., & Palmeri, T. J. (2010). Neurally constrained modeling of perceptual decision making. *Psychological Review*, 117(4), 1113–1143. <http://doi.org/10.1037/a0020311>

Article 4 (Étude 3) – Recherche visuelle dans le brouillard : attentes et accumulation d'évidence

- Purcell, B. A., Schall, J. D., Logan, G. D., & Palmeri, T. J. (2012). From Saliency to Saccades: Multiple-Alternative Gated Stochastic Accumulator Model of Visual Search. *Journal of Neuroscience*, *32*(10), 3433–3446. <http://doi.org/10.1523/JNEUROSCI.4622-11.2012>
- Quétard, B., Gana, R., Birem, M., Quinton, J.-C., Colomb, M., Bernardin, F., ... Pezzulo, G. (2014, May). *Saliency in the fog: Application to real & 2,5D pictures*. Poster session presented at the Transport Research Arena Conference, Paris, France.
- Quétard, B., Quinton, J.-C., Colomb, M., Pezzulo, G., Barca, L., Izaute, M., ... Mermillod, M. (2015). Combined effects of expectations and visual uncertainty upon detection and identification of a target in the fog. *Cognitive Processing*, *16*(S1), 343–348. <http://doi.org/10.1007/s10339-015-0673-1>
- Quétard, B., Quinton, J. C., Mermillod, M., Barca, L., Pezzulo, G., Colomb, M., & Izaute, M. (2016). Differential effects of visual uncertainty and contextual guidance on perceptual decisions: Evidence from eye and mouse tracking in visual search. *Journal of Vision*, *16*(11):28, 1:21. <http://doi.org/10.1167/16.11.28>
- Quinton, J.-C., Volpi, N. C., Barca, L., & Pezzulo, G. (2014). The Cat is on the Mat. or is it a Dog? Dynamic Competition in Perceptual Decision Making. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, *44*(5), 539–551. <http://doi.org/10.1109/TSMC.2013.2279664>
- Ratcliff, R., & McKoon, G. (2008). The Diffusion Decision Model: Theory and Data for Two-Choice Decision Tasks. *Neural Computation*, *20*(4), 873–922. <http://doi.org/10.1162/neco.2008.12-06-420>
- Renninger, L. W., Verghese, P., & Coughlan, J. (2007). Where to look next? Eye movements reduce local uncertainty. *Journal of Vision*, *7*(3):6, 1–17. <http://doi.org/10.1167/7.3.6>
- Röhrbein, F., Goddard, P., Schneider, M., James, G., & Guo, K. (2015). How does image noise affect actual and predicted human gaze allocation in assessing image quality? *Vision Research*, *112*, 11–25. <http://doi.org/10.1016/j.visres.2015.03.029>
- Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the symposium on Eye tracking research & applications - ETRA '00* (pp. 71–78). New York, USA: ACM Press. <http://doi.org/10.1145/355017.355028>
- Schall, J. D., Purcell, B. A., Heitz, R. P., Logan, G. D., & Palmeri, T. J. (2011). Neural mechanisms of saccade target selection: gated accumulator model of the visual-motor cascade. *European Journal of Neuroscience*, *33*(11), 1991–2002. <http://doi.org/10.1111/j.1460-9568.2011.07715.x>
- Song, J.-H., & Nakayama, K. (2009). Hidden cognitive states revealed in choice reaching tasks. *Trends in Cognitive Sciences*, *13*(8), 360–366. <http://doi.org/10.1016/j.tics.2009.04.009>
- Spivey, M. J., Grosjean, M., & Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences*, *102*(29), 10393–10398. <http://doi.org/www.pnas.org/cgi/doi/10.1073/pnas.0503903102>
- Spotorno, S., Malcolm, G. L., & Tatler, B. W. (2014). How context information and target information guide the eyes from the first epoch of search in real-world scenes. *Journal of Vision*, *14*(2):7, 1–21.

Article 4 (Étude 3) – Recherche visuelle dans le brouillard : attentes et accumulation d'évidence

<http://doi.org/10.1167/14.2.7>

Sprague, N., & Ballard, D. (2003). Eye Movements for Reward Maximization. In S. Thrun, L. Saul, & B. Schölkopf (Eds.), *Advances in Neural Information Processing Systems* (Vol. 16). MIT Press.

Tatler, B. W., Hayhoe, M. M., Land, M. F., & Ballard, D. H. (2011). Eye guidance in natural vision: Reinterpreting salience. *Journal of Vision*, *11*(5):5, 1–23. <http://doi.org/10.1167/11.5.5>

Torralla, A., Oliva, A., Castelhana, 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*(4), 766–786. <http://doi.org/10.1037/0033-295X.113.4.766>

Volpi, N. C., Quinton, J.-C., & Pezzulo, G. (2014). How active perception and attractor dynamics shape perceptual categorization: A computational model. *Neural Networks*, *60*, 1–16. <http://doi.org/10.1016/j.neunet.2014.06.008>

Wolfe, J. M. (2012). When do I quit? The search termination problem in visual search. In *Nebraska Symposium on Motivation* (Vol. 59, pp. 183–208). New York: Springer. http://doi.org/10.1007/978-1-4614-4794-8_8

Wu, C.-C., Wick, F. A., & Pomplun, M. (2014). Guidance of visual attention by semantic information in real-world scenes. *Frontiers in Psychology*, *5*:54. <http://doi.org/10.3389/fpsyg.2014.00054>

3.1. Résumé des résultats de l'Étude 3

À la suite de l'Étude 2, l'Étude 3 permet de confirmer le rôle de l'exploration visuelle dans le processus de prise de décision à la fois lorsque la tâche consiste à identifier la cible (Expérience 1) et à indiquer si une cible est présente ou absente (Expérience 2). Comme dans l'Étude 2, ce rôle est mis en avant via des analyses de médiation et modération, où les effets des variables causales (niveau de brouillard et attentes) sur les temps de détection et de vérification expliquent les effets de ces variables causales sur la prise de décision (durée de la décision et déviation de la souris). Ainsi, les résultats de ces expériences mettent en avant que la détection de la cible, dégradée par le niveau de brouillard, affecte la durée de la décision et, pour l'Expérience 2, la sélection de la réponse (déviation de la souris). La vérification de la cible, facilitée par les attentes sur l'objet présenté, affecte aussi ces deux variables. Par contre, lorsque l'objet présenté est un distracteur et que la réponse à donner est *cible absente*, la vérification de la cible ne médiatise pas l'effet des attentes sur les variables décisionnelles. De plus, dans l'Expérience 1, le niveau de brouillard dégrade la vérification de la cible, dégradant consécutivement durée de la décision et la sélection de réponse. Ce n'est pas le cas dans l'Expérience 2.

Ces résultats éclairent la manière dont l'évidence est accumulée lors de la recherche visuelle, et ce sur deux aspects. D'une part, ils mettent en avant le rôle central de la vérification de la cible (son processus d'identification) dans l'accumulation d'évidence en faveur de la réponse *cible présente* (lorsqu'un piéton est présenté). D'autre part, confirmant les résultats obtenus dans l'Étude 2, ils montrent que le temps passé à regarder la scène sans détecter la cible accumule de l'évidence en faveur d'une réponse *cible absente*. Ainsi, conformément à cette hypothèse formulée à la suite de l'Étude 2, lorsque l'objet présent dans la scène était un piéton (cible), des temps de détection élevés déviaient la trajectoire de la souris vers la réponse incorrecte *cible absente*. En revanche, lorsque l'objet présenté était un chevreuil (distracteur), des temps de détection élevés réduisaient la déviation vers la réponse incorrecte *cible présente* indiquant ainsi un biais vers la réponse *cible absente*. De manière générale, cela indique que le temps écoulé sans détecter l'objet biaise la prise de décision vers la réponse *cible absente*.

Chapitre 4

Discussion générale

Comme mentionné en introduction, ce travail de thèse s'inscrit dans le cadre du laboratoire d'excellence IMobS³, en collaboration avec le laboratoire de Clermont Ferrand du CEREMA, et porte ainsi des enjeux applicatifs en matière de perception en situation routière, en particulier en conditions de vision dégradée. Ces enjeux applicatifs sont directement mis en avant dans l'Étude 3, reproduisant une scène visuelle routière dans une situation de vision dégradée écologique, via la production d'un brouillard réaliste, reproduisant les caractéristiques physiques du brouillard naturel. Au delà des enjeux théoriques explorés dans cette étude, elle a ainsi permis d'élaborer une base standardisée de scènes visuelles plongées dans le brouillard (voir Annexe 2 pour des exemples) ainsi qu'une méthode et un programme permettant de manipuler les éléments composant ces scènes pour construire de nouvelles scènes avec des niveaux de brouillard réalistes (images synthétiques 2.5D, voir Annexe 1 : poster Quétard et al., 2014). À noter que cet outil a surtout été utilisé pour faire varier le panneau des scènes visuelles, le brouillard dans le reste de la scène et en particulier au niveau de la cible/distracteur étant laissés tels que photographiés (non modifiés via ce programme). S'il est évident que l'usage d'images statiques ainsi que d'une méthode expérimentale spécifique (tâche de *mouse-tracking*) éloigne grandement nos études de situations écologiques et empêche ainsi de mettre directement en parallèle nos résultats avec, par exemple, des situations de conduite réelles, les études présentées permettent de répondre à des enjeux applicatifs dans ce domaine, en particulier via les réponses qu'elles apportent au niveau fondamental.

Ainsi, ce travail s'inscrit dans un cadre théorique tentant de mettre en correspondance études expérimentales sur l'allocation du regard lors de l'exploration visuelle de scènes naturelles et approche computationnelle de la prise de décision perceptuelle. Par son inscription dans ce cadre théorique, les trois études présentées permettent d'apporter des éléments répondants à différentes problématiques applicatives. Ainsi, elles permettent d'adresser la question de la vision active en situation de vision dégradée dont un domaine applicatif, comme mentionné précédemment, concerne la conduite en situation de brouillard ou de pluie. Une seconde problématique applicative concerne l'intégration d'attentes pour l'allocation du regard. Ces attentes portent sur la localisation dans l'espace (Étude 2) et sur l'identité d'objets ou d'obstacles (Étude 3), toutes deux très importantes lorsque l'on aborde l'environnement routier. Dans cet environnement, nombre de stimuli visuels sont placés à des endroits standards, souvent normés (panneaux, lignes routières), et visent à donner des informations pertinentes pour la tâche de conduite, notamment générer des attentes sur les risques que peuvent rencontrer les conducteurs.

Chapitre 4 – Discussion générale

Enfin, nous avons mis en relation ces aspects perceptifs avec les mécanismes de prise de décision perceptuelle, permettant d'étudier ceux-ci dans une perspective active, dirigée vers un but. Cette orientation permet un rapprochement supplémentaire avec les situations routières dans la mesure où la perception, dans ce contexte, est portée sur l'action et souvent, la prise d'une décision (*e.g.*, détecter un obstacle et l'éviter vs. freiner, identifier un stimulus pour anticiper son mouvement et réagir en conséquence). La prise de décision perceptuelle offre ainsi un cadre théorique particulièrement intéressant pour la conduite automobile. De plus, le fait que ce cadre théorique repose en grande partie sur une approche computationnelle crée un pont naturel entre vision biologique et vision artificielle, cette dernière présentant des perspectives majeures pour la conduite automobile (véhicules et systèmes d'assistance intelligents).

Ce travail se positionne donc sur un versant expérimental et vise donc à recueillir des données empiriques mettant en parallèle perception visuelle et prise de décision de façon à mettre en avant des aspects actifs de la vision, et ainsi suggérer des architectures et mécanismes computationnels intégrant ces deux éléments. Nous avons donc employé une méthodologie originale s'appuyant sur deux tâches classiques de prise de décision : les tâches de catégorisation et de recherche visuelle. Notre méthodologie générale consiste à mesurer ou inférer de concert la prise de décision en s'appuyant sur des tâches de *mouse-tracking* et la perception visuelle, en particulier la collecte d'évidence sensorielle (vision active) via l'estimation du compromis vitesse-exactitude (CVE, Étude 1, catégorisation) et via les mouvements des yeux (Études 2 et 3, catégorisation et recherche visuelle).

Les tâches d'identification/catégorisation (et plus largement les tâches à choix forcés binaires, pour lesquelles deux réponses distinctes sont proposées) sont très largement employées dans le domaine de la prise de décision perceptuelle, mais seulement quelques études adressent le rôle du mouvement des yeux dans ce processus. *A contrario*, les tâches de recherche visuelles (et plus largement les tâches « présent/absent » ou « signal/bruit ») sont beaucoup moins abordées sous cet angle (hormis dans le cadre de la TDS) et le rôle des mouvements des yeux dans le processus de décision est à peu près ignoré. Ainsi, les tâches de recherche visuelle employant des scènes artificielles (cibles et distrateurs abstraits, singletons) servent traditionnellement à étudier les mécanismes d'attention visuelle et les tâches de recherche visuelle avec des scènes naturelles servent très majoritairement à étudier l'exploration visuelle elle-même (attention ouverte).

Enfin, au travers des tâches présentées, nous manipulons la force du stimulus via l'ajout de bruit (gaussien ou brouillard) et les attendons de manière à mettre en avant leur effet sur le processus de prise de décision et à l'expliquer par la manière dont ils déterminent le comportement de collecte d'information, en modulant soit le CVE, soit le comportement oculomoteur.

1. Prise de décision perceptuelle pour la catégorisation : influence du contexte et *mouse-tracking*

Dans l'Étude 1, nous avons tenté de mettre en évidence l'effet du contexte sur le processus de prise de décision à l'œuvre lors de la catégorisation, à différents niveaux de force de l'évidence visuelle. Les participants devaient ainsi catégoriser des animaux et des outils (Expériences 1 et 2) ou des véhicules (Expérience 3) plus ou moins bruités. Un contexte congruent ou incongruent avec ceux-ci était présenté préalablement. Dans ces expériences, nous utilisons les temps de réponse, temps d'initiation des mouvements et les profils de vitesse comme mesures de la vitesse des réponses, et la déviation des trajectoires des réponses en direction de la réponse incorrecte comme mesure de l'exactitude des réponses. Nos résultats sont cohérents au regard de la littérature classique sur la prise de décision perceptuelle, en particulier le cadre théorique général relatif aux modèles d'échantillonnage séquentiel, *i.e.* l'accumulation d'évidence jusqu'à atteindre un seuil de décision.

1.1. Dégradation de l'information visuelle et accumulation d'évidence

Nous obtenons une dégradation des performances lorsque la cible est bruitée, à la fois au niveau temporel et au niveau de la déviation de la souris. Dans les modèles de diffusion (et autres modèles à échantillonnage séquentiel), la force du stimulus ou de l'évidence contenue dans le stimulus à propos de l'une ou l'autre des réponses affecte positivement le taux de diffusion (J. Palmer et al., 2005; Ratcliff & McKoon, 2008; Smith & Ratcliff, 2004). Il en résulte un taux de diffusion plus faible lorsque l'information visuelle est bruitée, entraînant le besoin de prendre plus de temps pour collecter de l'évidence et dépasser le seuil de décision et ainsi, des temps de réponse plus élevés et profils de vitesse retardés (Drugowitsch et al., 2012; Gold & Shadlen, 2001; J. Palmer et al., 2005; Röhrbein et al., 2015). L'incertitude *bottom-up* générée par le bruit induit par ailleurs des réponses plus lentes, la souris atteignant une plus faible vitesse, et des trajectoires moins directes. Ces résultats font écho à l'idée que les mouvements de la souris puissent refléter en partie l'accumulation d'évidence, par exemple le taux de diffusion dans les modèles à échantillonnage séquentiel. Cela présuppose un lien entre accumulation dynamique d'évidence et préparation/exécution de l'action, que la communication entre ces deux fonctions soit vue comme continue (Lepora & Pezzulo, 2015; Song & Nakayama, 2008) ou intermittente (Friedman et al., 2013). Lorsque le stimulus est bruité et qu'ainsi, son identité est incertaine, la compétition entre les deux réponses est plus forte, entraînant une plus forte attraction par la réponse incorrecte que lorsque le stimulus n'est pas bruité, ou alternativement, un mouvement plus indécis, dirigé dans un premier temps vers le centre de l'écran.

1.2. Prise de décision en présence d'information contextuelle

Par ailleurs, nos résultats distinguent deux effets du contexte présenté préalablement à la cible (et simultanément pour l'Expérience 2) sur le processus décisionnel. D'une part, la présence d'un contexte, qu'il soit congruent ou incongruent avec la cible, précipite le processus de prise de décision (biais du CVE en faveur de la vitesse), initiant un mouvement de réponse précoce. Dans le cadre des modèles d'échantillonnage séquentiel, une augmentation globale du niveau de référence de l'accumulation d'évidence (dans les accumulateurs de chaque réponses alternatives) permet de modéliser ce type de variations du CVE (Bogacz et al., 2010; Heitz, 2014). Cet effet a été obtenu dans chacune des trois expériences, indiquant que l'information visuelle contextuelle est bien traitée et accumulée comme évidence. D'autre part, la congruence du contexte présenté entraîne une facilitation en faveur de la réponse qui lui est reliée. L'absence de cet effet lorsque les présentations du contexte et de la cible sont séparées au niveau moteur (Expérience 3, presser le bouton *start* entre les deux stimuli) pourrait indiquer que ce biais n'est pas dû à une facilitation de la reconnaissance de la cible grâce à l'activation, par le contexte congruent, de représentations liées à celle-ci, mais plutôt à biais de réponse (Hollingworth & Henderson, 1998, 1999). Ce type d'effets est usuellement modélisé via une modification du niveau de référence de l'accumulation d'évidence en faveur d'une réponse, ce qui correspond à l'addition d'une certaine quantité d'évidence pour cette réponse (Churchland, Kiani, & Shadlen, 2008; Forstmann et al., 2016; Ratcliff et al., 2016; Smith & Ratcliff, 2004).

Dans le cadre de cette étude, il peut être avancé que la perception du contexte accumule de l'évidence en faveur de la réponse congruente (*e.g.*, *vivant* pour un contexte naturel) avant le déclenchement du mouvement de réponse, biaisant ainsi sa trajectoire (Expériences 1 et 2). Lorsque presser le bouton *start* précède directement l'apparition de la cible (Expérience 3), la décision en cours de traitement lors de la présentation du contexte, est celle de presser ce bouton. Ainsi, lors de la décision suivante, portant sur la catégorie de la cible, la congruence du contexte ne provoque pas de biais, malgré le fait que de l'information sensorielle ait été traitée et accumulée.

Cependant, l'absence d'effet de congruence dans l'Expérience 3 est à confirmer. En effet, nous avons modifié les cibles non-vivantes dans cette expérience pour répondre à une remarque des reviewers de l'article. Nous avons remplacé les outils par des véhicules, sémantiquement plus proches des animaux sauvages que les outils (cibles pouvant être animées, contextes congruents situés à l'extérieur) et il est donc nécessaire de répliquer ces résultats avec le matériel des Expériences 1 et 2. Si cette étude n'a pas permis de mettre en évidence un effet de congruence entre le contexte et la cible, où le contexte facilite la

reconnaissance de la cible, elle met néanmoins en lumière la modulation de l'accumulation d'évidence via l'adaptation du CVE lorsque de l'information contextuelle est présentée. De par la présence d'information contextuelle potentiellement pertinente, la compétition entre les deux réponses alternatives commence plus tôt et mène à des réponses moins exactes (au sens où elles sont déviées vers la réponse alternative).

1.3. Compromis vitesse-exactitude dans le paradigme de *mouse-tracking*

Par ailleurs, notre méthodologie présente un certain nombre de différences avec les tâches classiques de prise de décision, notamment au regard du CVE. Dans les expériences de l'Étude 1, si la vitesse est évaluée en partie de manière similaire aux études classiques sur le CVE, via les temps de réponse, l'exactitude est mesurée d'une manière différente (Wickelgren, 1977). Nous n'avons pas pris en compte les taux d'erreur, ceux-ci étant trop faibles dans nos expériences et leur avons préféré la déviation de la trajectoire de la souris par rapport à une trajectoire idéale (ligne droite entre le bouton *start* et le bouton de réponse correct). Cette mesure présente l'avantage de pouvoir évaluer l'exactitude des réponses pour chaque essais, et ne nécessite donc pas d'échelonner la difficulté de la tâche dans le but d'obtenir un taux d'erreur suffisamment élevé pour pouvoir l'analyser.

Conceptuellement, elle peut refléter deux types de réponses. Certaines réponses sont dirigées vers la réponse incorrecte dans un premier temps puis corrigées, d'autres réponses sont dirigées vers le centre de l'écran, avant de bifurquer vers la réponse correcte. Ces deux types de réponses reflètent des processus différents (au moins en partie) (Freeman & Dale, 2013), le premier résultant du choix rapide et précoce de la réponse incorrecte et le second d'une compétition initiale accrue entre les deux réponses alternatives menant à une indécision dans les premiers temps de la réponse. Cependant, dans les deux cas, les trajectoires peuvent être interprétées comme des réponses moins exactes au sens où elles traduisent une incertitude plus forte quant à la réponse correcte. Les tâches classiques présentant le CVE ne prennent généralement en compte que le premier cas de figure, le second cas de figure résultant plutôt à des réponses correctes dans les tâches de temps de réaction.

Dans le cadre de cette étude, nous n'avons pas tenté de distinguer ces deux sources d'erreur de trajectoire, d'autant plus que pour une trajectoire unique, une combinaison de ces deux sources pourrait expliquer la déviation (choix initial de la réponse incorrecte associé à une forte incertitude sur l'exactitude de ce choix). Enfin, les tâches de *mouse-tracking* ne sont généralement pas neutres au regard du CVE. Comme elles contraignent temporellement les réponses pour permettre au mouvement de réponse d'être exécuté alors que la décision n'est pas terminée (début du mouvement le plus tôt possible et limite de temps), elles biaisent *de facto* le CVE vers la vitesse.

Chapitre 4 – Discussion générale

Cette étude, à notre connaissance, rapporte les premières expériences prenant en compte la relation entre vitesse et exactitude des réponses dans un paradigme de *mouse-tracking* et de futures études dans ce cadre sont nécessaires pour éclaircir les propriétés spatio-temporelles de réponses continues (plutôt que binaires comme dans les tâches classiques). Le CVE permettant de mettre en lumière des mécanismes de modulation active de l'accumulation d'évidence dans le processus de prise de décision, son étude dans ce paradigme permettrait une meilleure compréhension de l'articulation entre mouvement, perception et décision. À ce niveau, l'Étude 1 présente une limite majeure au regard de notre problématique portant sur l'implication de mécanismes de collecte d'évidence actifs dans le processus de prise de décision. Le CVE est généralement considéré comme un mécanisme actif permettant l'adaptation à diverses situations où la difficulté et la pression temporelle varient (Bogacz et al., 2010; Heitz, 2014; Wickelgren, 1977). D'une part, nous observons des variations du CVE dans des situations où ces éléments varient de manière indirecte. Une plus grande quantité d'information à traiter (le contexte) peut induire un accroissement de la difficulté –trier l'information selon sa pertinence– et/ou de la pression temporelle –plus d'information à traiter dans un délai fixe. D'autre part, notre méthode consiste à inférer la variation du CVE, et non à mesurer directement un mécanisme de collecte d'évidence actif. Ainsi, dans les Études 2 et 3, nous nous sommes focalisés sur l'allocation du regard lors de la recherche visuelle, mesurer les mouvements des yeux étant une manière plus directe de rendre compte d'une collecte de d'évidence active.

2. Détection et vérification de la cible dans la recherche visuelle

Les Études 2 et 3 se focalisent sur l'allocation du regard dans des scènes visuelles naturelles lors de la recherche visuelle de cible (par opposition aux tâches de recherche visuelle avec stimuli artificiels plus ou moins abstraits). L'Étude 3 présente aussi une tâche d'identification (Expérience 1). Le couplage *mouse-tracking*/oculométrie impose certaines adaptations à la fois par rapport aux tâches de recherche visuelle (scènes naturelles) classiquement employées dans la littérature et aux tâches de *mouse-tracking*. Dans les tâches de recherche visuelles présentant des scènes visuelles naturelles, les participants sont soumis à peu de pression temporelle et l'objectif de la tâche est de trouver et fixer la cible du regard (puis d'appuyer sur un bouton). Dans une tâche de *mouse-tracking*, la pression temporelle est nécessairement plus forte. Les participants doivent commencer leur réponse quelques centaines de millisecondes après l'apparition de la scène, et répondre dans un délai assez court. De plus le regard n'est pas seulement employé pour rechercher la cible, mais aussi pour contrôler le mouvement de la souris, et est généralement dirigé vers les boutons de réponse une fois la cible détectée.

La méthodologie *mouse-tracking* n'emploie habituellement pas de tâche de recherche visuelle et dans un but d'adaptation, les contraintes temporelles la caractérisant doivent être relâchées dans une certaine mesure, de façon à permettre l'exécution de plusieurs saccades. En ce sens, les tâches présentées dans les Études 2 et 3 sont des compromis. Imposer un délai avant lequel le mouvement de réponse doit commencer (délai d'initiation du mouvement) est une manipulation fréquente dans les tâches de *mouse-tracking* qui prend tout son sens pour la recherche visuelle. Il permet de maintenir une pression temporelle suffisante pour assurer que la réponse ne commence pas alors que le choix est déjà effectué, tout en laissant un délai suffisant pour trouver la cible et donner la réponse finale. Dans les Études 2 et 3 nous nous sommes focalisés sur les mouvements des yeux relatifs à deux temps du processus de recherche visuelle : la détection de la cible (temps écoulé avant la première fixation sur la cible) et sa vérification (durée de la première séquence de fixations sur la cible).

2.1. Effet des attentes quant à la localisation de la cible et à son identité

La distinction entre ces temps a été assez largement utilisée pour étudier l'effet d'attentes à propos de la localisation de la cible dans la scène et de son apparence (Castelhana et al., 2008; Malcolm & Henderson, 2009, 2010). De manière générale, les résultats obtenus dans les Études 2 (localisation de la cible) et 3 (apparence de la cible) sont cohérents avec la littérature. Ainsi, dans l'Étude 2, la cible est

Chapitre 4 – Discussion générale

détectée plus rapidement et avec un nombre de fixations moins élevé lorsque placée à un endroit habituel, de manière consistante avec des études antérieures (Castelhana & Heaven, 2010, 2011; Castelhana & Witherspoon, 2016; Malcolm & Henderson, 2010; Pereira & Castelhana, 2014; Spotorno et al., 2014, 2015). Comme présenté dans les Chapitres 1 et 2 et largement discuté dans la littérature (voir par exemple Brockmole & Võ, 2010; Ehinger et al., 2009; Spotorno et al., 2015; Torralba et al., 2006), ce résultat confirme l'orientation prioritaire du regard vers les zones de l'image constituant des localisations pertinentes et habituelles pour la cible. Toutes ces études manipulent diverses formes de connaissances sur la cible et sur l'environnement qui déterminent la probabilité *a priori* de trouver la cible à chaque endroit de la scène. Par exemple, notons la fonction des objets (Castelhana & Heaven, 2011; Castelhana & Witherspoon, 2016), la congruence entre les lois physiques (gravité) et la structure globale de la scène (Spotorno et al., 2014, 2015) ou bien à un niveau local, les relations spatiales de la cible avec d'autres objets (Brooks et al., 2010; Eckstein et al., 2006; Pereira & Castelhana, 2014). Sur ce plan, cette expérience ne permet pas d'être plus précis, les critères employés pour placer la cible à des localisations habituelles et inhabituelles étant assez larges et globaux.

Par ailleurs, nous n'avons pas obtenu d'effet des attentes quant à la localisation de la cible sur son temps de vérification. Ce résultat est cohérent avec certaines études antérieures (Castelhana & Heaven, 2010; Castelhana & Witherspoon, 2016; Pereira & Castelhana, 2014; Spotorno et al., 2014) mais d'autres études montrent une facilitation de la vérification par ce type d'attentes (Castelhana & Heaven, 2011; Malcolm & Henderson, 2010). Un rappel important est que dans un paradigme de *mouse-tracking*, après vérification de la cible, le regard se porte généralement vers les boutons de réponse (destinations possibles du mouvement), au contraire des expériences citées où les participants devaient fixer la cible et presser un bouton. De plus, la limite de temps imposée par le paradigme provoque une réduction du temps de vérification lorsque le temps de détection est étendu, comme par exemple lorsque la cible est placée à un endroit inhabituel. Ces éléments peuvent expliquer l'absence d'effet sur les temps de vérification.

Dans l'Étude 3, un autre type d'attentes est manipulé. Les panneaux routiers, indiquant quel objet va le plus probablement faire partie de la scène, génèrent des attentes sur l'identité de l'objet inclus dans la scène visuelle, piéton ou cerf/chevreuil. Dans ces deux expériences, l'apparence des objets ne varie pas d'un essai à l'autre et est donc rapidement apprise, construisant de ce fait une représentation visuelle de l'objet en mémoire de travail, qui peut être utilisée pour la recherche visuelle (Bravo & Farid, 2009, 2012). On peut donc rapprocher l'effet de ces attentes de l'effet de connaissances sur l'apparence (caractéristiques visuelles) des objets à rechercher. Un certain nombre d'études ont montré une facilitation de la détection et vérification de la cible par ces connaissances (Castelhana & Heaven, 2010; Castelhana et al., 2008; Hout & Goldinger, 2015; Malcolm & Henderson, 2009, 2010, Spotorno et al., 2014, 2015).

De manière générale, les résultats de l'Étude 3 ne sont que partiellement consistants avec les études mentionnées. Ainsi, nous n'avons obtenu qu'une facilitation de la vérification de l'objet recherché lorsqu'il est attendu. Cela suggère que ces attentes ne facilitent que l'identification de l'objet en vision centrale, via l'activation de représentations visuelles congruentes avec l'objet inclus dans la scène. Il faut cependant noter que dans les expériences montrant aussi une facilitation de la détection de la cible, la représentation visuelle de la cible est donnée aux participants en présentant une image de celle-ci (par opposition à son nom) alors que dans nos expériences, c'est la probabilité de voir apparaître l'objet qui est manipulée. La présentation d'une image de la cible pourrait faciliter sa détection périphérique par l'amorçage de ses caractéristiques visuelles, ce que notre manipulation ne permet pas.

2.2. Effet de l'incertitude sur l'information visuelle

Enfin, augmenter l'incertitude visuelle, que ce soit par l'ajout de bruit (Étude 2), d'un brouillard artificiel 3D consistant essentiellement en une atténuation du contraste (Étude 3, Expérience 1) ou de brouillard écologique (Étude 3, Expérience 2) affectait aussi les mouvements des yeux. À notre connaissance, peu d'études se sont penchées sur l'effet de l'incertitude visuelle sur l'allocation du regard lors de la perception de scènes visuelles. De manière générale, il a été montré que la durée des fixations ne dépend pas que de processus internes aveugles à l'information visuelle entrante, mais est aussi en partie dépendante de l'information visuelle captée sur l'image (Henderson & Smith, 2009; Nuthmann et al., 2010). Explorer du regard des scènes visuelles dégradées augmente la durée des fixations, réduit l'amplitude des saccades ainsi que l'efficacité du balayage pour trouver la cible (*scan-paths* plus longs), augmentant de ce fait le temps de recherche de la cible (Loschky & McConkie, 2002; Nuthmann, 2013, 2014; Pomplun, Reingold, & Shen, 2001; Röhrbein et al., 2015). Nous obtenons des résultats cohérents dans l'Étude 2.

En ce qui concerne les temps de détection et de vérification, nous obtenons des résultats différents dans les trois expériences présentées dans ces études. Dans l'Étude 2, le bruit augmente le temps de vérification mais pas le temps de détection. Dans l'Expérience 2 de l'Étude 3, on observe le pattern inverse avec l'augmentation du niveau de brouillard écologique. Dans l'Expérience 1 de l'Étude 3, le brouillard 3D augmente les deux mesures. Pour expliquer ces divergences, les différences fonctionnelles entre détection et vérification de la cible peuvent être mises en avant, en particulier au regard de la vision centrale et périphérique. Nuthmann (2014) a ainsi mis en évidence que le temps de détection était plus affecté par la dégradation de la vision périphérique, alors que le temps de vérification est plutôt sensible à la dégradation de la vision centrale.

Chapitre 4 – Discussion générale

L'information visuelle est dégradée de manière différente dans ces trois expériences. Dans l'Étude 2, le bruit additif gaussien employé semble dégrader plus fortement l'information locale que globale (gène de la vision centrale) alors que les brouillards manipulés dans l'Étude 3 pourraient affecter la vision périphérique comme centrale à cause de la réduction globale du contraste qu'ils provoquent (Caro, Cavallo, Marendaz, Boer, & Vienne, 2007; Cavallo, Caro, Dore, Colomb, & Dumont, 2007). L'absence d'effet du brouillard sur le temps de vérification dans l'Expérience 2 et sa présence dans l'Expérience 1 pourrait s'expliquer par la différence des stimuli des deux expériences en termes de richesse et complexité de l'information visuelle. Si les scènes des deux expériences sont composées des mêmes éléments, disposés de manière quasi-identique, l'Expérience 1 présente des scènes 3D et un brouillard artificiel uniforme, des textures lisses et une palette de couleurs ternes. L'Expérience 2 présente des scènes photographiées avec un brouillard réaliste moins uniforme formant des amas. Les objets sont plus détaillés, baignant dans une lumière plus chaude (due aux phares de la voiture dont l'observateur, le point de vue des photographies, fait office de pilote). Les halos de lumière (phares arrières de la voiture suivie par l'observateur, transmissiomètre) et la réflexion de la lumière (sur le piéton, le chevreuil, le marquage au sol, les panneaux) sont plus réalistes. Les images sont donc plus contrastées, et la dégradation visuelle de la cible provoquée par le brouillard n'est probablement pas équivalente entre les deux expériences. L'apparence des objets étant connue, repérer un détail diagnostique de leur identité suffit à les reconnaître. Les scènes de l'Expérience 2 étant globalement plus contrastées que les scènes de l'Expérience 1 pour les raisons évoquées, ce type de détail diagnostique peut ne pas être trop dégradé alors que l'objet l'est globalement, et ainsi ne pas provoquer la dégradation des performances en vision centrale (temps de vérification) escomptée pour l'Expérience 2.

De manière générale, il est donc difficile de tirer des conclusions précises sur les effets d'une dégradation de l'information visuelle sur les mouvements des yeux relatifs à la détection et à la vérification de la cible au travers de ces trois expériences étant donné les différences de méthodes employées. Il est cependant raisonnable de pointer que ces effets semblent dépendre de la manière dont cette dégradation affecte la vision périphérique et centrale. La première permettant de capter une information globale, grossière, serait cruciale pour la détection. La seconde permettant de capter une information locale, détaillée, serait cruciale dans la vérification de l'objet à rechercher (Nuthmann, 2014). Suivant cette logique, un bruit visuel affectant différemment l'information globale et locale affecte la détection de la cible ou sa vérification en conséquence.

2.3. Distinction entre processus de détection et de vérification

Nos résultats mettent en avant l'influence de l'incertitude de l'information visuelle et de différents types d'attentes sur l'allocation du regard dans la scène visuelle. Au travers des Études 2 et 3, deux processus d'exploration visuelle se distinguent de manière consistante, la détection de la cible et sa vérification, mesurées directement par segmentation des séquences de fixations effectuées à chaque essai. Ces deux processus sont influencés de manière différente par l'information visuelle et par les deux types d'attentes manipulés. La détection est plutôt influencée par des attentes sur la localisation de la cible, ce processus visant à trouver la localisation de la cible, alors que la vérification est plutôt influencée par des attentes sur l'identité de la cible, ce processus consistant à vérifier que le percept fixé correspond bien à la cible. Selon qu'elle affecte plutôt l'incertitude de l'information visuelle en vision périphérique ou centrale, la dégradation de l'image influence plutôt la détection ou la vérification de la cible.

Cette distinction a déjà été mise en avant précédemment (Castelhana et al., 2008; Malcolm & Henderson, 2009; Pannasch & Velichkovsky, 2009; Tatler & Vincent, 2008). Pour rappel, Pannasch et Velichkovsky (2009) distinguent deux comportements d'exploration visuelle, les modes de traitement ambiant et local caractérisés par, respectivement, des saccades de grande amplitude/fixations de courte durée et des saccades de courte amplitude/fixations longues. L'alternance entre ces types de comportements (ou la gradation) dépend des demandes de la tâche en cours, le mode ambiant permettant la relocalisation du regard vers la prochaine aire pertinente et le mode local permettant la captation d'information plus précise dans cette aire (Tatler & Vincent, 2008). La distinction entre détection de cible et vérification de cible peut être vue comme une application de la distinction mise en avant par Pannasch et Velichkovsky à la recherche visuelle (avec la nuance qu'un comportement de traitement local puisse être effectué durant la période de détection pour vérifier que la cible ne se trouve pas à un endroit particulier). Cette distinction indique plus précisément que l'exploration visuelle est un comportement actif, s'adaptant aux demandes de la tâche en cours d'exécution sur la base de l'information visuelle disponible comme des attentes se construisant avec la captation de nouvelles informations.

3. Prise de décision perceptuelle dans la recherche visuelle

L'originalité des expériences présentées dans les Études 2 et 3 réside dans l'étude des liens entre trois niveaux –*i.e.* l'intégration de diverses sources d'information *bottom-up* et *top-down*, l'exploration visuelle des scènes visuelles et leur contribution à la tâche de recherche visuelle. La section précédente s'est focalisée sur l'intégration d'information pour diriger l'exploration visuelle. Pour mesurer les effets des deux premiers niveaux sur le troisième, nous nous sommes placés dans le cadre de la prise de décision perceptuelle. Comme présenté dans le Chapitre 2 et discuté précédemment, ce cadre théorique et méthodologique est extensivement employé pour étudier la perception visuelle. La reconnaissance et la catégorisation d'objets et de stimuli a bénéficié de beaucoup d'attention dans ce cadre, ce qui n'est pas le cas de la recherche visuelle dans des scènes naturelles. Dans ce domaine, les recherches antérieures se sont plutôt focalisées sur les sources d'information déterminant les mouvements des yeux lors de l'exploration visuelle (Tatler et al., 2011; Torralba et al., 2006; Vö & Wolfe, 2015; Wu et al., 2014).

Loin de mettre de côté cette approche plus classique de la recherche visuelle, ce travail visait à l'intégrer dans le cadre de la prise de décision perceptuelle, considérant l'exploration visuelle comme l'un des principaux mécanismes de collecte active d'information visuelle permettant l'échantillonnage et l'accumulation d'évidence. En cela, ces expériences contribuent à l'étude de la recherche visuelle sur deux aspects. Premièrement, elles mettent en valeur les contributions distinctes de deux processus –la détection et la vérification de la cible– à la prise de décision. Deuxièmement, elles mettent en lumière la manière dont l'évidence est accumulée dans le processus de prise de décision présidant à la recherche visuelle. Dans les tâches de catégorisation/identification, le stimulus présente des pièces d'évidence en faveur (ou en défaveur) d'une catégorie et/ou de la catégorie opposée et l'accumulation d'évidence en faveur des différentes réponses est symétrique (au sens que les règles d'accumulation et de décision sont identiques quelle que soit la réponse). Dans la recherche visuelle, l'accumulation pour la réponse *présent* se solde par la détection et reconnaissance de la cible, alors que l'accumulation pour la réponse *absent* ne peut pas reposer sur l'extraction d'évidence provenant d'un stimulus particulier, puisqu'elle repose sur l'absence de la cible. L'accumulation d'évidence est donc asymétrique.

Comme exposé précédemment, les temps de détection et de vérification étaient influencés par différentes attentes. Notamment, nous avons observé que le temps de détection était plutôt influencé par des attentes sur la localisation de la cible (Étude 2) alors que le temps de vérification était plutôt influencé par des attentes sur l'identité de la cible (Étude 3). De même, nous avons observé que différents types de bruit dégradant l'information visuelle (chacun manipulé dans une expérience différente) influençaient

différemment les temps de détection et de vérification, et nous avançons qu'une forme de bruit dégradant plutôt le champ visuel périphérique affectera plutôt le premier alors qu'une forme de bruit dégradant plutôt le champ visuel central affectera plutôt le second.

Nos analyses de médiation suggèrent que ces processus, influencés par l'intégration d'attentes et d'information visuelle plus ou moins bruitée, ont à leur tour une influence sur un processus de prise de décision dynamique, accumulant progressivement de l'évidence jusqu'à déterminer la présence ou l'absence de la cible. À ce titre, l'Étude 2 permet d'établir que les attentes sur la localisation de la cible influencent la décision (temps de réponse comme déviation de la souris) en majeure partie via l'effet qu'elles ont sur le comportement oculomoteur relatif à la détection de la cible. L'Étude 3 montre que les attentes sur l'identité de la cible influencent la décision en partie via leur effet sur le processus de vérification. L'incertitude visuelle quant à elle, affecte le processus de prise de décision par l'effet qu'elle a sur le processus de détection (Étude 3 Expérience 2) ou de vérification (Étude 2). Cela suggère donc un certain contrôle de l'accumulation d'évidence par le comportement d'exploration visuelle, lui-même déterminé (en partie) par différentes formes d'attentes, et s'adaptant à l'incertitude présente dans l'information visuelle. Nos résultats soulignent différents mécanismes de perception active permettant l'accumulation d'évidence en faveur d'une réponse *cible présente*, via la reconnaissance de la cible à l'endroit fixé du regard, et d'une réponse *cible absente*, en atteignant un certain niveau de confiance dans le fait que la scène ne contienne pas la cible.

3.1. Accumulation d'évidence pour la réponse *cible présente*

Pour accumuler de l'évidence en faveur de la réponse *cible présente*, le processus de vérification semble crucial, étant donné qu'il permet d'accumuler de l'évidence sur l'identité de la cible par l'exploration (la fixation) de celle-ci. Comme mis en avant dans l'Étude 3, les attentes sur l'identité de la cible influent sur la décision en partie parce qu'elles facilitent l'identification de la cible. Cette forme d'attentes n'a pas que cet effet sur le processus de décision. Elles peuvent aussi influencer la décision et la réponse motrice associée via un biais de réponse ne reposant pas sur l'intégration d'information visuelle, mais favorisant directement la réponse congruente au panneau (Hollingworth & Henderson, 1998, 1999; Summerfield & de Lange, 2014) et générant de l'anticipation motrice biaisant les mouvements de la souris (Bruhn, 2013; Bruhn, Huette, & Spivey, 2014). L'Étude 3 montre donc qu'au-delà d'un biais de réponse de ce genre, la décision dépend en partie de la facilitation de la reconnaissance de l'objet à trouver par la génération d'attentes préalables sur son identité. Cela est vrai à la fois lorsque la décision porte sur l'identification de la cible (Expérience 1) comme lorsqu'elle porte sur la recherche de cette cible (Expérience 2). Sur cette base,

Chapitre 4 – Discussion générale

on peut tracer un parallèle entre accumulation d'évidence pour la réponse *présent* en recherche visuelle et accumulation d'évidence pour l'identification d'une cible. Dans les deux cas, la fixation de la cible permet d'accumuler de l'évidence pour son identification. Néanmoins, si nous trouvons ce parallèle raisonnable (et très intuitif), nous ne le considérons pas forcément conclusif sur cette question, de par la différence de matériel entre les deux expériences (bien que l'organisation des scènes visuelles soit très proche) d'une part, et d'autre part, car la similitude entre ces patterns de résultats n'implique pas forcément des mécanismes identiques. Des études futures pourraient tenter de confirmer cette similitude entre accumulation d'évidence pour la présence d'une cible dans une tâche de recherche visuelle et pour la reconnaissance d'un stimulus dans une tâche d'identification/catégorisation, permettant de justifier l'emploi d'un cadre théorique commun, comme nous l'avons fait ici.

Enfin, l'incertitude de l'information visuelle affecte le temps passé à vérifier la cible, et cet effet affecte le temps passé à prendre une décision en faveur de la réponse *cible présente*, tel qu'obtenu dans l'Étude 2. Ainsi, une cible dégradée est plus difficilement reconnue, augmentant le temps de fixation requis (Nuthmann, 2014; Röhrbein et al., 2015) et ralentissant de ce fait l'accumulation d'évidence. En conséquence, nous avons observé une décision plus longue, et des mouvements plus lents. Cependant, la déviation du mouvement de réponse n'était pas affectée par l'augmentation du temps de vérification due au bruit. Cela indique que même si la reconnaissance est plus compliquée et longue, l'évidence accumulée alors que le processus de vérification se déroule, ne fait pas dévier la décision vers la réponse *cible absente*. L'incertitude générée par le fait que la cible soit dégradée n'induit pas non plus de mouvements dont les trajectoires sont dirigées vers le centre de l'écran (sur l'axe x) qui pourraient indiquer une indécision. L'incertitude dans la réponse lorsque la cible est dégradée, observable dans nos données, se traduit plutôt par une diminution de la vitesse du mouvement (Hehman et al., 2015).

Globalement, ces résultats peuvent indiquer que la réponse *absent* est rejetée durant la vérification de la cible mais avec une confiance dans la réponse réduite dû au bruit (jusqu'à la fin de la vérification). En revanche, dans l'Expérience 1 de l'Étude 3 –qui est centrée sur la reconnaissance de la cible (réponses *piéton* ou *cerf*)– la dégradation du processus de vérification par le brouillard augmente la déviation du mouvement vers la réponse incorrecte. Cette réponse incorrecte n'indiquant pas l'absence d'une cible mais l'identité concurrente à celle correspondant à la cible, l'incertitude quant à l'identité du stimulus fait que la réponse incorrecte est rejetée plus tardivement et influence donc la trajectoire du mouvement (Dale et al., 2007; Freeman & Ambady, 2011). Ici encore, à cause de la différence entre les scènes et le type de bruit utilisé dans les deux expériences, cette différence entre recherche visuelle et identification de cible est à confirmer dans le futur.

Enfin, un indice supplémentaire pointant le rôle particulier de la vérification de la cible dans

l'accumulation pour la réponse *présent*, est l'absence d'effet du temps de détection sur la déviation de la souris lorsque la tâche porte sur la reconnaissance de la cible (Étude 3, Expérience 1). Dans cette tâche, tant que la cible n'est pas détectée, la sélection de la réponse n'est pas particulièrement biaisée. Si ce résultat n'est pas présenté explicitement dans l'Article 3, il est déductible du fait qu'il y a un effet du brouillard sur le temps de détection mais pas de médiation de l'effet du brouillard sur les aires-sous-la-courbe, indiquant donc que l'effet du temps de détection sur les aires sous la courbe n'est pas significatif (à titre indicatif, la pente de cet effet et son intervalle de confiance –évalué via la procédure de bootstrap permettant de tester la médiation– sont les suivants : $b = 0.0004$, IC 95% = [-0.00006, 0.0007]).

De manière générale, nos études n'ont pas pu clairement montrer d'accumulation d'évidence en faveur la réponse *cible présente* durant la période de détection. Cela ne veut pas dire qu'elle n'existe pas. Notre méthodologie impliquait d'exclure les essais pour lesquels la cible n'était pas fixée directement au moins sur une fixation, ce qui représentait entre 10.8% (Étude 3 Expérience 2) et 15.5% (Étude 2) des essais. Cette limitation était nécessaire pour pouvoir calculer les périodes de détection et de vérification et travailler sur des essais valides au regard de nos hypothèses. Si dans certains de ces essais, le regard a pu se porter aux environs de la cible sans être dans le périmètre défini et ainsi accumuler de l'évidence sur la cible en vision parafovéale, il est aussi possible que suffisamment d'évidence ait pu être accumulée en vision périphérique, sans fixer directement la cible. Nos études n'excluent donc pas cette possibilité et si elles ne peuvent conclure sur un rôle exclusif de la vérification de la cible sur l'accumulation en faveur de sa présence dans la scène visuelle, au moins pointent-elles son rôle central.

3.2. Accumulation d'évidence pour la réponse *cible absente*

La recherche visuelle a la particularité de présenter un processus de prise de décision reposant sur une accumulation d'évidence asymétrique où converger vers une réponse *cible présente* ne repose pas sur l'accumulation du même type d'évidence que converger vers une réponse *cible absente*. Ainsi, plutôt que détecter et reconnaître la cible (comparer l'information visuelle intégrée à une représentation de la cible), le choix de la réponse *cible absente* résulte, au contraire, de l'échec à trouver son emplacement. Nos résultats vont en ce sens. Dans l'Étude 2, nous avons observé la déviation des mouvements de réponse vers la réponse *absent* (réponse incorrecte dans cette expérience) alors que temps de détection augmente (effet du temps de détection sur la déviation), cette déviation débutant à des intervalles de temps antérieurs à la détection de la cible. Ce résultat peut refléter l'accumulation d'évidence en faveur de la réponse *absent* au fur et à mesure que du temps est consacré à rechercher la cible infructueusement. Néanmoins, l'expérience de l'Étude 2 n'était pas construite pour tester précisément cette explication et ne permettait donc que de la

Chapitre 4 – Discussion générale

suggérer. Notamment, l'effet décrit peut aussi être expliqué par le fait que les participants débutent prudemment leur mouvement de réponse le long de l'axe vertical central tant qu'ils n'ont pas trouvé la cible ou échoué à la trouver dans un temps raisonnable. Cette explication alternative ne présuppose pas une accumulation d'information au cours du processus de détection.

L'expérience 2 de l'Étude 3 permettait d'exclure cette explication alternative. Dans cette étude, si l'objet faisant partie de la scène était un distracteur (chevreuil) ou s'il n'y avait pas d'objet dans la scène, les participants devaient choisir la réponse *absent* alors que si c'était une cible (piéton), ils devaient choisir la réponse *présent*. Ainsi, si de l'évidence est accumulée en faveur de la réponse *absent* au fur et à mesure du processus de détection, la déviation du mouvement de réponse vers cette réponse devait être observable qu'une cible ou un distracteur soit présenté. Lorsque la cible est présentée, cela résulte en une augmentation de la déviation vers la réponse incorrecte *absent* avec l'augmentation des temps de détection. Lorsque le distracteur est présenté, cela résulte en une diminution de la déviation vers la réponse incorrecte *présent* avec l'augmentation des temps de détection (ou une absence d'effet pour des raisons de métrique de la mesure de déviation). L'hypothèse d'un mouvement prudent le long de l'axe central postulerait une augmentation de la déviation avec l'augmentation des temps de détection quel que soit l'objet présenté. Les résultats de cette expérience confirment l'hypothèse d'une accumulation d'évidence en faveur de la réponse *absent* durant le temps où la cible est cherchée du regard mais pas (encore) trouvée ; la déviation vers la réponse incorrecte augmentait avec les temps de détection lorsque la cible faisait partie de la scène et elle diminuait avec l'augmentation des temps de détection lorsqu'un distracteur faisait partie de la scène.

Cette accumulation d'évidence permet d'orienter progressivement la décision vers la possibilité que la cible soit absente. Ces résultats apportent de nouveaux éléments à un problème assez rarement abordé dans les modèles de recherche visuelle, le problème de l'arrêt de la recherche (*search termination problem*; Wolfe, 2012, 2013), soit la décision de stopper une activité de recherche visuelle lorsque la cible est introuvable. Dans le cadre de la recherche visuelle dans des scènes visuelles artificielles, plusieurs modèles ont été proposés pour expliquer ce problème. Par exemple, Cousineau et Shiffrin (2004) soutiennent l'idée que la probabilité de stopper la recherche dépend du nombre d'items examinés par rapport au nombre total d'items (cible et distracteurs). Voir Moran, Zehetleitner, Liesefeld, Müller et Usher (2015) pour des implémentations supplémentaires de cette idée. Alternativement, Wolfe (2012) propose l'établissement d'une limite temporelle ou d'une limite de fixations sur la base du nombre d'items estimé grossièrement au début de la recherche visuelle (sur la base d'information globale) et de la probabilité *a priori* que la cible fasse partie de la scène. Pour une implémentation d'un tel mécanisme, voir la revue critique des modèles sur le sujet de Wolfe (2012). Dans des scènes visuelles artificielles, le nombre d'items (singletons, stimuli

abstraites...) est fixe, et les emplacements possibles saillants car occupés soit par une cible, soit par un distracteur. Dans une scène visuelle naturelle, le nombre d'items composant la scène est difficile à estimer (qu'est ce qui est un item et qu'est ce qui ne l'est pas) et le nombre de localisations où une cible pourrait se trouver est potentiellement infini.

Cette contrainte impose au système visuel de hiérarchiser les endroits de la scène à explorer successivement, de manière à optimiser l'information visuelle intégrée en vision centrale comme périphérique. Dans nos expériences, le nombre de fixations effectuées sur les scènes visuelles était très faible, notamment par rapport au nombre de positions possibles (par exemple, dans l'Étude 3 où les scènes sont très simples, les essais avec plus de deux fixations durant la période de détection incluant la latence de la première saccade, sont très rares). Le nombre d'endroits fixés n'est donc pas la seule variable déterminante dans cette accumulation d'évidence. Nos résultats appuient sur l'importance des aspects temporels de l'exploration visuelle dans l'accumulation d'évidence, l'information visuelle étant intégrée progressivement durant chaque fixation, ce dont les modèles précédents ne rendent pas compte (voir Chapitre 1 Section 1., ou pour une revue de littérature, Hegdé, 2008). Sur ce plan, l'incertitude sur l'information visuelle augmente le temps de fixation nécessaire à l'extraction d'évidence, ralentissant donc la détection de la cible, et accumulant de l'évidence en faveur de la réponse *cible absente* au fur et à mesure de l'exploration, comme observé dans l'Étude 3.

Enfin, les attentes sur la localisation de la cible (Étude 2) influençaient ce mécanisme d'accumulation d'évidence vers la réponse *absent*. Les attentes sur la localisation affectaient la détection de la cible (détection plus longue pour une cible à un endroit inhabituel) et modéraient son effet sur la déviation du mouvement vers la réponse *cible absente*. Les longues périodes de détection déviaient les trajectoires dans une plus grande mesure lorsque la cible était à un endroit inhabituel que lorsqu'elle était à un endroit habituel. Cette interaction, ainsi que le rôle de l'incertitude visuelle dans la convergence vers une réponse *cible absente* (Étude 3) suggèrent que le temps passé sans trouver la cible, lorsque conçu via le simple établissement d'une limite temporelle au moment où la scène est présentée, tel que décrit par Wolfe (2012), ne peut pas expliquer à lui seul le choix d'une réponse *absent*. L'engagement dans un comportement de collecte active d'évidence, via l'exploration visuelle de la scène, serait crucial dans la décision de stopper la recherche, accumulant de l'évidence jusqu'à statuer que la cible est *absente* lorsqu'il s'avère infructueux.

3.3. Modéliser l'accumulation d'évidence pour l'absence de la cible

L'adaptation des modèles à échantillonnage séquentiel à la recherche visuelle en scènes naturelles

Chapitre 4 – Discussion générale

repose fondamentalement sur comment implémenter l'accumulation d'évidence pour une réponse *cible absente*. L'idée sous-jacente à nos résultats est que l'absence d'évidence pour une réponse *cible présente* peut servir d'évidence pour la réponse *cible absente*. Cette idée a été exploitée avec succès par Dufau, Grainger et Ziegler (2012) dans un modèle à accumulateurs (basé sur Usher et McClelland, 2001) simulant une tâche de décision lexicale opposant des mots et des non-mots (pseudo-mots). L'évidence sensorielle pour une réponse *non-mot* à chaque intervalle de temps correspondait à une constante *moins* l'évidence sensorielle pour une réponse *mot*. Ainsi, la convergence vers la réponse *non-mot* dépend à la fois du temps écoulé et d'information lexicale. Ce modèle base la décision de stopper la recherche sur l'accumulation d'évidence sensorielle, et s'oppose ainsi à des modèles qui n'en rendraient compte que par la mise en place d'une limite temporelle ou de signaux d'urgence magnifiant le taux de déviation ou réduisant le seuils de la réponse *absent* en fonction du temps (voir par exemple Thura, Beauregard-Racine, Fradet, & Cisek, 2012; Winkel, Keuken, van Maanen, Wagenmakers, & Forstmann, 2014; Wolfe, 2012). Il faut cependant noter que de tels mécanismes d'urgence ne sont pas incompatibles avec l'accumulation d'évidence sensorielle et ils peuvent permettre d'implémenter l'adaptation aux limites temporelles d'une tâche ou bien une fonction de coût de l'accumulation d'évidence, lorsque appliqués à toutes les réponses dans un modèle à échantillonnage séquentiel (Ditterich, 2006; Drugowitsch et al., 2012; Ratcliff et al., 2016).

De la même manière que dans le modèle de (Dufau et al., 2012), les travaux présentés dans les Études 2 et 3 suggèrent que le choix de la réponse *absent* dépend à la fois du temps écoulé et de l'accumulation d'information visuelle sur la présence de la cible dans la scène. Ces données vont plus loin, plaçant l'exploration visuelle comme permettant d'échantillonner l'information via l'hétérogénéité entre vision centrale et périphérique et les mouvements des yeux, et ainsi de collecter activement de l'évidence. Cependant, elles ne permettent pas de conclure sur l'impact de l'activité motrice et attentionnelle sous-jacente à l'exploration visuelle sur l'accumulation d'évidence. En plus de collecter activement de l'évidence supplémentaire, poursuivre l'exploration visuelle (sans la trouver) peut avoir un coût en matière de ressources attentionnelles et motrices (sélectionner les localisations et planifier des saccades oculaires supplémentaires) pouvant biaiser la décision vers la réponse *absent* (Drugowitsch et al., 2012; Wolfe, 2013). Ainsi, la décision de poursuivre ou stopper la recherche dépendrait du gain d'évidence sur la présence ou l'absence de la cible si la recherche est poursuivie d'une part et du coût anticipé en termes de ressources motrices/attentionnelles et temporelles (*e.g.*, délai de réponse), entraîné par la poursuite de l'exploration visuelle d'autre part. Nos travaux ne permettent pas de conclure sur cette question, notre protocole ne permettant pas de manipuler ces coûts.

4. Limites et perspectives pour la prise de décision lors de la recherche visuelle

4.1. Limites méthodologiques et théoriques

Les expériences présentées dans ces trois études mettent ainsi en avant le rôle de la vision active et de l'exploration visuelle dans la prise de décision perceptuelle. Les deux types de tâches ont été abordées dans cette thèse, sans doute inégalement en faveur de la recherche visuelle, avec pour objectif de les étudier dans un cadre et avec des mécanismes communs. Le cadre de la prise de décision perceptuelle semble être un cadre théorique et pratique pertinent pour satisfaire à cet objectif. En particulier, nous avons cherché à mettre en avant l'influence de l'incertitude de l'information visuelle (bruit visuel, brouillard) et des attentes sur le processus de prise de décision.

L'Étude 1, à défaut d'avoir pu montrer un effet de facilitation de la reconnaissance d'une cible par le contexte au-delà d'un biais de réponse, a permis de mettre en avant l'adaptation de l'accumulation d'évidence, notamment via le CVE, à la quantité d'information et à la temporalité de sa présentation, soulignant un mécanisme actif. Les Études 2 et 3 ont établi le rôle de l'exploration visuelle dans le processus de prise de décision à l'œuvre lors de la recherche visuelle, et dans une moindre mesure lors de l'identification d'un objet. Ces études ont permis de distinguer les rôles respectifs des processus de détection et vérification de la cible dans le processus de prise de décision, en particulier dans l'accumulation d'évidence en faveur, respectivement, des réponses *cible absente* et *cible présente*.

Il est important de noter que les conclusions de ces études devraient être confirmées dans de futurs travaux, certaines des expériences les composant incluant des biais. En particulier, une partie des conclusions de l'Étude 3 repose d'une part sur les similarités et différences de résultats entre les deux expériences, la première portant plutôt sur l'identification et la seconde sur la recherche visuelle. Seulement, les images employées, bien qu'ayant une configuration spatiale similaire et variant selon les mêmes critères (position de l'objet, brouillard, panneaux) ne sont pas construites de la même manière (3D vs. photographies). Les résultats de l'Expérience 1 devraient être répliqués avec des photographies et des paramètres identiques à l'Expérience 2, permettant ainsi une comparaison plus fiable de la prise de décision lors de l'identification et lors de la recherche visuelle. De même, les résultats de l'Expérience 2 révèlent une asymétrie entre réponses *présent* et *absent*. Cependant, le piéton est toujours la cible et le chevreuil le distracteur. Une réplification contrebalançant les statuts des deux objets est nécessaire pour vérifier que cette asymétrie ne soit pas due aux caractéristiques visuelles des deux objets.

Les Études 2 et 3 permettent de mettre en évidence la manière dont pourrait être accumulée

Chapitre 4 – Discussion générale

l'évidence dans les tâches de recherche visuelle dans des scènes visuelles naturelles, la réponse *cible présente* reposant sur la reconnaissance de la cible alors qu'accumuler activement de l'évidence sur son *absence* en la recherchant activement des yeux (sans la trouver) pourrait déterminer le choix de la réponse correspondante. Comme souligné, si nos résultats suggèrent un tel rôle pour l'exploration visuelle, ils ne permettent pas de conclure sur l'impact du coût moteur et/ou attentionnel de la poursuite du comportement d'exploration visuelle sur la décision (Drugowitsch et al., 2012). Nos expériences ne permettaient pas de distinguer une hypothèse où seule l'évidence accumulée contre la présence de la cible accumule de l'évidence vers la réponse *absent* et une hypothèse où s'ajoute un compromis entre l'incertitude restante dans l'environnement, les coûts associés à l'erreur et les coûts associés à l'exécution d'une saccade supplémentaire. Dans le même ordre d'idée, il a été montré que le coût d'une action biaise la prise de décision vers la réponse la moins coûteuse (Hagura, Haggard, & Diedrichsen, 2017), et proposé que l'engagement moteur pour une réponse (tel que mesuré dans un paradigme Mouse-Tracker) puisse biaiser la décision vers la réponse en question, poussant à la poursuite du mouvement plutôt qu'à sa modification (e.g., déviation de sa trajectoire) ou annulation, car un changement d'avis entraînerai des coûts moteurs supplémentaires (Catenacci Volpi et al., 2014; Lepora & Pezzulo, 2015). Dans l'optique de parfaire notre compréhension des interaction entre prise de décision, perception active et (oculo-) motricité, déterminer l'influence de ces coûts est important dans la mesure où ils conditionnent le comportement de collecte active d'évidence.

Enfin, nous avons mis en place des protocoles où les mécanismes de collecte d'évidence, que ce soient le CVE (Étude 1) ou les mouvements des yeux (Études 2 et 3), étaient mesurés (et non manipulés) et l'influence de ces mécanismes sur la prise de décision était évaluée via des méthodes de régression (notamment via des analyses de médiation). Ainsi, cet aspect de nos résultat est de nature corrélacionnelle, empêchant de tirer des conclusions claires sur la chaîne causale déterminant la décision. En particulier, les Études 2 et 3 visaient à tester un modèle à sens unique où information visuelle et attentes influencent les mouvements des yeux, et l'information récoltée via les mouvements des yeux affecte à son tour la décision que ce soient pour la recherche visuelle comme dans ces expériences, ou pour la reconnaissance de stimuli (Kietzmann et al., 2011). Cependant, à chaque instant, l'état du processus de décision lui-même peut influencer la stratégie d'exploration visuelle employée pour effectuer la tâche, par exemple porter le regard vers des endroits du stimulus visant à confirmer son identité (Kietzmann et al., 2011; Quinton et al., 2014) ou effectuer un comportement de balayage (global, ambient, exploratoire) ou de vérification (local, d'exploitation) (Kietzmann & König, 2015). L'état de l'accumulation d'évidence pourrait donc permettre de formuler des hypothèses sur l'environnement, et de tester ces hypothèses via les mouvements des yeux (Friston et al., 2012).

Dans la section suivante, nous décrivons un modèle théorique général sur l'interaction entre exploration visuelle et prise de décision lors de tâches de catégorisation et de recherche visuelle, pouvant rendre compte de ces données. Si ce modèle est majoritairement une reformulation de modèles computationnels existants, il constitue une modeste tentative de proposer une architecture générale permettant de rendre compte de la prise de décision et de la perception active à la fois lors du choix forcé entre deux alternatives (ou plus, tel que pour la catégorisation) et lors de la recherche visuelle ou la détection (présent/absent, oui/non, signal/bruit). En tout cas, intégrant différents modèles avec les données originales présentées dans ce document, il vise à résumer les aspects de la dynamique entre perception, action et décision mis en lumière au cours de cette thèse.

4.2. Perspectives pour la modélisation de la recherche visuelle comme un processus de prise de décision

De manière générale, ce travail se positionnait en continuité, au moins conceptuellement, d'une étude antérieure de Quinton et al. (2014). Cette étude propose un modèle à accumulateurs intégrant des principes de codage prédictif, dans lequel le système oculomoteur oriente activement la collecte d'évidence sur les parties diagnostiques d'animaux schématiques ambigus (chiens, girafes...) pour en déterminer la catégorie. Les mouvements des yeux sont ainsi effectués pour tester activement des hypothèses formulées *a priori* sur la catégorie du stimulus (basées sur l'état de la compétition entre les différentes catégories) en orientant le regard, et mettre à jour ces prédictions une fois l'information captée (voir Donnarumma, Costantini, Ambrosini, Friston, & Pezzulo, 2017; Friston et al., 2012 pour d'autres implémentations de ce principe). En plus de permettre la reconnaissance de la cible, cette accumulation active d'évidence guide en continu la réponse motrice effectuée via le mouvement de la souris. Cependant, les simulations effectuées avec ce modèle n'ont été comparées qu'aux trajectoires de souris des réponses effectuées par des participants humains effectuant la même tâche, pas à leurs mouvements des yeux.

D'autres modèles intégrant prise de décision perceptuelle et mouvements des yeux ont été proposés. Parmi eux, Lewis, Shvartsman et Singh (2013), dans une tâche consistant à déterminer si une suite de chaînes de caractères ne contient que des mots, proposent le décours en parallèle de deux processus de décision. Le premier (*saccade decision*) est destiné à déclencher une saccade vers la chaîne suivante (impliquant un système oculomoteur indépendant) si l'évidence captée durant la fixation n'est pas suffisante pour le choix (entre les réponses *que des mots* et *pas que des mots*). Le second accumule l'évidence captée durant les fixations effectuées successivement jusqu'à converger vers une réponse. L'approche neuro-inspirée de Purcell et collaborateurs (Purcell et al., 2010, 2012; Schall et al., 2011),

Chapitre 4 – Discussion générale

présentée dans le Chapitre 2 Section 1.2.2., intègre aussi une distinction entre la captation d'évidence perceptive d'une part et l'accumulation de cette évidence pour générer une réponse motrice d'autre part, par deux populations de neurones dits *visuels* et *de mouvement*, respectivement. L'information captée par les neurones visuels n'est accumulée par les neurones de mouvement que si elle produit une activité minimale au niveau des neurones visuels (mécanisme de palier).

La synthèse de ces modèles permet de décrire un cadre théorique général pour la prise de décision dans des tâches impliquant la vision active. Ce modèle général est illustré par la Figure 6. Se distinguent donc deux systèmes, l'un intégrant l'évidence perceptive et l'autre accumulant cette évidence pour prendre une décision. Lewis et al. (2013) et Quinton et al. (2014) distinguent en plus un système oculomoteur. Le système perceptif permet ainsi de capter l'information pertinente (*i.e.*, de l'évidence au regard de la décision) durant chaque fixation du regard (1). Il peut-être vu comme intégrant cette évidence progressivement (Hegdé, 2008; Purcell et al., 2010, 2012). Le système oculomoteur permet de diriger le regard vers le prochain endroit de la scène pertinent pour la tâche en cours. Il implique d'intégrer de l'information visuelle progressivement captée durant une fixation (2), notamment en vision périphérique, avec les connaissances quant à l'apparence et la position de la cible (7) pour orienter le regard vers des caractéristiques visuelles pertinentes (Friston et al., 2012; Quinton et al., 2014). Le système décisionnel accumule progressivement l'évidence intégrée par le système perceptif de manière à prendre la décision (4). Suivant Lewis et al. (2013), lorsque suffisamment d'information a été accumulée au niveau du système perceptif (et que poursuivre la fixation n'apporte pas significativement plus d'évidence) sans qu'au niveau décisionnel, l'accumulation d'évidence ne franchisse le seuil, la préparation d'une saccade oculaire est entamée (3).

Par ailleurs, dans le modèle de Purcell et al. (2010, 2012), l'accumulation d'évidence au niveau du système décisionnel (neurones du mouvement) ne débute qu'à partir de l'intégration d'un minimum d'évidence au niveau du système perceptif (activité des neurones perceptifs). Ce mécanisme de palier est aisément implémentable mais son utilité dans le cadre des tâches de prise de décision de plus haut niveau employées ici est à statuer. Enfin, le système décisionnel est en lien avec les effecteurs permettant de produire une réponse motrice, que ce soit pour appuyer sur un bouton comme dans les tâches traditionnelles, effectuer une saccade oculaire (Purcell et al., 2010, 2012; Schall et al., 2011) ou produire un mouvement du bras et de la main comme dans le présent travail (Freeman, Dale, et al., 2011; Song & Nakayama, 2009).

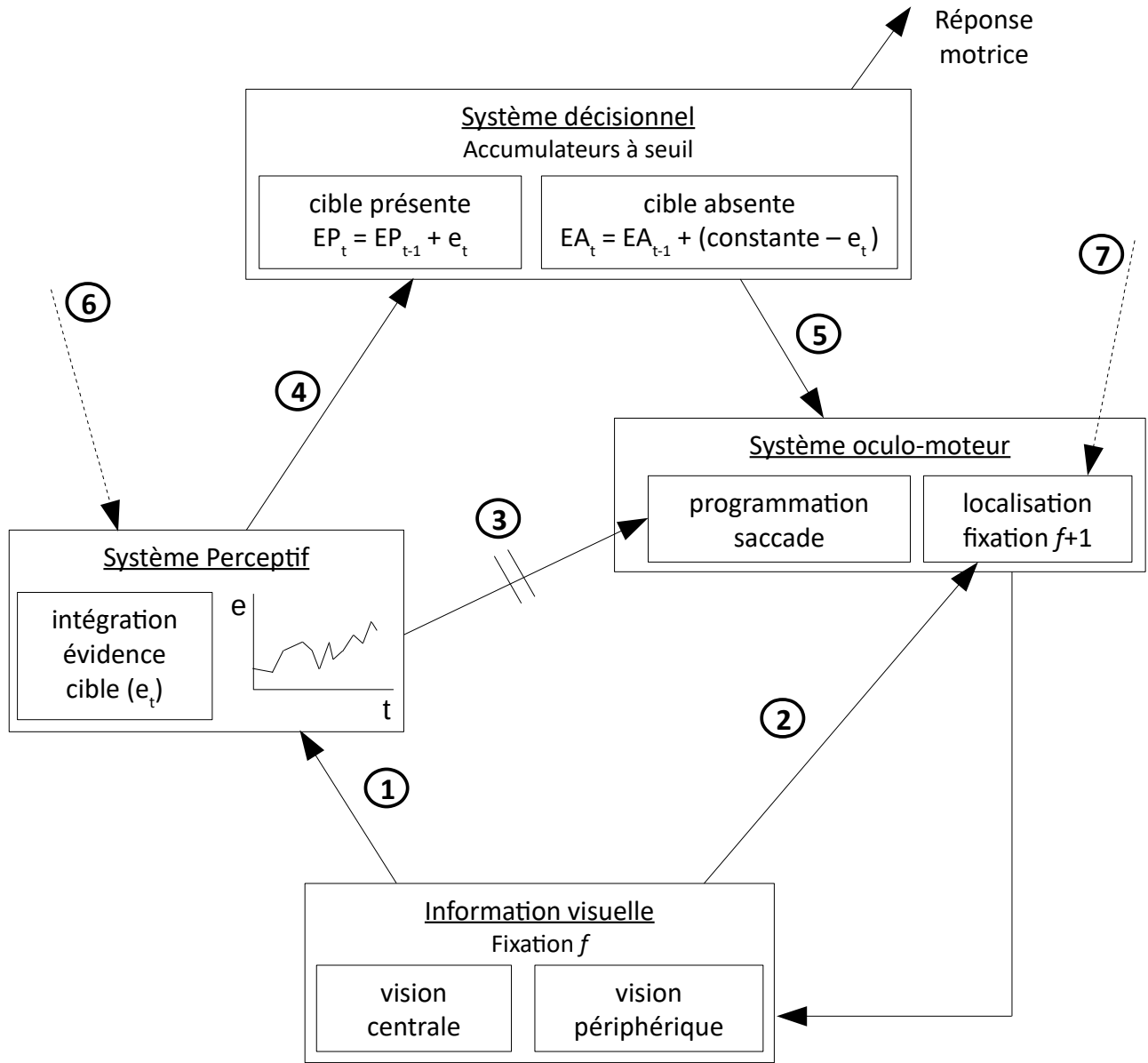


Figure 6: Schématisation d'un modèle général tentant d'intégrer intégration d'information visuelle, comportement d'exploration visuelle et prise de décision perceptuelle. Le modèle est composé de trois systèmes : le système oculomoteur oriente le regard pour capter de l'information à diverses localisations, le système perceptif intègre l'information visuelle pertinente dans le champ visuel, le système décisionnel accumule cette évidence pour prendre la décision. La fixation en cours est notée f , la prochaine fixation (en préparation) est notée $f+1$. L'intervalle de temps présent est noté t et l'intervalle de temps précédant le temps t est noté $t-1$. L'évidence en faveur de la présence de la cible (e.g., la similarité entre l'information visuelle et l'apparence attendue de la cible) intégrée à l'instant t , est notée e_t . L'évidence totale accumulée pour les réponses *présent* et *absent* sont notées respectivement EP et EA . Les règles d'accumulation d'évidence proposées le sont à titre d'exemple. Les liens indexés par des numéros sont explicités dans le corps du texte. Les flèches en pointillés indiquent l'intégration d'attentes (e.g., apparence ou localisation de la cible). Les attentes quant à la probabilité *a priori* de chaque réponse ne sont pas représentées mais sont simplement modélisables par un biais de réponse (baseline de l'accumulateur de la réponse attendue ou réduction de seuil).

Chapitre 4 – Discussion générale

Les systèmes perceptifs et décisionnels peuvent paraître redondants, mais cette distinction devient d'autant plus pertinente dans le cadre des tâches de recherche visuelle (et plus généralement, les tâches où les mouvements des yeux sont nécessaires) qu'il faut distinguer l'évidence intégrée lors d'une fixation (système perceptif) et l'évidence accumulée au cours des fixations effectuées jusqu'à lors (système décisionnel). Dans ce cadre, le système décisionnel exerce un contrôle sur le comportement oculomoteur (5). Ainsi, dans la recherche visuelle, de l'évidence est intégrée progressivement par le système perceptif à chaque fixation sur la base de la similarité entre l'information visuelle captée et l'apparence attendue de la cible. Dans la Figure 6, un exemple de règle d'accumulation d'évidence basée sur (Dufau et al., 2012) est proposé à titre indicatif pour une tâche de recherche visuelle, où une similarité forte accumulera beaucoup d'évidence pour la réponse *présent* et peu pour la réponse *absent* (constante - évidence). Si cette similarité est trop faible, de l'évidence pour une réponse *absent* est accumulée au niveau du système décisionnel et une saccade est effectuée, poursuivant le processus de détection de la cible. Si la similarité est forte, la fixation du regard est maintenue ou bien une saccade vers autre endroit de l'objet est effectuée pour capter plus d'information sur l'objet (s'il est assez grand), ou pour centrer l'objet en vision centrale (s'il a été capté en périphérie), de manière à vérifier la cible et accumuler suffisamment d'évidence pour une réponse *présent*.

Le maintien de la fixation peut notamment impliquer l'inhibition d'une saccade déjà programmée comme mis en évidence par Henderson et Smith (2009) (pour une modélisation de la durée des fixations implémentant un tel mécanisme d'inhibition, voir Nuthmann et al., 2010). Une fois la décision prise, le système décisionnel peut relâcher son contrôle sur le système oculomoteur, pour qu'il puisse être mobilisé sur une autre tâche (tel que le contrôle visuomoteur du mouvement de réponse), ou bien le maintenir pour accumuler plus d'évidence et augmenter la confiance dans la réponse (Kiani & Shadlen, 2009; Zylberberg, Fetsch, & Shadlen, 2016)

Ce modèle théorique général rend compte des résultats obtenus dans la littérature ainsi que dans les Études 2 et 3. Il correspond à l'idée que les fixations du regard permettent de réduire, en particulier, l'incertitude locale dans une scène visuelle (Renninger et al., 2007). L'intégration progressive d'évidence durant les fixations implique que la durée d'une fixation est contrôlée (au moins en partie) par l'information visuelle captée durant cette fixation (Nuthmann et al., 2010; J. Palmer et al., 2005). Ainsi, si l'information visuelle est bruitée, l'intégration au niveau perceptif sera plus lente, augmentant la durée des fixations en général (Nuthmann, 2013, 2014; Röhrbein et al., 2015), la durée des fixations durant la période de détection (Étude 2) et le temps de vérification (Études 2 et 3, Expérience 1). Cet effet est permis par le lien 3, où la quantité d'évidence intégrée influence la génération d'une saccade. L'information visuelle bruitée entraîne de ce fait une augmentation de la durée de la décision (évidence accumulée plus lentement, lien 4)

Chapitre 4 – Discussion générale

et un biais vers une réponse *absent* (si augmentation du temps de détection, Étude 3 Expérience 2).

Un tel modèle rend aussi compte du contrôle exercé par les attentes sur l'exploration visuelle et la reconnaissance (Henderson, 2017; Hwang, Wang, & Pomplun, 2011; Kietzmann et al., 2011; Vö & Wolfe, 2015), bien que cet aspect n'ait pas été au centre de la présente discussion. Le système décisionnel détermine le but de la tâche sous la forme d'une compétition entre alternatives (catégories, identités, présence ou absence d'un objet). Il permet ainsi la comparaison entre information visuelle et apparence attendue des stimuli (flèche 6) à reconnaître au niveau du système perceptif (effet des attentes de l'identité de la cible sur la vérification, Étude 3). Il contraint aussi les localisations vers lesquelles porter le regard pour capter de l'information diagnostique (lien 5, flèche 7), par exemple les localisations probables de la cible en recherche visuelle (et leur effet sur la détection de la cible, Étude 2) ou les endroits de l'objet où se trouvent un élément distinguant deux catégories possibles (Friston et al., 2012; Kietzmann et al., 2011; Kietzmann & König, 2015; Quinton et al., 2014). Sur la Figure 6, ces influences *top-down* sont cependant représentées de manière assez grossières, pour rester le plus général possible.

Conclusion

Ces recherches ont été effectuées dans le cadre du laboratoire d'excellence IMobS³ et avec la collaboration du CEREMA, apportant au projet des enjeux applicatifs dans le domaine de la sécurité routière et de la conduite en situation de vision dégradée. De manière générale, ce travail de thèse a permis de rendre compte de la relation entre prise de décision perceptuelle et mécanismes de collecte d'évidence. Dans ces deux processus, nous nous sommes focalisés sur l'intégration d'information visuelle plus ou moins bruitée avec des attentes à propos de la cible de la tâche en cours d'exécution. Les différentes études ont été placées dans un cadre théorique classique, où la prise de décision est vue comme résultante de l'accumulation progressive d'évidence jusqu'à un seuil, où la décision est prise. Ce cadre permet notamment d'expliquer via des mécanismes communs deux types de tâches classiques en psychologie, les tâches reposant sur un choix forcé entre différentes alternatives (dont la catégorisation et l'identification de stimuli), et la recherche visuelle reposant sur l'absence ou la présence d'un stimulus. Les mécanismes de collecte d'évidence active, tels que le compromis vitesse-exactitude et l'exploration visuelle, sont vus comme permettant de moduler l'accumulation d'évidence en adaptant le temps attribué à la collecte d'information et en l'orientant vers de l'information pertinente dans le cadre de la décision à effectuer. Ces mécanismes permettent d'une part d'adapter le processus de décision à des contraintes imposées par la situation dans laquelle est effectuée la tâche (contraintes temporelles, difficulté), et de l'adapter à un environnement visuel complexe où sélectionner l'information à intégrer est nécessaire.

Un tel cadre théorique centré sur la prise de décision ouvre des applications directes dans le champ de l'activité de conduite. La conduite est une tâche complexe, dans laquelle on peut distinguer diverses sous-tâches pouvant être exécutées les unes après les autres mais aussi, dans certaines situations, en parallèle, créant une situation de multi-tâche. Par exemple, arriver à une intersection requiert de piloter le véhicule (vérifier la distance avec les autres véhicules, la vitesse, les obstacles sur la route) et de naviguer dans l'environnement routier (décider de la direction à prendre, s'arrêter ou non au croisement) dans un laps de temps réduit (Lemonnier et al., 2015). Ces tâches impliquent fortement la vision et pour la plupart, peuvent être conceptualisées comme des tâches de prise de décision requérant un contrôle exécutif plus ou moins fort. Piloter le véhicule implique une activité de veille soutenue mais plutôt passive, plutôt automatique (dans une certaine mesure), débouchant sur une décision si une irrégularité survient –*e.g.*, un animal traverse la route, le véhicule précédent freine. Naviguer requiert de prendre des décisions plus planifiées et intentionnelles –*e.g.*, déterminer la trajectoire à prendre dans une intersection complexe en fonction des panneaux, du marquage au sol et des autres usagers. Dans l'Étude 3, nous nous étions basés sur une situation de conduite plus proche du pilotage du véhicule (détection d'un obstacle et réaction

Conclusion

consécutives de freiner). Cependant, la tâche que les participants avaient explicitement à faire n'y correspondait pas, *i.e.* rechercher activement du regard l'objet. Les conclusions et modèles que nous présentons rendent difficilement compte d'une situation réelle de détection d'obstacle lors de la conduite d'un véhicule dans la mesure où la décision est différente et où le comportement oculomoteur demandé par notre tâche, s'il peut être vaguement rapproché d'une détection d'obstacle, s'en détache néanmoins sur un point fondamental, la situation de veille (regarder devant soi jusqu'à l'apparition d'un obstacle potentiel dans le champ visuel).

Cependant, nos résultats peuvent apporter des éléments intéressants à l'étude du comportement des conducteurs, notamment oculomoteur, sur deux points : la perception en situation de brouillard et la recherche visuelle lors de la navigation dans l'environnement routier. Ainsi, cette thèse propose, à notre connaissance, une des premières études examinant les mouvements des yeux dans un brouillard écologique, ou au moins reproduisant un tel brouillard. Dans notre expérience, le brouillard dégradait la détection de la cible en vision périphérique notamment en augmentant la durée des fixations durant cette période, ce qui, comme mis en avant en discussion (Chapitre 4 Section 2.2.), est consistant avec des résultats plus généraux sur la dégradation de l'information visuelle par le bruit. Nos résultats ne sont pas conclusifs concernant la dégradation de la vérification de la cible par le brouillard. Nous avons avancé que lorsque la dégradation est en mesure d'affecter la vision centrale (comme dans l'Étude 2), la vérification de la cible est dégradée. Le design de l'expérience (un seul exemplaire de cible et distracteur répétés d'un essai sur l'autre) et l'organisation des stimuli (scènes peu complexes, caractéristiques distinctives de la cible et du distracteur pouvant probablement être intégrées rapidement en vision centrale) font que la reconnaissance de l'objet a pu n'être que faiblement affectée par le niveau de brouillard. De plus amples études devront statuer sur les effets du bruit sur la vérification de la cible d'une part, mais aussi sur l'influence du brouillard sur le mouvement de la cible à la fois pour sa détection et reconnaissance. Le fait que nous ayons utilisé des stimuli statiques limite nos conclusions.

En situation de conduite, la détection périphérique d'obstacles, en particulier de piétons et animaux sauvages, repose sur leur mouvement. Le rôle de la vérification dans la reconnaissance d'un obstacle mouvant est aussi une perspective importante, dans la mesure où en situation de conduite réelle, l'identification de l'obstacle pourrait être secondaire par rapport à la détermination de la trajectoire et de la vitesse de son mouvement. En plus de dégrader l'identification, le brouillard, mais aussi les intempéries de manière plus générale, pourraient dégrader l'évaluation de la vitesse et de la direction du mouvement d'un obstacle potentiel (Caro et al., 2007). Sur ce plan, les méthodes que nous avons employées, combinant *mouse-tracking* et oculométrie, pourraient permettre à de futures études une évaluation dynamique des processus décisionnels lors de la perception de scènes de conduite en mouvement.

Conclusion

Enfin, l'originalité de cette étude est de mettre en avant l'impact de la dégradation de l'information visuelle par le brouillard et l'impact des attentes sur un processus de prise de décision, notamment via leur influence sur l'exploration visuelle. Comme nous l'avons vu, le processus de décision étudié est difficilement adaptable (du moins, pas directement) au processus de prise de décision à l'œuvre lors de la détection d'un obstacle (et plus généralement lors du pilotage du véhicule), celle-ci étant plus réactive qu'active. Par contre, le paradigme de recherche visuelle s'applique mieux à la prise de décision lors de manœuvres de navigation, où il est fréquent que la recherche active d'évidence visuelle soit mise en place pour s'informer sur une direction, sur des directives permettant de visualiser l'organisation d'un carrefour ou d'une portion de route compliquée ou sur la position des autres usagers. Cette prise d'information permet de planifier la séquence de mouvements permettant de conduire le véhicule au travers d'une situation complexe (notamment visuelle) et génératrice d'incertitude. Déterminer s'il faut s'arrêter à un croisement nécessite ainsi de s'informer de la directive s'appliquant à la voie empruntée en recherchant du regard le panneau ou le marquage au sol donnant cette information, puis en fonction de la règle s'appliquant, rechercher du regard d'autres usagers sur les autres voies pour décider *in fine* de s'arrêter ou non.

Cette situation en particulier a été étudiée par Lemonnier et al. (2014, 2015) en simulateur de conduite, mettant en avant la distribution du regard entre l'exploration de l'environnement (en particulier les autres voies) pour supporter la décision de s'arrêter ou non, et le contrôle du véhicule. Leurs résultats montrent que le comportement d'exploration occupe une part plus grosse lorsqu'un choix devait être fait (panneau *cédez le passage*) ou que la situation pouvait présenter un risque (panneau *priorité* et/ou trafic dense) par rapport à une situation sans incertitude sur la décision à prendre (panneau *stop*) et/ou sur la complexité de l'environnement (pas de trafic) (Lemonnier et al., 2015). Ils montrent par ailleurs comment cette collecte d'évidence est déployée pour prendre la décision initiale avant l'arrivée au croisement et anticiper le comportement à déployer au moment venu, puis pour consolider cette décision et vérifier qu'elle est toujours valide à l'arrivée au croisement (Lemonnier et al., 2014). Sur ce plan, nos travaux peuvent être vus comme supplémentant les études de Lemonnier et collaborateurs dans la mesure où une décision complexe telle que décider de s'arrêter ou poursuivre à un croisement implique de vérifier la présence de véhicules, panneaux et autres stimuli visuels pertinents, et de les identifier, de façon à les intégrer comme évidence dans une décision plus complexe.

Nous avons proposé, comme perspectives théoriques en matière de modélisation (Chapitre 4 Section 4.2.), un modèle basé sur des principes devenus classiques dans le domaine de la prise de décision perceptuelle (échantillonnage et accumulation d'évidence), que nous avons plus particulièrement adapté à la recherche visuelle. Ce modèle est basé sur l'interaction entre un mécanisme d'identification de la cible (système perceptif) et un mécanisme de réallocation du regard (système oculomoteur) lorsque celle-ci ne se

Conclusion

trouve pas à l'endroit observé, doublé de l'implémentation d'un mécanisme d'accumulation d'évidence pour une réponse *cible absente* (système décisionnel). Un tel modèle serait aisément adaptable à diverses tâches constituant de l'activité de conduite automobile. Nous avons évoqué les tâches relatives à la navigation, reposant directement sur de la recherche visuelle de divers indices visuels tels que des panneaux directionnels (pouvant s'apparenter à une recherche de plusieurs cibles en parallèle). Tel que présenté, le modèle pourrait rendre compte (avec quelques adaptations) de l'activité de veille pour la détection d'obstacles (*e.g.*, piétons traversant la route, voiture se rabattant devant l'utilisateur), où les mouvements des yeux permettent essentiellement de vérifier un danger potentiel si le détecter en vision périphérique ne permet pas d'atteindre le seuil de la réponse *obstacle présent*, et d'accumuler de l'information supplémentaire utile au contrôle de l'action effectuée en réaction.

Par extension, un tel cadre permet d'aborder naturellement des processus méta-cognitifs plus complexes tels que la confiance dans la décision (Kiani & Shadlen, 2009; Zylberberg et al., 2016) ou les jugements de préférence (Konovalov & Krajbich, 2016; Krajbich, Armel, & Rangel, 2010). Cette versatilité apporte de nombreuses perspectives applicatives dans un vaste champ de recherche –notamment le domaine routier– pour la modélisation du traitement de l'information et de la prise de décision chez l'être humain (*e.g.*, conducteurs ou autres usagers). Des extensions possibles au travail proposé ici incluent des applications pour la vision et l'intelligence artificielle. Ces technologies sont aujourd'hui au centre des préoccupations dans tous les domaines, par exemple pour le développement de robots assistant les patients dans le domaine médical ou les ouvriers dans l'industrie. Dans le domaine routier, avoir une meilleure compréhension des interactions entre perception et décision chez l'être biologique est capital car il permet non seulement le développement de véhicules autonomes anticipant efficacement les dangers en provenance d'usagers humains, mais aussi de mieux prédire l'occurrence d'accidents à des endroits particuliers et/ou sous des conditions climatiques particulières.

Bibliographie

- Armel, K. C., Beaumel, A., & Rangel, A. (2008). Biasing simple choices by manipulating relative visual attention. *Judgment and Decision Making*, 3(5), 396–403.
- Ashby, F. G., & Soto, F. A. (2015). Multidimensional signal detection theory. In J. R. Busemeyer, Z. Wang, J. T. Townsend, & A. Eidels (Eds.), *Oxford handbook of computational and mathematical psychology* (pp. 13–34). New York: Oxford University Press. <http://doi.org/10.1093/oxfordhb/9780199957996.013.2>
- Auckland, M. E., Cave, K. R., & Donnelly, N. (2007). Nontarget objects can influence perceptual processes during object recognition. *Psychonomic Bulletin & Review*, 14(2), 332–337. <http://doi.org/10.3758/BF03194073>
- Ballard, D. H., & Hayhoe, M. M. (2009). Modelling the role of task in the control of gaze. *Visual Cognition*, 17(6–7), 1185–1204. <http://doi.org/10.1080/13506280902978477>
- Baluch, F., & Itti, L. (2011). Mechanisms of top-down attention. *Trends in Neurosciences*, 34(4), 210–224. <http://doi.org/10.1016/j.tins.2011.02.003>
- Banno, H., & Saiki, J. (2015). The use of higher-order statistics in rapid object categorization in natural scenes. *Journal of Vision*, 15(2):4, 1–20. <http://doi.org/10.1167/15.2.4>
- Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience*, 5(8), 617–629. <http://doi.org/10.1038/nrn1476>
- Bar, M. (2007). The proactive brain: Using analogies and associations to generate predictions. *Trends in Cognitive Sciences*, 11(7), 280–289. <http://doi.org/10.1016/j.tics.2007.05.005>
- Barca, L., Benedetti, F., & Pezzulo, G. (2016). The effects of phonological similarity on the semantic categorisation of pictorial and lexical stimuli: Evidence from continuous behavioural measures. *Journal of Cognitive Psychology*, 28(2), 159–170. <http://doi.org/10.1080/20445911.2015.1101117>
- Barca, L., & Pezzulo, G. (2012). Unfolding visual lexical decision in time. *PLoS ONE*, 7(4):e35932. <http://doi.org/10.1371/journal.pone.0035932>
- Barca, L., & Pezzulo, G. (2015). Tracking second thoughts: Continuous and discrete revision processes during visual lexical decision. *PLoS ONE*, 10(2):e0116193. <http://doi.org/10.1371/journal.pone.0116193>
- Battaglia, P. W., & Schrater, P. R. (2007). Humans trade off viewing time and movement duration to improve visuomotor accuracy in a fast reaching task. *Journal of Neuroscience*, 27(26), 6984–6994. <http://doi.org/10.1523/JNEUROSCI.1309-07.2007>
- Beck, D. M., & Kastner, S. (2005). Stimulus context modulates competition in human extrastriate cortex. *Nature Neuroscience*, 8(8), 1110–1116. <http://doi.org/10.1038/nn1501>
- Beck, D. M., & Kastner, S. (2009). Top-down and bottom-up mechanisms in biasing competition in the

Bibliographie

- human brain. *Vision Research*, 49(10), 1154–1165. <http://doi.org/10.1016/j.visres.2008.07.012>
- Biederman, I., Rabinowitz, J. C., Glass, A. L., & Stacy, E. W. (1974). On the information extracted from a glance at a scene. *Journal of Experimental Psychology*, 103(3), 597–600. <http://doi.org/10.1037/h0037158>
- Billot, R., El Faouzi, N.-E., & De Vuyst, F. (2009). Multilevel assessment of the impact of rain on drivers' behavior. *Transportation Research Record: Journal of the Transportation Research Board*, 2107, 134–142. <http://doi.org/10.3141/2107-14>
- Bogacz, R., Wagenmakers, E.-J., Forstmann, B. U., & Nieuwenhuis, S. (2010). The neural basis of the speed–accuracy tradeoff. *Trends in Neurosciences*, 33(1), 10–16. <http://doi.org/10.1016/j.tins.2009.09.002>
- Borji, A., & Itti, L. (2013). State-of-the-art in visual attention modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(1), 185–207. <http://doi.org/10.1109/TPAMI.2012.89>
- Borji, A., Tavakoli, H. R., Sihite, D. N., & Itti, L. (2013). Analysis of scores, datasets, and models in visual saliency prediction. In *2013 IEEE International Conference on Computer Vision* (pp. 921–928). IEEE. <http://doi.org/10.1109/ICCV.2013.118>
- Bravo, M. J., & Farid, H. (2009). The specificity of the search template. *Journal of Vision*, 9(1):34, 1–9. <http://doi.org/10.1167/9.1.34>
- Bravo, M. J., & Farid, H. (2012). Task demands determine the specificity of the search template. *Attention, Perception, & Psychophysics*, 74(1), 124–131. <http://doi.org/10.3758/s13414-011-0224-5>
- Broadbent, D. E. (1958). *Perception and communication*. Elmsford, NY, US: Pergamon Press, Inc. <http://doi.org/10.1037/10037-000>
- Brockmole, J. R., & Henderson, J. M. (2005a). Object appearance, disappearance, and attention prioritization in real-world scenes. *Psychonomic Bulletin & Review*, 12(6), 1061–1067. <http://doi.org/10.3758/BF03206444>
- Brockmole, J. R., & Henderson, J. M. (2005b). Prioritization of new objects in real-world scenes: Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance*, 31(5), 857–868. <http://doi.org/10.1037/0096-1523.31.5.857>
- Brockmole, J. R., & Henderson, J. M. (2006). Recognition and attention guidance during contextual cueing in real-world scenes: Evidence from eye movements. *The Quarterly Journal of Experimental Psychology*, 59(7), 1177–1187. <http://doi.org/10.1080/17470210600665996>
- Brockmole, J. R., & Vö, M. L.-H. (2010). Semantic memory for contextual regularities within and across scene categories: Evidence from eye movements. *Attention, Perception & Psychophysics*, 72(7), 1803–1813. <http://doi.org/10.3758/APP.72.7.1803>
- Brooks, D. I., Rasmussen, I. P., & Hollingworth, A. (2010). The nesting of search contexts within natural scenes: Evidence from contextual cuing. *Journal of Experimental Psychology: Human Perception and Performance*, 36(6), 1406–1418. <http://doi.org/10.1037/a0019257>

Bibliographie

- Bruhn, P. (2013). Emergence of spontaneous anticipatory hand movements in a probabilistic environment. *Advances in Cognitive Psychology*, 9(2), 62–73. <http://doi.org/10.2478/v10053-008-0132-y>
- Bruhn, P., Huette, S., & Spivey, M. (2014). Degree of certainty modulates anticipatory processes in real time. *Journal of Experimental Psychology: Human Perception and Performance*, 40(2), 525–538. <http://doi.org/10.1037/a0034365>
- Buc Calderon, C., Verguts, T., & Gevers, W. (2015). Losing the boundary: Cognition biases action well after action selection. *Journal of Experimental Psychology: General*, 144(4), 737–743. <http://doi.org/10.1037/xge0000087>
- Bullier, J. (2001). Integrated model of visual processing. *Brain Research Reviews*, 36, 96–107. [http://doi.org/10.1016/S0165-0173\(01\)00085-6](http://doi.org/10.1016/S0165-0173(01)00085-6)
- Buswell, G. T. (1935). *How people look at pictures*. Chicago: University of Chicago Press.
- Caro, S., Cavallo, V., Marendaz, C., Boer, E., & Vienne, F. (2007). The influence of fog on motion discrimination thresholds in car following. In *Proceedings of the 4th International Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design* (pp. 446–451).
- Carrasco, M. (2011). Visual attention: The past 25 years. *Vision Research*, 51(13), 1484–1525. <http://doi.org/10.1016/j.visres.2011.04.012>
- Castelhana, M. S., & Heaven, C. (2010). The relative contribution of scene context and target features to visual search in scenes. *Attention, Perception & Psychophysics*, 72(5), 1283–1297. <http://doi.org/10.3758/APP.72.5.1283>
- Castelhana, M. S., & Heaven, C. (2011). Scene context influences without scene gist: Eye movements guided by spatial associations in visual search. *Psychonomic Bulletin & Review*, 18(5), 890–896. <http://doi.org/10.3758/s13423-011-0107-8>
- Castelhana, M. S., & Henderson, J. M. (2007). Initial scene representations facilitate eye movement guidance in visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 33(4), 753–763. <http://doi.org/10.1037/0096-1523.33.4.753>
- Castelhana, M. S., Pollatsek, A., & Cave, K. R. (2008). Typicality aids search for an unspecified target, but only in identification and not in attentional guidance. *Psychonomic Bulletin & Review*, 15(4), 795–801. <http://doi.org/10.3758/PBR.15.4.795>
- Castelhana, M. S., & Witherspoon, R. L. (2016). How you use it matters: Object function guides attention during visual search in scenes. *Psychological Science*, 27(5), 606–621. <http://doi.org/10.1177/0956797616629130>
- Catenacci Volpi, N., Quinton, J.-C., & Pezzulo, G. (2014). How active perception and attractor dynamics shape perceptual categorization: A computational model. *Neural Networks*, 60, 1–16. <http://doi.org/10.1016/j.neunet.2014.06.008>
- Cavallo, V., Caro, S., Dore, J., Colomb, M., & Dumont, E. (2007). Risky driving in fog: Psychological explanations. In *Proceedings of the Road Safety on Four Continents Conference* (Vol. 14, pp. 12–12).

Bibliographie

- Cavanagh, P. (2011). Visual cognition. *Vision Research*, *51*(13), 1538–1551. <http://doi.org/10.1016/j.visres.2011.01.015>
- Cave, K. R., & Wolfe, J. M. (1990). Modeling the role of parallel processing in visual search. *Cognitive Psychology*, *22*, 225–271.
- Chen, Z. (2012). Object-based attention: A tutorial review. *Attention, Perception, & Psychophysics*, *74*(5), 784–802. <http://doi.org/10.3758/s13414-012-0322-z>
- Churchland, A. K., Kiani, R., & Shadlen, M. N. (2008). Decision-making with multiple alternatives. *Nature Neuroscience*, *11*(6), 693–702. <http://doi.org/10.1038/nn.2123>
- Cousineau, D., & Shiffrin, R. (2004). Termination of a visual search with large display size effects. *Spatial Vision*, *17*(4–5), 327–352. <http://doi.org/10.1163/1568568041920104>
- Dacey, D. M., & Petersen, M. R. (1992). Dendritic field size and morphology of midget and parasol ganglion cells of the human retina. *Proceedings of the National Academy of Sciences*, *89*, 9666–9670.
- Dale, R., Kehoe, C., & Spivey, M. J. (2007). Graded motor responses in the time course of categorizing atypical exemplars. *Memory & Cognition*, *35*(1), 15–28. <http://doi.org/10.3758/BF03195938>
- Davenport, J. L., & Potter, M. C. (2004). Scene consistency in object and background perception. *Psychological Science*, *15*(8), 559–564. <http://doi.org/10.1111/j.0956-7976.2004.00719.x>
- Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*, 193–222. <http://doi.org/10.1146/annurev.ne.18.030195.001205>
- Ditterich, J. (2006). Evidence for time-variant decision making. *European Journal of Neuroscience*, *24*(12), 3628–3641. <http://doi.org/10.1111/j.1460-9568.2006.05221.x>
- Donnarumma, F., Costantini, M., Ambrosini, E., Friston, K., & Pezzulo, G. (2017). Action perception as hypothesis testing. *Cortex*, *89*, 45–60. <http://doi.org/10.1016/j.cortex.2017.01.016>
- Droll, J. A., & Eckstein, M. P. (2009). Gaze control and memory for objects while walking in a real world environment. *Visual Cognition*, *17*(6–7), 1159–1184. <http://doi.org/10.1080/13506280902797125>
- Droll, J. A., Hayhoe, M. M., Triesch, J., & Sullivan, B. T. (2005). Task demands control acquisition and storage of visual information. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(6), 1416–1438. <http://doi.org/10.1037/0096-1523.31.6.1416>
- Drugowitsch, J., Moreno-Bote, R., Churchland, A. K., Shadlen, M. N., & Pouget, A. (2012). The cost of accumulating evidence in perceptual decision making. *Journal of Neuroscience*, *32*(11), 3612–3628. <http://doi.org/10.1523/JNEUROSCI.4010-11.2012>
- Dufau, S., Grainger, J., & Ziegler, J. C. (2012). How to say “no” to a nonword: A leaky competing accumulator model of lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(4), 1117–1128. <http://doi.org/10.1037/a0026948>
- Eckstein, M. P., Drescher, B. A., & Shimozaki, S. S. (2006). Attentional cues in real scenes, saccadic targeting, and bayesian priors. *Psychological Science*, *17*(11), 973–980. <http://doi.org/10.1111/j.1467->

Bibliographie

9280.2006.01815.x

- Ehinger, K. A., Hidalgo-Sotelo, B., Torralba, A., & Oliva, A. (2009). Modelling search for people in 900 scenes: A combined source model of eye guidance. *Visual Cognition*, 17(6–7), 945–978. <http://doi.org/10.1080/13506280902834720>
- Eriksen, C. W., & St. James, J. D. (1986). Visual attention within and around the field of focal attention: A zoom lens model. *Perception & Psychophysics*, 40(4), 225–240. <http://doi.org/10.3758/BF03211502>
- Fabre-Thorpe, M. (2011). The characteristics and limits of rapid visual categorization. *Frontiers in Psychology*, 2:243. <http://doi.org/10.3389/fpsyg.2011.00243>
- Farmer, T., Cargill, S., Hindy, N., Dale, R., & Spivey, M. (2007). Tracking the continuity of language comprehension: Computer mouse trajectories suggest parallel syntactic processing. *Cognitive Science: A Multidisciplinary Journal*, 31(5), 889–909. <http://doi.org/10.1080/03640210701530797>
- Faulkenberry, T. J. (2014). Hand movements reflect competitive processing in numerical cognition. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, 68(3), 147–151. <http://doi.org/10.1037/cep0000021>
- Feldman, H., & Friston, K. J. (2010). Attention, uncertainty, and free-energy. *Frontiers in Human Neuroscience*, 4:215. <http://doi.org/10.3389/fnhum.2010.00215>
- Fernandez-Duque, D., & Johnson, M. L. (2002). Cause and effect theories of attention: The role of conceptual metaphors. *Review of General Psychology*, 6(2), 153–165. <http://doi.org/10.1037/1089-2680.6.2.153>
- Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance*, 18(4), 1030–1044. <http://doi.org/10.1037/0096-1523.18.4.1030>
- Forstmann, B. U., Ratcliff, R., & Wagenmakers, E.-J. (2016). Sequential sampling models in cognitive neuroscience: Advantages, applications, and extensions. *Annual Review of Psychology*, 67, 641–666. <http://doi.org/10.1146/annurev-psych-122414-033645>
- Foulsham, T., & Underwood, G. (2008). What can saliency models predict about eye movements? Spatial and sequential aspects of fixations during encoding and recognition. *Journal of Vision*, 8(2):6, 1–17. <http://doi.org/10.1167/8.2.6>
- Freeman, J. B., & Ambady, N. (2009). Motions of the hand expose the partial and parallel activation of stereotypes. *Psychological Science*, 20(10), 1183–1188. <http://doi.org/10.1111/j.1467-9280.2009.02422.x>
- Freeman, J. B., & Ambady, N. (2010). MouseTracker: Software for studying real-time mental processing using a computer mouse-tracking method. *Behavior Research Methods*, 42(1), 226–241. <http://doi.org/10.3758/BRM.42.1.226>
- Freeman, J. B., & Ambady, N. (2011). Hand movements reveal the time-course of shape and pigmentation processing in face categorization. *Psychonomic Bulletin & Review*, 18(4), 705–712.

Bibliographie

<http://doi.org/10.3758/s13423-011-0097-6>

- Freeman, J. B., Ambady, N., Rule, N. O., & Johnson, K. L. (2008). Will a category cue attract you? Motor output reveals dynamic competition across person construal. *Journal of Experimental Psychology: General*, *137*(4), 673–690. <http://doi.org/10.1037/a0013875>
- Freeman, J. B., & Dale, R. (2013). Assessing bimodality to detect the presence of a dual cognitive process. *Behavior Research Methods*, *45*(1), 83–97. <http://doi.org/10.3758/s13428-012-0225-x>
- Freeman, J. B., Dale, R., & Farmer, T. A. (2011). Hand in motion reveals mind in motion. *Frontiers in Psychology*, *2*:59. <http://doi.org/10.3389/fpsyg.2011.00059>
- Freeman, J. B., Ma, Y., Han, S., & Ambady, N. (2013). Influences of culture and visual context on real-time social categorization. *Journal of Experimental Social Psychology*, *49*(2), 206–210. <http://doi.org/10.1016/j.jesp.2012.10.015>
- Freeman, J. B., Pauker, K., Apfelbaum, E. P., & Ambady, N. (2010). Continuous dynamics in the real-time perception of race. *Journal of Experimental Social Psychology*, *46*(1), 179–185. <http://doi.org/10.1016/j.jesp.2009.10.002>
- Freeman, J. B., Penner, A. M., Saperstein, A., Scheutz, M., & Ambady, N. (2011). Looking the part: social status cues shape race perception. *PLoS ONE*, *6*(9):e25107. <http://doi.org/10.1371/journal.pone.0025107>
- Friedman, J., Brown, S., & Finkbeiner, M. (2013). Linking cognitive and reaching trajectories via intermittent movement control. *Journal of Mathematical Psychology*, *57*(3–4), 140–151. <http://doi.org/10.1016/j.jmp.2013.06.005>
- Friston, K., Adams, R. A., Perrinet, L., & Breakspear, M. (2012). Perceptions as hypotheses: Saccades as experiments. *Frontiers in Psychology*, *3*:151. <http://doi.org/10.3389/fpsyg.2012.00151>
- Gerhard, H. E., Wichmann, F. A., & Bethge, M. (2013). How sensitive is the human visual system to the local statistics of natural images? *PLoS Computational Biology*, *9*(1), e1002873. <http://doi.org/10.1371/journal.pcbi.1002873>
- Gersch, T. M., Kowler, E., Schnitzer, B. S., & Doshier, B. A. (2009). Attention during sequences of saccades along marked and memorized paths. *Vision Research*, *49*(10), 1256–1266. <http://doi.org/10.1016/j.visres.2007.10.030>
- Gilbert, C. D., & Li, W. (2013). Top-down influences on visual processing. *Nature Reviews Neuroscience*, *14*(5), 350–363. <http://doi.org/10.1038/nrn3476>
- Gold, J. I., & Shadlen, M. N. (2001). Neural computations that underlie decisions about sensory stimuli. *Trends in Cognitive Sciences*, *5*(1), 10–16. [http://doi.org/10.1016/S1364-6613\(00\)01567-9](http://doi.org/10.1016/S1364-6613(00)01567-9)
- Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annual Review of Neuroscience*, *30*, 535–574. <http://doi.org/10.1146/annurev.neuro.29.051605.113038>
- Greene, M. R., & Oliva, A. (2009). The briefest of glances. *Psychological Science*, *20*(4), 464–472.

Bibliographie

<http://doi.org/10.1111/j.1467-9280.2009.02316.x>

- Hagura, N., Haggard, P., & Diedrichsen, J. (2017). Perceptual decisions are biased by the cost to act. *eLife*, 6, 1–20. <http://doi.org/10.7554/eLife.18422>
- Hanks, T. D., Mazurek, M. E., Kiani, R., Hopp, E., & Shadlen, M. N. (2011). Elapsed decision time affects the weighting of prior probability in a perceptual decision task. *Journal of Neuroscience*, 31(17), 6339–6352. <http://doi.org/10.1523/JNEUROSCI.5613-10.2011>
- Harley, E. M., Dillon, A. M., & Loftus, G. R. (2004). Why is it difficult to see in the fog? How stimulus contrast affects visual perception and visual memory. *Psychonomic Bulletin & Review*, 11(2), 197–231. <http://doi.org/10.3758/BF03196564>
- Heekeren, H. R., Marrett, S., & Ungerleider, L. G. (2008). The neural systems that mediate human perceptual decision making. *Nature Reviews Neuroscience*, 9(6), 467–479. <http://doi.org/10.1038/nrn2374>
- Hegd e, J. (2008). Time course of visual perception: Coarse-to-fine processing and beyond. *Progress in Neurobiology*, 84(4), 405–439. <http://doi.org/10.1016/j.pneurobio.2007.09.001>
- Helman, E., Stolier, R. M., & Freeman, J. B. (2015). Advanced mouse-tracking analytic techniques for enhancing psychological science. *Group Processes & Intergroup Relations*, 18(3), 384–401. <http://doi.org/10.1177/1368430214538325>
- Heitz, R. P. (2014). The speed-accuracy tradeoff: History, physiology, methodology, and behavior. *Frontiers in Neuroscience*, 8:150. <http://doi.org/10.3389/fnins.2014.00150>
- Heitz, R. P., & Schall, J. D. (2012). Neural mechanisms of speed-accuracy tradeoff. *Neuron*, 76(3), 616–628. <http://doi.org/10.1016/j.neuron.2012.08.030>
- Heitz, R. P., & Schall, J. D. (2013). Neural chronometry and coherency across speed-accuracy demands reveal lack of homomorphism between computational and neural mechanisms of evidence accumulation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368(1628), 20130071. <http://doi.org/10.1098/rstb.2013.0071>
- Henderson, J. M. (2017). Gaze control as prediction. *Trends in Cognitive Sciences*, 21(1), 15–23. <http://doi.org/10.1016/j.tics.2016.11.003>
- Henderson, J. M., Brockmole, J. R., Castelhano, M. S., & Mack, M. (2007). Visual saliency does not account for eye movements during visual search in real-world scenes. In R. P. G. Van Gompel, M. H. Fischer, W. S. Murray, & R. L. Hill (Eds.), *Eye movements: A window on mind and brain* (pp. 537–562). Amsterdam: Elsevier.
- Henderson, J. M., Malcolm, G. L., & Schandl, C. (2009). Searching in the dark: Cognitive relevance drives attention in real-world scenes. *Psychonomic Bulletin & Review*, 16(5), 850–856. <http://doi.org/10.3758/PBR.16.5.850>
- Henderson, J. M., & Smith, T. J. (2009). How are eye fixation durations controlled during scene viewing? Further evidence from a scene onset delay paradigm. *Visual Cognition*, 17(6–7), 1055–1082. <http://doi.org/10.1080/13506280802685552>

Bibliographie

- Hohwy, J. (2012). Attention and conscious perception in the hypothesis testing brain. *Frontiers in Psychology*, 3:96. <http://doi.org/10.3389/fpsyg.2012.00096>
- Hollingworth, A. (2009). Two forms of scene memory guide visual search: Memory for scene context and memory for the binding of target object to scene location. *Visual Cognition*, 17(1–2), 273–291. <http://doi.org/10.1080/13506280802193367>
- Hollingworth, A., & Henderson, J. M. (1998). Does consistent scene context facilitate object perception? *Journal of Experimental Psychology. General*, 127(4), 398–415.
- Hollingworth, A., & Henderson, J. M. (1999). Object identification is isolated from scene semantic constraint: Evidence from object type and token discrimination. *Acta Psychologica*, 102(2–3), 319–343. [http://doi.org/10.1016/S0001-6918\(98\)00053-5](http://doi.org/10.1016/S0001-6918(98)00053-5)
- Hout, M. C., & Goldinger, S. D. (2015). Target templates: The precision of mental representations affects attentional guidance and decision-making in visual search. *Attention, Perception, & Psychophysics*, 77(1), 128–149. <http://doi.org/10.3758/s13414-014-0764-6>
- Hwang, A. D., Wang, H.-C., & Pomplun, M. (2011). Semantic guidance of eye movements in real-world scenes. *Vision Research*, 51(10), 1192–1205. <http://doi.org/10.1016/j.visres.2011.03.010>
- Itti, L., & Baldi, P. (2009). Bayesian surprise attracts human attention. *Vision Research*, 49(10), 1295–1306. <http://doi.org/10.1016/j.visres.2008.09.007>
- Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, 40, 1489–1506. [http://doi.org/10.1016/S0042-6989\(99\)00163-7](http://doi.org/10.1016/S0042-6989(99)00163-7)
- Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience*, 2(3), 194–203. <http://doi.org/10.1038/35058500>
- Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11), 1254–1259. <http://doi.org/10.1109/34.730558>
- Johnson, K. L., Freeman, J. B., & Pauker, K. (2012). Race is gendered: How covarying phenotypes and stereotypes bias sex categorization. *Journal of Personality and Social Psychology*, 102(1), 116–131. <http://doi.org/10.1037/a0025335>
- Kanan, C., Tong, M. H., Zhang, L., & Cottrell, G. W. (2009). SUN: Top-down saliency using natural statistics. *Visual Cognition*, 17(6–7), 979–1003. <http://doi.org/10.1080/13506280902771138>
- Kaplan, E. (2004). The M, P and K pathways of the primate visual system. In L. M. Chalupa & J. M. Werner (Eds.), *The visual neurosciences* (pp. 481–493). Cambridge, MA: MIT Press.
- Kaplan, E. (2012). The M, P and K pathways of the primate visual system revisited. In J. Werner & L. Chalupa (Eds.), *The new visual neuroscience* (pp. 215–227). Cambridge, MA: MIT Press.
- Kauffmann, L., Chauvin, A., Guyader, N., & Peyrin, C. (2015). Rapid scene categorization: Role of spatial frequency order, accumulation mode and luminance contrast. *Vision Research*, 107, 49–57.

Bibliographie

<http://doi.org/10.1016/j.visres.2014.11.013>

Kauffmann, L., Ramanoël, S., & Peyrin, C. (2014). The neural bases of spatial frequency processing during scene perception. *Frontiers in Integrative Neuroscience*, 8:37.

<http://doi.org/10.3389/fnint.2014.00037>

Kelly, S. P., & O'Connell, R. G. (2013). Internal and external influences on the rate of sensory evidence accumulation in the human brain. *The Journal of Neuroscience*, 33(50), 19434–19441.

<http://doi.org/10.1523/JNEUROSCI.3355-13.2013>

Kiani, R., & Shadlen, M. N. (2009). Representation of confidence associated with a decision by neurons in the parietal cortex. *Science*, 324(5928), 759–764. <http://doi.org/10.1126/science.1169405>

Kietzmann, T. C., Geuter, S., & König, P. (2011). Overt visual attention as a causal factor of perceptual awareness. *PLoS ONE*, 6(7):e22614. <http://doi.org/10.1371/journal.pone.0022614>

Kietzmann, T. C., & König, P. (2015). Effects of contextual information and stimulus ambiguity on overt visual sampling behavior. *Vision Research*, 110, 76–86. <http://doi.org/10.1016/j.visres.2015.02.023>

Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology*, 4, 219–227.

Konovalov, A., & Krajbich, I. (2016). Gaze data reveal distinct choice processes underlying model-based and model-free reinforcement learning. *Nature Communications*, 7, 12438.

<http://doi.org/10.1038/ncomms12438>

Konstantopoulos, P., Chapman, P., & Crundall, D. (2010). Driver's visual attention as a function of driving experience and visibility. Using a driving simulator to explore drivers' eye movements in day, night and rain driving. *Accident Analysis & Prevention*, 42(3), 827–834. <http://doi.org/10.1016/j.aap.2009.09.022>

Kowler, E. (2011). Eye movements: The past 25 years. *Vision Research*, 51(13), 1457–1483.

<http://doi.org/10.1016/j.visres.2010.12.014>

Krajbich, I., Armel, C., & Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nature Neuroscience*, 13(10), 1292–1298. <http://doi.org/10.1038/nn.2635>

Krajbich, I., Hare, T., Bartling, B., Morishima, Y., & Fehr, E. (2015). A common mechanism underlying food choice and social decisions. *PLOS Computational Biology*, 11(10):e1004371.

<http://doi.org/10.1371/journal.pcbi.1004371>

Krajbich, I., Lu, D., Camerer, C., & Rangel, A. (2012). The attentional drift-diffusion model extends to simple purchasing decisions. *Frontiers in Psychology*, 3:193. <http://doi.org/10.3389/fpsyg.2012.00193>

Kveraga, K., Ghuman, A. S., & Bar, M. (2007). Top-down predictions in the cognitive brain. *Brain and Cognition*, 65(2), 145–168. <http://doi.org/10.1016/j.bandc.2007.06.007>

Land, M. F. (2009). Vision, eye movements, and natural behavior. *Visual Neuroscience*, 26, 51–62.

<http://doi.org/10.1017/S0952523808080899>

Land, M. F., & Hayhoe, M. (2001). In what ways do eye movements contribute to everyday activities? *Vision*

Bibliographie

- Research*, 41(25–26), 3559–3565. [http://doi.org/10.1016/S0042-6989\(01\)00102-X](http://doi.org/10.1016/S0042-6989(01)00102-X)
- Lemonnier, S., Brémond, R., & Baccino, T. (2014). Discriminating cognitive processes with eye movements in a decision-making driving task. *Journal of Eye Movement Research*, 7(4), 1–14. <http://doi.org/10.16910/jemr.7.4.3>
- Lemonnier, S., Brémond, R., & Baccino, T. (2015). Gaze behavior when approaching an intersection: Dwell time distribution and comparison with a quantitative prediction. *Transportation Research Part F: Traffic Psychology and Behaviour*, 35, 60–74. <http://doi.org/10.1016/j.trf.2015.10.015>
- Lepora, N. F., & Pezzulo, G. (2015). Embodied choice: how action influences perceptual decision making. *PLoS Computational Biology*, 11(4):e1004110. <http://doi.org/10.1371/journal.pcbi.1004110>
- Lewis, R. L., Shvartsman, M., & Singh, S. (2013). The adaptive nature of eye movements in linguistic tasks: How payoff and architecture shape speed-accuracy trade-offs. *Topics in Cognitive Science*, 5(3), 581–610. <http://doi.org/10.1111/tops.12032>
- Loschky, L. C., & McConkie, G. W. (2002). Investigating spatial vision and dynamic attentional selection using a gaze-contingent multiresolutional display. *Journal of Experimental Psychology: Applied*, 8(2), 99–117. <http://doi.org/10.1037//1076-898X.8.2.99>
- Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Malcolm, G. L., & Henderson, J. M. (2009). The effects of target template specificity on visual search in real-world scenes: Evidence from eye movements. *Journal of Vision*, 9(11):8, 1–13. <http://doi.org/10.1167/9.11.8>
- Malcolm, G. L., & Henderson, J. M. (2010). Combining top-down processes to guide eye movements during real-world scene search. *Journal of Vision*, 10(2):4, 1–11. <http://doi.org/10.1167/10.2.4>
- Malcolm, G. L., Nuthmann, A., & Schyns, P. G. (2014). Beyond gist: Strategic and incremental information accumulation for scene categorization. *Psychological Science*, 25(5), 1087–1097. <http://doi.org/10.1177/0956797614522816>
- Malcolm, G. L., & Shomstein, S. (2015). Object-based attention in real-world scenes. *Journal of Experimental Psychology: General*, 144(2), 257–263. <http://doi.org/10.1037/xge0000060>
- Marghetis, T., Núñez, R., & Bergen, B. K. (2014). Doing arithmetic by hand: Hand movements during exact arithmetic reveal systematic, dynamic spatial processing. *The Quarterly Journal of Experimental Psychology*, 67(8), 1579–1596. <http://doi.org/10.1080/17470218.2014.897359>
- Matsukura, M., Brockmole, J. R., Boot, W. R., & Henderson, J. M. (2011). Oculomotor capture during real-world scene viewing depends on cognitive load. *Vision Research*, 51(6), 546–552. <http://doi.org/10.1016/j.visres.2011.01.014>
- Matsukura, M., Brockmole, J. R., & Henderson, J. M. (2009). Overt attentional prioritization of new objects and feature changes during real-world scene viewing. *Visual Cognition*, 17(6–7), 835–855. <http://doi.org/10.1080/13506280902868660>

Bibliographie

- McCarley, J. S., Steelman, K. S., & Horrey, W. J. (2014). The view from the driver's seat: What good is salience? *Applied Cognitive Psychology, 28*(1), 47–54. <http://doi.org/10.1002/acp.2955>
- Mermillod, M., Bonin, P., Mondillon, L., Alleysson, D., & Vermeulen, N. (2010). Coarse scales are sufficient for efficient categorization of emotional facial expressions: Evidence from neural computation. *Neurocomputing, 73*(13–15), 2522–2531. <http://doi.org/10.1016/j.neucom.2010.06.002>
- Mermillod, M., Guyader, N., & Chauvin, A. (2005). The coarse-to-fine hypothesis revisited: Evidence from neuro-computational modeling. *Brain and Cognition, 57*(2), 151–157. <http://doi.org/10.1016/j.bandc.2004.08.035>
- Mills, M., Hollingworth, A., Van der Stigchel, S., Hoffman, L., & Dodd, M. D. (2011). Examining the influence of task set on eye movements and fixations. *Journal of Vision, 11*(8):17, 1–15. <http://doi.org/10.1167/11.8.17>
- Moore, T., Armstrong, K. M., & Fallah, M. (2003). Visuomotor origins of covert spatial attention. *Neuron, 40*(4), 671–683. [http://doi.org/10.1016/S0896-6273\(03\)00716-5](http://doi.org/10.1016/S0896-6273(03)00716-5)
- Moran, R., Zehetleitner, M., Liesefeld, H. R., Müller, H. J., & Usher, M. (2016). Serial vs. parallel models of attention in visual search: Accounting for benchmark RT-distributions. *Psychonomic Bulletin & Review, 23*(5), 1300–1315. <http://doi.org/10.3758/s13423-015-0978-1>
- Morett, L. M., & MacWhinney, B. (2013). Syntactic transfer in English-speaking Spanish learners. *Bilingualism: Language and Cognition, 16*, 132–151. <http://doi.org/10.1017/S1366728912000107>
- Mueller, A. S., & Trick, L. M. (2012). Driving in fog: The effects of driving experience and visibility on speed compensation and hazard avoidance. *Accident Analysis & Prevention, 48*, 472–479. <http://doi.org/10.1016/j.aap.2012.03.003>
- Muller, J. R., Philiastides, M. G., & Newsome, W. T. (2005). Microstimulation of the superior colliculus focuses attention without moving the eyes. *Proceedings of the National Academy of Sciences, 102*(3), 524–529. <http://doi.org/10.1073/pnas.0408311101>
- Nuthmann, A. (2013). On the visual span during object search in real-world scenes. *Visual Cognition, 21*(7), 803–837. <http://doi.org/10.1080/13506285.2013.832449>
- Nuthmann, A. (2014). How do the regions of the visual field contribute to object search in real-world scenes? Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance, 40*(1), 342–360. <http://doi.org/10.1037/a0033854>
- Nuthmann, A., & Henderson, J. M. (2010). Object-based attentional selection in scene viewing. *Journal of Vision, 10*(8):20, 1–19. <http://doi.org/10.1167/10.8.20>
- Nuthmann, A., Smith, T. J., Engbert, R., & Henderson, J. M. (2010). CRISP: A computational model of fixation durations in scene viewing. *Psychological Review, 117*(2), 382–405. <http://doi.org/10.1037/a0018924>
- Oliva, A., & Schyns, P. G. (1997). Coarse blobs or fine edges? Evidence that information diagnosticity changes the perception of complex visual stimuli. *Cognitive Psychology, 34*, 72–107. <http://doi.org/10.1006/cogp.1997.0667>

Bibliographie

- Oliva, A., & Torralba, A. (2006). Building the gist of a scene: The role of global image features in recognition. In S. Martinez-Conde, S. L. Macknik, L. M. Martinez, J.-M. Alonso, & P. U. Tse (Eds.), *Progress in brain research* (Vol. 155, pp. 23–36). Elsevier. [http://doi.org/10.1016/S0079-6123\(06\)55002-2](http://doi.org/10.1016/S0079-6123(06)55002-2)
- Orquin, J. L., & Mueller Loose, S. (2013). Attention and choice: A review on eye movements in decision making. *Acta Psychologica*, *144*(1), 190–206. <http://doi.org/10.1016/j.actpsy.2013.06.003>
- Palmer, J., Huk, A. C., & Shadlen, M. N. (2005). The effect of stimulus strength on the speed and accuracy of a perceptual decision. *Journal of Vision*, *5*, 376–404. <http://doi.org/10.1167/5.5.1>
- Palmer, S. E. (1975). The effects of contextual scenes on the identification of objects. *Memory & Cognition*, *3*(5), 519–526. <http://doi.org/10.3758/BF03197524>
- Palmeri, T. J., Schall, J. D., & Logan, G. D. (2015). Neurocognitive modeling of perceptual decision making. In J. R. Busemeyer, Z. Wang, J. T. Townsend, & A. Eidels (Eds.), *The Oxford handbook of computational and mathematical psychology* (pp. 320–340). New York: Oxford University Press. <http://doi.org/10.1093/oxfordhb/9780199957996.013.15>
- Pannasch, S., & Velichkovsky, B. M. (2009). Distractor effect and saccade amplitudes: Further evidence on different modes of processing in free exploration of visual images. *Visual Cognition*, *17*(6–7), 1109–1131. <http://doi.org/10.1080/13506280902764422>
- Papesh, M. H., & Goldinger, S. D. (2012). Memory in motion: Movement dynamics reveal memory strength. *Psychonomic Bulletin & Review*, *19*(5), 906–913. <http://doi.org/10.3758/s13423-012-0281-3>
- Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research*, *42*, 107–123. [http://doi.org/10.1016/S0042-6989\(01\)00250-4](http://doi.org/10.1016/S0042-6989(01)00250-4)
- Pereira, E. J., & Castelhana, M. S. (2014). Peripheral guidance in scenes: The interaction of scene context and object content. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(5), 2056–2072. <http://doi.org/10.1037/a0037524>
- Peyrin, C., Michel, C. M., Schwartz, S., Thut, G., Seghier, M., Landis, T., ... Vuilleumier, P. (2010). The neural substrates and timing of top-down processes during coarse-to-fine categorization of visual scenes: A combined fMRI and ERP study. *Journal of Cognitive Neuroscience*, *22*(12), 2768–2780. <http://doi.org/10.1162/jocn.2010.21424>
- Pleskac, T. J., & Busemeyer, J. R. (2010). Two-stage dynamic signal detection: A theory of choice, decision time, and confidence. *Psychological Review*, *117*(3), 864–901. <http://doi.org/10.1037/a0019737>
- Pomplun, M., Reingold, E. M., & Shen, J. (2001). Peripheral and parafoveal cueing and masking effects on saccadic selectivity in a gaze-contingent window paradigm. *Vision Research*, *41*(21), 2757–2769. [http://doi.org/10.1016/S0042-6989\(01\)00145-6](http://doi.org/10.1016/S0042-6989(01)00145-6)
- Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*, 3–25. <http://doi.org/10.1080/00335558008248231>
- Posner, M. I., Snyder, C. R. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General*, *109*(2), 160–174. <http://doi.org/10.1037/0096-3445.109.2.160>

Bibliographie

- Pradhan, A. K., Pollatsek, A., Knodler, M., & Fisher, D. L. (2009). Can younger drivers be trained to scan for information that will reduce their risk in roadway traffic scenarios that are hard to identify as hazardous? *Ergonomics*, *52*(6), 657–673. <http://doi.org/10.1080/00140130802550232>
- Purcell, B. A., Heitz, R. P., Cohen, J. Y., Schall, J. D., Logan, G. D., & Palmeri, T. J. (2010). Neurally constrained modeling of perceptual decision making. *Psychological Review*, *117*(4), 1113–1143. <http://doi.org/10.1037/a0020311>
- Purcell, B. A., Schall, J. D., Logan, G. D., & Palmeri, T. J. (2012). From salience to saccades: Multiple-alternative gated stochastic accumulator model of visual search. *Journal of Neuroscience*, *32*(10), 3433–3446. <http://doi.org/10.1523/JNEUROSCI.4622-11.2012>
- Purves, D., Augustine, G. J., Fitzpatrick, D., Katz, L. C., LaMantia, A.-S., McNamara, J. O., & Williams, S. M. (2001). *Neuroscience*. Sunderland (MA): Sinauer Associates.
- Quinton, J.-C., Catenacci Volpi, N., Barca, L., & Pezzulo, G. (2014). The Cat is on the mat. Or is it a dog? Dynamic competition in perceptual decision making. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, *44*(5), 539–551. <http://doi.org/10.1109/TSMC.2013.2279664>
- Rajashekar, U., Bovik, A. C., & Cormack, L. K. (2006). Visual search in noise: Revealing the influence of structural cues by gaze-contingent classification image analysis. *Journal of Vision*, *6*(4), 379–386. <http://doi.org/10.1167/6.4.7>
- Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87. <http://doi.org/10.1038/4580>
- Rao, R. P. N., Zelinsky, G. J., Hayhoe, M. M., & Ballard, D. H. (2002). Eye movements in iconic visual search. *Vision Research*, *42*, 1447–1463.
- Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation*, *20*(4), 873–922. <http://doi.org/10.1162/neco.2008.12-06-420>
- Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science*, *9*(5), 347–356. <http://doi.org/10.1111/1467-9280.00067>
- Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion decision model: Current issues and history. *Trends in Cognitive Sciences*, *20*(4), 260–281. <http://doi.org/10.1016/j.tics.2016.01.007>
- Rayner, K., McConkie, G. W., & Ehrlich, S. (1978). Eye movements and integrating information across fixations. *Journal of Experimental Psychology: Human Perception and Performance*, *4*(4), 529–544.
- Reeder, R. R., & Peelen, M. V. (2013). The contents of the search template for category-level search in natural scenes. *Journal of Vision*, *13*(3):13, 1–13. <http://doi.org/10.1167/13.3.13>
- Renninger, L. W., Verghese, P., & Coughlan, J. (2007). Where to look next? Eye movements reduce local uncertainty. *Journal of Vision*, *7*(3):6, 1–17. <http://doi.org/10.1167/7.3.6>
- Resulaj, A., Kiani, R., Wolpert, D. M., & Shadlen, M. N. (2009). Changes of mind in decision-making. *Nature*,

Bibliographie

- 461(7261), 263–266. <http://doi.org/10.1038/nature08275>
- Reynolds, J. H., & Chelazzi, L. (2004). Attentional modulation of visual processing. *Annual Review of Neuroscience*, 27, 611–647. <http://doi.org/10.1146/annurev.neuro.26.041002.131039>
- Rizzolatti, G., Riggio, L., Dascola, I., & Umiltá, C. (1987). Reorienting attention across the horizontal and vertical meridians: Evidence in favor of a premotor theory of attention. *Neuropsychologia*, 25(1), 31–40. [http://doi.org/10.1016/0028-3932\(87\)90041-8](http://doi.org/10.1016/0028-3932(87)90041-8)
- Röhrbein, F., Goddard, P., Schneider, M., James, G., & Guo, K. (2015). How does image noise affect actual and predicted human gaze allocation in assessing image quality? *Vision Research*, 112, 11–25. <http://doi.org/10.1016/j.visres.2015.03.029>
- Rosenholtz, R., Huang, J., Raj, A., Balas, B. J., & Ilie, L. (2012). A summary statistic representation in peripheral vision explains visual search. *Journal of Vision*, 12(4):14, 1–17. <http://doi.org/10.1167/12.4.14>
- Rothkopf, C. A., & Ballard, D. H. (2010). Credit assignment in multiple goal embodied visuomotor behavior. *Frontiers in Psychology*, 1:173. <http://doi.org/10.3389/fpsyg.2010.00173>
- Rothkopf, C. A., Ballard, D. H., & Hayhoe, M. M. (2016). Task and context determine where you look. *Journal of Vision*, 7(14):16, 1–20. <http://doi.org/10.1167/7.14.16>
- Rousselet, G. A., Joubert, O. R., & Fabre-Thorpe, M. (2005). How long to get to the “gist” of real-world natural scenes? *Visual Cognition*, 12(6), 852–877. <http://doi.org/10.1080/13506280444000553>
- Saffarian, M., Happee, R., & de Winter, J. C. F. (2012). Why do drivers maintain short headways in fog? A driving-simulator study evaluating feeling of risk and lateral control during automated and manual car following. *Ergonomics*, 55(9), 971–985. <http://doi.org/10.1080/00140139.2012.691993>
- Schall, J. D., Purcell, B. A., Heitz, R. P., Logan, G. D., & Palmeri, T. J. (2011). Neural mechanisms of saccade target selection: Gated accumulator model of the visual-motor cascade. *European Journal of Neuroscience*, 33(11), 1991–2002. <http://doi.org/10.1111/j.1460-9568.2011.07715.x>
- Schyns, P. G., & Oliva, A. (1994). From blobs to boundary edges: Evidence for time- and spatial-scale-dependent scene recognition. *Psychological Science*, 5(4), 195–201.
- Séré, B., Marendaz, C., & Héroult, J. (2000). Nonhomogeneous resolution of images of natural scenes. *Perception*, 29(12), 1403–1412. <http://doi.org/10.1068/p2991>
- Shadlen, M. N., & Kiani, R. (2013). Decision making as a window on cognition. *Neuron*, 80(3), 791–806. <http://doi.org/10.1016/j.neuron.2013.10.047>
- Smith, P. L., & Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. *Trends in Neurosciences*, 27(3), 161–168. <http://doi.org/10.1016/j.tins.2004.01.006>
- Song, J.-H., & Nakayama, K. (2008). Target selection in visual search as revealed by movement trajectories. *Vision Research*, 48(7), 853–861. <http://doi.org/10.1016/j.visres.2007.12.015>
- Song, J.-H., & Nakayama, K. (2009). Hidden cognitive states revealed in choice reaching tasks. *Trends in*

Bibliographie

- Cognitive Sciences*, 13(8), 360–366. <http://doi.org/10.1016/j.tics.2009.04.009>
- Spivey, M. J., Grosjean, M., & Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences*, 102(29), 10393–10398. <http://doi.org/10.1073/pnas.0503903102>
- Spotorno, S., Malcolm, G. L., & Tatler, B. W. (2014). How context information and target information guide the eyes from the first epoch of search in real-world scenes. *Journal of Vision*, 14(2):7, 1–21. <http://doi.org/10.1167/14.2.7>
- Spotorno, S., Malcolm, G. L., & Tatler, B. W. (2015). Disentangling the effects of spatial inconsistency of targets and distractors when searching in realistic scenes. *Journal of Vision*, 15(2):12, 1–21. <http://doi.org/10.1167/15.2.12>
- Sprague, N., & Ballard, D. (2003). Eye movements for reward maximization. In S. Thrun, L. Saul, & B. Schölkopf (Eds.), *Advances in neural information processing systems* (Vol. 16). MIT Press.
- Summerfield, C., & de Lange, F. P. (2014). Expectation in perceptual decision making: Neural and computational mechanisms. *Nature Reviews Neuroscience*, 15(11), 745–756. <http://doi.org/10.1038/nrn3838>
- Summerfield, C., & Egnér, T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences*, 13(9), 403–409. <http://doi.org/10.1016/j.tics.2009.06.003>
- Tatler, B. W. (2007). The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. *Journal of Vision*, 7(14):4, 1–17. <http://doi.org/10.1167/7.14.4>
- Tatler, B. W. (2009). Current understanding of eye guidance. *Visual Cognition*, 17(6–7), 777–789. <http://doi.org/10.1080/13506280902869213>
- Tatler, B. W., Hayhoe, M. M., Land, M. F., & Ballard, D. H. (2011). Eye guidance in natural vision: Reinterpreting salience. *Journal of Vision*, 11(5):5, 1–23. <http://doi.org/10.1167/11.5.5>
- Tatler, B. W., & Vincent, B. T. (2008). Systematic tendencies in scene viewing. *Journal of Eye Movement Research*, 2(2), 1–18.
- Tatler, B. W., & Vincent, B. T. (2009). The prominence of behavioural biases in eye guidance. *Visual Cognition*, 17(6–7), 1029–1054. <http://doi.org/10.1080/13506280902764539>
- Theeuwes, J. (1989). Effects of location and form cuing on the allocation of attention in the visual field. *Acta Psychologica*, 72(2), 177–192. [http://doi.org/10.1016/0001-6918\(89\)90043-7](http://doi.org/10.1016/0001-6918(89)90043-7)
- Theeuwes, J. (2013). Feature-based attention: It is all bottom-up priming. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368(1628), 20130055. <http://doi.org/10.1098/rstb.2013.0055>
- Theeuwes, J., & Van der Burg, E. (2013). Priming makes a stimulus more salient. *Journal of Vision*, 13(3):21, 1–11. <http://doi.org/10.1167/13.3.21>
- Thura, D., Beauregard-Racine, J., Fradet, C.-W., & Cisek, P. (2012). Decision making by urgency gating: Theory

Bibliographie

- and experimental support. *Journal of Neurophysiology*, 108(11), 2912–2930.
<http://doi.org/10.1152/jn.01071.2011>
- Torralba, A., Oliva, A., Castelhana, 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(4), 766–786. <http://doi.org/10.1037/0033-295X.113.4.766>
- Trapp, S., & Bar, M. (2015). Prediction, context, and competition in visual recognition. *Annals of the New York Academy of Sciences*, 1339, 190–198. <http://doi.org/10.1111/nyas.12680>
- Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12(1), 97–136. [http://doi.org/10.1016/0010-0285\(80\)90005-5](http://doi.org/10.1016/0010-0285(80)90005-5)
- Turner, B. M., Van Zandt, T., & Brown, S. (2011). A dynamic stimulus-driven model of signal detection. *Psychological Review*, 118(4), 583–613. <http://doi.org/10.1037/a0025191>
- Underwood, G., Chapman, P., Bowden, K., & Crundall, D. (2002). Visual search while driving: Skill and awareness during inspection of the scene. *Transportation Research Part F: Traffic Psychology and Behaviour*, 5(2), 87–97. [http://doi.org/10.1016/S1369-8478\(02\)00008-6](http://doi.org/10.1016/S1369-8478(02)00008-6)
- Underwood, G., & Foulsham, T. (2006). Visual saliency and semantic incongruity influence eye movements when inspecting pictures. *The Quarterly Journal of Experimental Psychology*, 59(11), 1931–1949. <http://doi.org/10.1080/17470210500416342>
- Underwood, G., Foulsham, T., & Humphrey, K. (2009). Saliency and scan patterns in the inspection of real-world scenes: Eye movements during encoding and recognition. *Visual Cognition*, 17(6–7), 812–834. <http://doi.org/10.1080/13506280902771278>
- Underwood, G., Foulsham, T., van Loon, E., Humphreys, L., & Bloyce, J. (2006). Eye movements during scene inspection: A test of the saliency map hypothesis. *European Journal of Cognitive Psychology*, 18(3), 321–342. <http://doi.org/10.1080/09541440500236661>
- Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review*, 108(3), 550–592. <http://doi.org/10.1037/0033-295X.108.3.550>
- Usher, M., Tsetsos, K., Yu, E. C., & Lagnado, D. A. (2013). Dynamics of decision-making: From evidence accumulation to preference and belief. *Frontiers in Psychology*, 4:758. <http://doi.org/10.3389/fpsyg.2013.00758>
- Van der Stigchel, S., & Theeuwes, J. (2007). The relationship between covert and overt attention in endogenous cuing. *Perception & Psychophysics*, 69(5), 719–731.
- Võ, M. L.-H., & Henderson, J. M. (2009). Does gravity matter? Effects of semantic and syntactic inconsistencies on the allocation of attention during scene perception. *Journal of Vision*, 9(3):24, 1–15. <http://doi.org/10.1167/9.3.24>
- Võ, M. L.-H., & Henderson, J. M. (2011). Object–scene inconsistencies do not capture gaze: Evidence from the flash-preview moving-window paradigm. *Attention, Perception, & Psychophysics*, 73(6), 1742–

Bibliographie

1753. <http://doi.org/10.3758/s13414-011-0150-6>
- Vö, M. L.-H., & Wolfe, J. M. (2015). The role of memory for visual search in scenes. *Annals of the New York Academy of Sciences*, 1339, 72–81. <http://doi.org/10.1111/nyas.12667>
- Voss, A., Nagler, M., & Lerche, V. (2013). Diffusion models in experimental psychology: A practical introduction. *Experimental Psychology*, 60(6), 385–402. <http://doi.org/10.1027/1618-3169/a000218>
- Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica*, 41, 67–85.
- Winkel, J., Keuken, M. C., van Maanen, L., Wagenmakers, E.-J., & Forstmann, B. U. (2014). Early evidence affects later decisions: Why evidence accumulation is required to explain response time data. *Psychonomic Bulletin & Review*, 21(3), 777–784. <http://doi.org/10.3758/s13423-013-0551-8>
- Wolfe, J. M. (1994). Guided Search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review*, 1(2), 202–238. <http://doi.org/10.3758/BF03200774>
- Wolfe, J. M. (2012). When do I quit? The search termination problem in visual search. In *Nebraska symposium on motivation* (Vol. 59, pp. 183–208). New York: Springer. http://doi.org/10.1007/978-1-4614-4794-8_8
- Wolfe, J. M. (2013). When is it time to move to the next raspberry bush? Foraging rules in human visual search. *Journal of Vision*, 13(3):10, 1–17. <http://doi.org/10.1167/13.3.10>
- Wu, C.-C., Wick, F. A., & Pomplun, M. (2014). Guidance of visual attention by semantic information in real-world scenes. *Frontiers in Psychology*, 5(FEB), 54. <http://doi.org/10.3389/fpsyg.2014.00054>
- Yarbus, A. L. (1967). *Eye movements and vision*. New York: Plenum Press.
- Zelinsky, G. J. (2008). A theory of eye movements during target acquisition. *Psychological Review*, 115(4), 787–835. <http://doi.org/10.1037/a0013118>
- Zelinsky, G. J., & Schmidt, J. (2009). An effect of referential scene constraint on search implies scene segmentation. *Visual Cognition*, 17(6–7), 1004–1028. <http://doi.org/10.1080/13506280902764315>
- Zelinsky, G. J., Zhang, W., Yu, B., Chen, X., & Samaras, D. (2006). The role of top-down and bottom-up processes in guiding eye movements during visual search. In Y. Weiss, B. Scholkopf, & J. Platt (Eds.), *Advances in neural information processing systems 18* (pp. 1569–1576). Cambridge, MA: MIT Press.
- Zhao, M., Gersch, T. M., Schnitzer, B. S., Doshier, B. A., & Kowler, E. (2012). Eye movements and attention: The role of pre-saccadic shifts of attention in perception, memory and the control of saccades. *Vision Research*, 74, 40–60. <http://doi.org/10.1016/j.visres.2012.06.017>
- Zylberberg, A., Fetsch, C. R., & Shadlen, M. N. (2016). The influence of evidence volatility on choice, reaction time and confidence in a perceptual decision. *eLife*, 5, 1–31. <http://doi.org/10.7554/eLife.17688>

Annexes

Annexe 1

Saliency in the fog : Application to real & 2.5D pictures

Poster présenté à la conférence Transport Research Arena 2014, à Paris.

Quétard, Boris

Gana, Racha

Birem, Merwan

Quinton, Jean-Charles

Colomb, Michèle

Bernardin, Frédéric

Mermillod, Martial

Izaute, Marie

Pezzulo, Giovanni



SALIENCY IN THE FOG: APPLICATION TO REAL & 2.5D PICTURES

Quétard, Boris, LabEx IMobS3, France
Birem, Merwan, Institut Pascal, France
Colomb, Michèle, CEREMA, France
Mermillod, Martial, LPNC, France
Pezzulo, Giovanni, ISTC-CNR, Italy

Gana, Racha, LabEx IMobS3, France
Quinton*, J-Charles, Institut Pascal, France
Bernardin, Frédéric, CEREMA, France
Izaute, Marie, LAPSCO, France
* j-charles.quinton@univ-bpclermont.fr



Introduction

Driving is mostly a vision-based task involving complex cognitive processing so as to adapt the environment and conditions which may change rapidly. Within the context of the *LabEx IMobS3 / PEVMA* project, we study critical situations combining:

- *Reduced visibility* (night, fog, rain, glare)
- *Rich driving environments* (obstacles, road users, distractors)

Critical targets –like pedestrians– must be detected on time to avoid accidents, but may be hardly visible in degraded conditions. Solutions for mobility in the future, including driver assistance and autonomous systems, must be robust to such conditions.

Experimental context

We aim at combining psychophysical experiments and computational developments to evaluate and model how humans deal with uncertainty in visual information, manipulating both:

- *Saliency* of the stimuli (reactive bottom-up information)
- *Driving context* (predictive top-down influences)

We then record dynamical measures of the subjects behavior:

- *Eye movements* (reflecting information gathering strategies)
- *Hand movements* (reflecting dynamical decision-making)

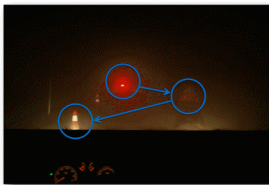


Illustration of stimuli and scanpaths (EyeTracking)

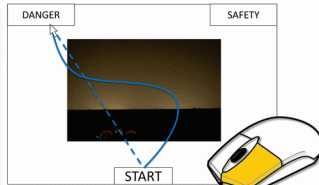


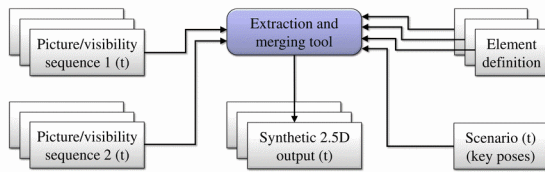
Illustration of dynamic decision-making (MouseTracking [1])

Picture database

The *CEREMA* in Clermont-Ferrand can generate fog in an up to 30m long tunnel, where various driving situations can be reproduced [2]. The density of fog can be precisely controlled and the associated visibility measured by a transmissometer.

Because of time and experimental constraints, we may want to generate synthetic pictures from real images captured in the fog by applying the following procedure:

- *Capture real pictures and record visibility* (synchronously)
- *Design static/dynamic scenarios* (e.g. with a 3D CAD software)
- For any image in the sequence to generate
 - *Extract elements* from adequate pictures (fog level, position...)
 - *Merge and adapt* these elements (resize, normalize and blend)
 - *Save* the resulting 2.5D synthetic image



Data flow diagram of the 2.5D synthetic image generation

Saliency maps

Saliency maps are the output of neuro-inspired algorithms developed to mimic the way humans extract information from their visual environment. We here limit ourselves to the bottom-up version originally introduced in [3], extracting multi-scale luminance and color contrasts, as well as oriented contours from input images. The objectives of applying such algorithm are:

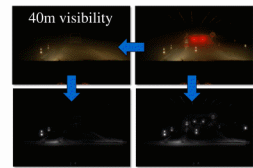
- *Applicability of saliency* for night / fog images
- *Modeling and analysis of scanpaths* (eye tracker)
- *Control of biases* in visual attention
- *Ecological validity of 2.5D images*

Results

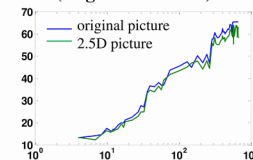
Following the analysis of saliency maps on the original and generated pictures, the following results have been obtained:

- Intensity chain* (luminance contrasts) gives the stronger response (visibility <100m) in night and fog conditions.
- Exponential attenuation of light* through a scattering atmosphere is reflected by the maximal saliency.
- Consistent results* ($\Delta < 2\%$ for original vs 2.5D picture) are obtained, despite the broad visibility range and basic merging method used.

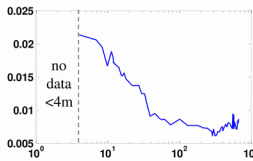
The high variability for some high visibility values may reflect the presence of eddies on the transmissometer path.



A. Saliency from pictures (original and 2.5D)



B. Maximal saliency function of visibility (m)



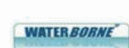
C. Mean saliency difference function of visibility (m)

Conclusion

The database and methodology are yet to be validated by psychology experiments, but multi-scale intensity-based *saliency algorithms* seem to consistently and efficiently extract information from fog pictures. Because the presence of fog simplifies the algorithmic merging process, *2.5D synthetic images* reflect identical properties as their real camera counterparts. However, objects must be clearly separated and reciprocal illumination limited in the original images to avoid discrepancies in the output pictures. Better algorithms and segmentation methods should thus be implemented to improve the results and their ecological validity.

Main references

- Freeman, J. B. & Ambady, N. (2010). MouseTracker: Software for studying real-time mental processing using a computer mouse-tracking method. *Behavior Research Methods*, 42(1), 226-241.
- Cavallo, V., Colomb, M. & Doré, J. (2001). "Distance perception of rear lights in fog", *Human Factors*, vol. 43:3, 442-45.
- Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 20(11), 1254-1259.



Annexe 2

Exemples de scènes visuelles plongées dans le brouillard, employées dans l'Étude 3-article 2

Prises de vue construites et photographiées au CEREMA, Laboratoire de Clermont-Ferrand
Avec la collaboration de Philippe Morange et Jean-Luc Bicard, techniciens au CEREMA
Et sous la supervision de Michèle Colomb et Frédéric Bernardin

Merci à M. et Mme. Mialon pour le prêt d'animaux empaillés (notamment un chevreuil).



Chevreuil – Haut : brouillard faible, panneau congruent – Bas : brouillard élevé, panneau incongruent





Chevreuil – Haut : brouillard faible, panneau incongruent – Bas : brouillard élevé, panneau neutre





Piéton – Haut : brouillard faible, panneau congruent – Bas : brouillard élevé, panneau incongruent





Piéton – Haut : brouillard faible, panneau incongruent – Bas : brouillard élevé, panneau neutre





Aucune cible – Haut : brouillard faible, panneau neutre – Bas : brouillard élevé, panneau enfants



