



HAL
open science

Etude de l'influence des processus prédictifs sur la perception visuelle subjective

Pauline Rossel

► **To cite this version:**

Pauline Rossel. Etude de l'influence des processus prédictifs sur la perception visuelle subjective. Psychologie. Université Grenoble Alpes [2020-..], 2023. Français. NNT: 2023GRALS035. tel-04415153

HAL Id: tel-04415153

<https://theses.hal.science/tel-04415153v1>

Submitted on 24 Jan 2024

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.

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

École doctorale : ISCE - Ingénierie pour la Santé la Cognition et l'Environnement

Spécialité : PCN - Sciences cognitives, psychologie et neurocognition

Unité de recherche : Laboratoire de Psychologie et Neuro Cognition

Etude de l'influence des processus prédictifs sur la perception visuelle subjective

Investigation of the influence of predictive processes on subjective visual perception

Présentée par :

Pauline ROSSEL

Direction de thèse :

Carole PEYRIN

DIRECTRICE DE RECHERCHE, CNRS DELEGATION ALPES

Directrice de thèse

Louise KAUFFMANN

MAITRE DE CONFERENCES, Université Grenoble Alpes

Co-encadrante de thèse

Rapporteurs :

JEROME SACKUR

PROFESSEUR DES UNIVERSITES, ECOLE HAUTES ETUDES SCIENCES SOC PARIS

THERESE COLLINS

PROFESSEURE DES UNIVERSITES, UNIVERSITE PARIS CITE

Thèse soutenue publiquement le **8 novembre 2023**, devant le jury composé de :

CAROLE PEYRIN

DIRECTRICE DE RECHERCHE, CNRS DELEGATION ALPES

Directrice de thèse

VALERIE GOFFAUX

PROFESSEURE, UNIVERSITE CATHOLIQUE DE LOUVAIN

Examinatrice

CHRISTOPHER MOULIN

PROFESSEUR DES UNIVERSITES, UNIVERSITE GRENOBLE ALPES

Président

JEROME SACKUR

PROFESSEUR DES UNIVERSITES, ECOLE HAUTES ETUDES SCIENCES SOC PARIS

Rapporteur

THERESE COLLINS

PROFESSEURE DES UNIVERSITES, UNIVERSITE PARIS CITE

Rapporteuse

Invités :

LOUISE KAUFFMANN

MAITRESSE DE CONFERENCES, UNIVERSITE GRENOBLE ALPES



Résumé

Les modèles actuels de la perception visuelle considèrent la perception comme un processus proactif. Dans ce contexte, percevoir consisterait à faire correspondre en permanence les caractéristiques d'un stimulus visuel à des attentes ou prédictions sur celui-ci, élaborées sur la base des expériences antérieures et des régularités apprises dans l'environnement. Cependant, les mécanismes par lesquels ces attentes (ou prédictions) influencent le traitement de l'information et façonnent in fine l'expérience subjective des stimuli visuels restent débattus. Alors que certains travaux suggèrent que la perception serait biaisée en faveur de ce qui peut être prédit et attendu dans l'environnement visuel – conduisant le plus souvent à un percept fiable de celui-ci – d'autres travaux suggèrent au contraire que la perception serait dominée par ce qui est inattendu – et donc plus informatif. L'objectif de ce travail de thèse a été de déterminer dans quelle mesure et de quelle manière la perception subjective de stimuli visuels complexes tels que les scènes et les objets est susceptible d'être modifiée en fonction de leur caractère attendu ou inattendu, ainsi que de différentes contraintes visuelles. Nous avons pour cela mené 6 expériences psychophysiques. Les trois premières expériences nous ont permis de démontrer que les prédictions influencent qualitativement la perception. Plus précisément, nous avons montré que des scènes et objets flous sont subjectivement perçus comme plus nets lorsqu'ils peuvent être prédits sur la base d'informations contextuelles disponibles dans le pattern visuel que lorsque ces informations contextuelles sont absentes, alors même que ces stimuli sont physiquement les mêmes dans les deux conditions. Les résultats de la quatrième expérience ont par ailleurs mis en évidence que ces effets perceptifs des prédictions variaient en fonction de la qualité des signaux visuels : ces effets sont d'autant plus forts que les stimuli visuels sont bruités (i.e., flous). Dans la cinquième expérience, nous avons plus spécifiquement l'objectif d'étudier de quelle manière ces effets perceptifs sont influencés par la validité des prédictions (prédictions correctes ou erronées). Nous avons tout d'abord observé que lorsque le signal visuel est bruité (par exemple, très flou), des objets attendus (présentés dans un contexte congruent conduisant à des prédictions correctes) sont perçus comme plus nets que ces mêmes objets inattendus (présentés dans un contexte incongruent conduisant à des prédictions erronées). La validité des prédictions viendrait considérablement aider la perception de stimuli bruités. Par contre, lorsque le signal visuel était peu bruité (par exemple, des objets relativement nets), les objets inattendus étaient cette fois perçus comme plus nets que les objets attendus. Dans l'Expérience 6, nous avons voulu tester si ces effets dépendaient également de contraintes temporelles mais nos résultats n'ont pas mis en évidence de modulation de l'effet des prédictions selon le temps de présentation des stimuli. Dans l'ensemble, les résultats de ces travaux permettent de préciser les mécanismes par lesquels les prédictions influencent qualitativement la perception en soulignant le rôle de la validité des prédictions (stimulus attendu vs. inattendu), ainsi que le rôle de la fiabilité du signal visuel (par exemple, le niveau de flou). Nous les avons interprétés dans le cadre de modèles théoriques récents selon lesquels les mécanismes prédictifs permettent d'optimiser le traitement de l'information visuelle de façon à construire des percepts à la fois fiables et informatifs.

Abstract

Current models of visual perception consider perception as a proactive process. Under this view, perceiving consists of continuously matching the characteristics of a visual stimulus to expectations or predictions we have about it, built upon previous perceptual experiences and learned regularities in the environment. However, the mechanisms by which these expectations (or predictions) influence information processing and ultimately shape the subjective experience of visual stimuli remain debated. While some work suggests that perception is biased towards what can be predicted and expected in the visual environment - most often leading to a veridical percept - other work suggests that perception is dominated by what is unexpected - and therefore more informative. The aim of this thesis was to determine to what extent and in what way the subjective perception of complex visual stimuli such as scenes and objects can be modified as a function of their expected or unexpected status and of different visual constraints. To this end, we conducted six psychophysical experiments. In the first three experiments, we demonstrated that predictions qualitatively influence perception. Specifically, we showed that blurred scenes and objects are subjectively perceived as sharper when they can be predicted on the basis of contextual information available in the visual pattern than when this contextual information is absent, even though these stimuli are physically the same in both conditions. The results of the fourth experiment also showed that these perceptual effects of the predictions varied according to the quality of the visual signals: the noisier (i.e., blurred) the visual stimuli were, the stronger these effects were. In the fifth experiment, we specifically studied how these perceptual effects are influenced by the validity of the predictions (correct or incorrect predictions). We first observed that when the visual signal is noisy (e.g., very blurred), expected objects (presented in a congruent context leading to correct predictions) are perceived as sharper than the same unexpected objects (presented in an incongruent context leading to incorrect predictions). The validity of the predictions would therefore significantly improve the perception of noisy stimuli. In contrast, when the visual signal was clear and unambiguous (e.g., relatively sharp objects), unexpected objects were perceived as sharper than expected objects. In Experiment 6, we tested whether these effects also depend on temporal constraints, but our results did not reveal any modulation of the effects by the presentation time of the stimuli. Overall, the results of this work clarify the mechanisms by which predictions qualitatively influence perception by highlighting the role of the validity of predictions (expected vs. unexpected stimuli), as well as the role of the reliability of the visual signal (e.g., the blur level). We interpret them in the context of recent theoretical models according to which predictive mechanisms allow optimizing visual information processing in order to construct percepts that are both veridical and informative.

Remerciements

Je remercie très sincèrement Thérèse Collins, Jérôme Sackur, Valérie Goffaux et Chris Moulin d'avoir accordé de leur temps à l'évaluation de cette thèse. Je remercie aussi Nicolas Poirel et Richard Palluel-Germain pour avoir suivi le déroulement de mes travaux de thèse par le biais du comité de suivi individuel.

Je tiens à remercier chaleureusement le laboratoire pour avoir permis la réalisation de cette thèse. Merci notamment à toutes celles et ceux qui m'ont accordé leur confiance pour que je puisse enseigner en TD et même en CM pendant ces trois années. Merci également aux étudiants avec qui j'ai pu partager des discussions passionnantes.

Mes remerciements vont évidemment à Carole Peyrin, ma directrice de thèse et Louise Kauffmann, ma co-encadrante qui a fait bien plus que me co-encadrer. Si cette thèse a vu le jour c'est en grande partie grâce à elle et sa fascination pour le codage prédictif. Merci Louise pour tout. Ta disponibilité, tes conseils, ta compréhension dans les moments de galère. Je te dois une fière chandelle pour toutes tes brillantes idées qui ont permis de mener cette thèse jusqu'ici. Je suis très heureuse d'avoir partagé ces trois années à tes côtés. Carole, merci à toi pour ton accompagnement depuis le master. Je dois aussi cette réussite à ta pédagogie et à ta rigueur qui m'ont permis de progresser pendant toutes ces années. Tu étais toujours là quand on était dans l'impasse avec Louise et tu as toujours été présente quand j'en avais besoin. Pour tout ça, mille mercis à toutes les deux !

Je n'oublie pas non plus les précieux conseils que m'ont donné Alexia Roux-Sibilon et Audrey Trouilloud et qui m'ont permis de surmonter (entre autres) les corrections rédactionnelles de Carole depuis le master, nécessaires mais plutôt douloureuses, surtout au début.

Merci à Aurélie Campagne, Raouf Zouglech et Romain Grandchamp pour leur aide sur l'expérience en EEG. Je ne m'en serais pas sortie sans votre expertise. Je souhaite également remercier Dominique Müller pour ces cours de stats si clairs et si complets pendant le master qui m'ont permis de faire plus que des ANOVA pour cette thèse.

Un grand merci aussi à toutes celles et ceux qui ont participé aux expériences de cette thèse et sans qui il y aurait eu beaucoup de théorie mais pas beaucoup de résultats. Je tiens

également à remercier les étudiants de licence et de master qui ont fait un petit bout de chemin avec moi pendant cette thèse : Merci pour votre aide Clara, Amaury et Mahé. J'espère que vous trouverez et que vous suivrez votre voie.

Evidemment, je remercie tous ceux (enfin surtout toutes celles) du bureau E120. Même si je n'y ai pas été très longtemps, c'était quand même chouette. Je veux aussi remercier tout spécialement Lise qui m'a aidée bien plus qu'elle ne peut l'imaginer. Tu as su m'écouter et trouver les mots justes. Pour ça, je ne pourrai jamais te remercier assez.

Enfin, merci à ma famille et surtout à Dylan pour ton soutien infaillible et ton amour inconditionnel. Le Turing de mon Enigma, la lumière de ma vie.

Table des Matières

| | |
|----------------------------------------------------------------------------------------------------------------------------------------------|----|
| <i>Résumé</i> | 2 |
| <i>Abstract</i> | 3 |
| <i>Remerciements</i> | 4 |
| <i>Table des Matières</i> | 6 |
| <i>Table des Figures</i> | 12 |
| <i>Abréviations</i> | 14 |
| <i>Préambule</i> | 14 |
| | |
| <i>Chapitre 1 - Contexte théorique : Influence des mécanismes prédictifs sur la perception visuelle</i> | 16 |
| 1. <i>Influence des connaissances et des attentes sur la perception visuelle</i> | 16 |
| 1.1. Les connaissances issues de l'information visuelle | 16 |
| 1.2. La perception comme processus proactif..... | 17 |
| 1.3. Arguments comportementaux en faveur de l'influence des connaissances sur la perception visuelle | 19 |
| 1.3.1. Influence des connaissances sur la reconnaissance de stimuli visuels complexes | 19 |
| 1.3.2. Influence des connaissances sur le contenu et l'apparence de la perception visuelle subjective | 21 |
| 1.3.3. Critiques et controverses théoriques | 26 |
| 2. <i>Par quels mécanismes les attentes basées sur les connaissances a priori sont-elles susceptibles d'influencer la perception ?</i> | 34 |

| | |
|---------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 2.1. Rappels sur l'organisation anatomique et fonctionnelle des voies visuelles | 35 |
| 2.1.1. De la rétine au cortex visuel | 35 |
| 2.1.2. Au-delà du cortex strié : Les voies ventrale et dorsale | 36 |
| 2.1.3. Rôle des connexions descendantes..... | 37 |
| 2.2. Modèles neurocognitifs proactifs de la perception visuelle de scènes et d'objets..... | 41 |
| 2.3. Modèles du codage prédictif | 45 |
| 2.3.1. Modèle computationnel du codage prédictif | 45 |
| 2.3.2. Corrélats cérébraux du codage prédictif : Mécanismes de Sharpening vs. Dampening | 48 |
| 2.3.3. Modulation du poids relatif des prédictions/de l'erreur de prédiction et opposing process theory..... | 55 |
| 3. <i>Objectif et problématique de la thèse</i> | 60 |
| <i>Chapitre 2 – Contribution empirique</i> | 63 |
| 1. <i>Etude de l'influence des prédictions sur la perception subjective de scènes et d'objets</i> | 63 |
| Objectif des Expériences 1, 2 et 3 | 63 |
| 1.1. Expérience 1 – Ajustement du flou lors de la perception subjective d'objets prédictibles ou non prédictibles sur la base de leur contexte | 66 |
| 1.1.1. Méthode..... | 66 |
| 1.1.1.1 Participants et Design | 66 |
| 1.1.1.2. Stimuli..... | 66 |
| 1.1.1.3. Procédure | 69 |
| 1.1.2. Analyse des données | 72 |
| 1.1.3. Résultats | 73 |
| 1.1.4. Discussion de l'Expérience 1 | 77 |
| 1.2. Expérience 2 – Ajustement du flou lors de la perception subjective de scènes prédictibles et non prédictibles | 79 |

| | |
|-------------------------------------------------------------------------------------------------------------------------------------|-----|
| 1.2.1. Méthode..... | 79 |
| 1.2.1.1. Participants et Design | 79 |
| 1.2.1.2. Stimuli..... | 79 |
| 1.2.1.3. Procédure | 80 |
| 1.2.2. Analyse des données | 83 |
| 1.2.3. Résultats | 83 |
| 1.2.4. Discussion de l'Expérience 2 | 87 |
| 1.3. Expérience 3 – Comparaison explicite du flou lors de la perception subjective de scènes prédictibles et non prédictibles | 89 |
| 1.3.1. Méthode..... | 89 |
| 1.3.1.1. Participants et Design | 89 |
| 1.3.1.2. Stimuli et Procédure | 90 |
| 1.3.2. Analyse des données | 93 |
| 1.3.2.1. Données comportementales | 93 |
| 1.3.2.2. Données des mouvements oculaires | 95 |
| 1.3.3. Résultats | 96 |
| 1.3.3.1. Résultats comportementaux..... | 96 |
| 1.3.3.2. Résultats des mouvements oculaires | 99 |
| 1.3.4. Discussion de l'Expérience 3 | 99 |
| 2. <i>Facteurs influençant l'effet des prédictions sur la perception subjective</i> | 102 |
| 2.2. Expérience 4 – Etude de l'effet des prédictions en fonction de la fiabilité du signal | 102 |
| 2.2.1. Méthode..... | 103 |
| 2.2.1.1. Participants et Design | 103 |
| 2.2.1.2. Stimuli..... | 104 |
| 2.2.1.3. Procédure | 105 |
| 2.2.2. Analyse des données | 105 |

| | |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| 2.2.3. Résultats | 106 |
| 2.2.4. Discussion de l'Expérience 4 | 108 |
| 2.3. Expérience 5 – Etude de l'effet des prédictions en fonction de la validité des prédictions et de la fiabilité du signal..... | 110 |
| Objectif des Expériences 5 et 6 | 110 |
| 2.3.1. Méthode..... | 112 |
| 2.3.1.1. Participants et Design | 112 |
| 2.3.1.2. Stimuli..... | 112 |
| 2.3.1.3. Procédure | 116 |
| 2.3.2. Analyse des données | 119 |
| 2.3.3. Résultats | 121 |
| 2.3.4. Discussion de l'Expérience 5 | 124 |
| 2.4. Expérience 6 – Etude de l'effet des prédictions en fonction de la validité des prédictions et de contraintes temporelles sur le traitement du signal visuel..... | 128 |
| 2.4.1. Méthode..... | 130 |
| 2.4.1.1. Participants et Design | 130 |
| 2.4.1.2. Stimuli..... | 130 |
| 2.4.1.3. Procédure | 132 |
| 2.4.2. Analyse des données | 134 |
| 2.4.3. Résultats | 136 |
| 2.4.4. Discussion de l'Expérience 6 | 138 |
| | |
| <i>Chapitre 3 – Discussion générale et perspectives</i> | 140 |
| 1. <i>Apports des résultats aux questions théoriques et limites</i> | 140 |
| 1.1. Dans quelle mesure ce que nous savons influence-t-il ce que nous voyons ? | 140 |
| 1.1.1. Synthèse et apport des résultats..... | 140 |
| 1.1.2. Limites et perspectives | 143 |

| | |
|---------------------------------------------------------------------------------------------------------------------------------|-----|
| 1.2. Modulation des effets des prédictions sur la perception subjective..... | 148 |
| 1.2.1. Synthèse et apport des résultats..... | 148 |
| 1.2.2. Limites et perspectives | 153 |
| 2. <i>Perspectives théoriques et implications cliniques</i> | 161 |
| 2.1. Application des effets perceptifs des prédictions à la vision active et leurs conséquences pour la reconnaissance | 161 |
| 2.2. Implications cliniques | 166 |
| 3. <i>Conclusion</i> | 167 |
| <i>ANNEXE A – Article Rossel et al. (2022)</i> | 169 |
| <i>ANNEXE B - Prétest sur la linéarité de la perception subjective de flou</i> | 190 |
| Méthode..... | 190 |
| Participants et Design | 190 |
| Stimuli | 191 |
| Procédure | 192 |
| Analyse des données | 193 |
| Résultats | 193 |
| <i>ANNEXE C – Post-test sur la congruence sémantique et physique des stimuli de l'Expérience 5</i> | 195 |
| Méthode..... | 195 |
| Participants et Design | 195 |
| Stimuli | 196 |
| Procédure | 196 |
| Analyse des données | 196 |

| | |
|------------------------------------------------------|-----|
| Résultats | 199 |
| <i>ANNEXE D – Article Rossel et al. (2023)</i> | 201 |
| <i>Bibliographie</i> | 211 |

Table des Figures

| | |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Figure 1.1 <i>Exemple d'Image Mooney.</i> | 18 |
| Figure 1.2 <i>Exemple de Stimuli Utilisés par Teufel et al. (2018).</i> | 23 |
| Figure 1.3 <i>Exemples de Stimuli Utilisés Pour Etudier l'Influence des Connaissances sur l'Apparence des Stimuli.</i> | 25 |
| Figure 1.4 <i>Paradigme d'Ajustement Perceptif de Lupyan (2017).</i> | 31 |
| Figure 1.5 <i>Paradigme de Spaak et al. (2022).</i> | 33 |
| Figure 1.6 <i>Représentation Schématique des Caractéristiques Extraites Dans les Aires Cérébrales le Long des Voies Ventrale et Dorsale.</i> | 37 |
| Figure 1.7 <i>Latences de Réponse de Neurones Dans Différentes Aires Corticales.</i> | 39 |
| Figure 1.8 <i>Modèle Intégré du Traitement de l'Information Visuelle de Bullier (2001).</i> | 40 |
| Figure 1.9 <i>Schéma du Modèle de Bar.</i> | 42 |
| Figure 1.10 <i>Illustration du Modèle Cortical de la Catégorisation de Scènes Basé sur les Fréquences Spatiales.</i> | 43 |
| Figure 1.11 <i>Illustration du Modèle Fréquentiel de la Catégorisation de Scènes.</i> | 44 |
| Figure 1.12 <i>Illustration du Fonctionnement des Modèles du Codage Prédicatif.</i> | 46 |
| Figure 1.13 <i>Stimuli et Résultats de l'Expérience de Murray et al. (2002).</i> | 50 |
| Figure 1.14 <i>Illustration des Mécanismes de Dampening et de Sharpening au Niveau Cérébral.</i> | 51 |
| Figure 1.15 <i>Illustration de l'Influence des Prédications au Niveau Perceptif.</i> | 58 |
| Figure 2.1 <i>Stimuli des Expériences 1, 2 et 3.</i> | 64 |
| Figure 2.2 <i>Exemples d'un Objet à Différents Niveaux de Flou.</i> | 67 |
| Figure 2.3 <i>Illustration des Conditions Expérimentales et du Déroulement de l'Expérience 1.</i> | 71 |
| Figure 2.4 <i>Matched Blur Level Moyen pour Chaque Condition de l'Expérience 1.</i> | 74 |
| Figure 2.5 <i>Matched Blur Level Moyen pour les Paires Target-Sample Différentes et Identiques en Fonction du Niveau de Flou de l'Objet en Target.</i> | 76 |
| Figure 2.6 <i>Exemples d'une Scène à Différents Niveaux de Flou.</i> | 80 |

| | |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Figure 2.7 <i>Illustration des Conditions Expérimentales et du Déroulement de l'Expérience 2.</i> | 82 |
| Figure 2.8 <i>Matched Blur Level Moyen pour Chaque Condition de l'Expérience 2.</i> | 84 |
| Figure 2.9 <i>Matched Blur Level Moyen pour les Paires Target-Sample Différentes et Identiques en Fonction du Niveau de Flou de la Scène en Target.</i> | 86 |
| Figure 2.10 <i>Illustration des Conditions Expérimentales et du Déroulement de l'Expérience 3.</i> | 92 |
| Figure 2.11 <i>Proportion Moyenne de Scènes Jugées Comme Etant les Plus Floues Selon le Niveau de Flou Pour les Essais Test et Taux de Réponses Correctes Moyens Selon le Type de Paire et l'Orientaion de la Scène Objectivement la Plus Floue Pour les Essais Catch.</i> | 98 |
| Figure 2.12 <i>Exemple d'une Scène Orientée à l'Endroit et à l'Envers à Différents Niveaux de Flou.</i> | 104 |
| Figure 2.13 <i>Matched Blur Level Moyen pour Chaque Condition de l'Expérience 4.</i> | 107 |
| Figure 2.14 <i>Illustration des Conditions Expérimentales et du Déroulement de l'Expérience 5.</i> | 114 |
| Figure 2.15 <i>Illustration du Niveau de Flou des Stimuli dans les Différentes Conditions de Fiabilité.</i> | 118 |
| Figure 2.16 <i>Matched Blur Level Moyen pour Chaque Condition de l'Expérience 5.</i> | 123 |
| Figure 2.17 <i>Exemple de Stimulus de l'Expérience 5.</i> | 128 |
| Figure 2.18 <i>Exemple de Paires d'Objets Présentées en Conditions Congruente et Incongruente dans l'Expérience 6.</i> | 130 |
| Figure 2.19 <i>Procédure Expérimentale de l'Expérience 6.</i> | 134 |
| Figure 2.20 <i>Taux de Réponses Correctes et Temps de Réponse Correcte Pour Chaque Condition de l'Expérience 6.</i> | 137 |
| Figure 3.1 <i>Exemple des Stimuli de Williams et al. (2022).</i> | 145 |
| Figure 3.2 <i>Exemple d'Image Morph Animal-Meuble en Contexte.</i> | 147 |
| Figure 3.3 <i>Illustration du Déroulement de l'Expérience.</i> | 155 |
| Figure 3.4 <i>Illustration des Conditions Expérimentales et du Déroulement de l'Expérience EEG.</i> | 158 |
| Figure 3.5 <i>Illustration des Hypothèses Dérivées des Mécanismes de Sharpening et Dampening sur la Distinction de Deux Catégories d'Objets.</i> | 160 |

Abréviations

ANR Agence Nationale de la Recherche

BFS Basses Fréquences Spatiales

CC0 licence Creative Common Zero

EDISCE Ecole Doctorale d'Ingénierie pour la Santé, la Cognition et l'Environnement

EEG Electroencéphalographie

EP unité d'Erreur de Prédiction

HFS Hautes Fréquences Spatiales

IRMf Imagerie par Résonance Magnétique fonctionnelle

LOC Lateral Occipital Complex

LPNC Laboratoire de Psychologie et Neurocognition

MBL Matched Blur Level

MEG Magnétoencéphalographie

MT middle temporal visual area

MVPA Multi-voxel Pattern Analysis

P unité de Prédiction

RMS Root Mean Square

TMS Transcranial Magnetic Stimulation

V1 aire visuelle primaire

V2 aire visuelle secondaire

Préambule

Les travaux présentés dans cette thèse ont été réalisés au sein du Laboratoire de Psychologie et Neurocognition (LPNC, CNRS UMR 5105) et de l'Ecole Doctorale d'Ingénierie pour la Santé, la Cognition et l'Environnement (EDISCE), grâce à un financement du Ministère de la Recherche. Ce travail de thèse s'inscrit dans les thématiques de recherche de l'équipe Vision & Emotion du LPNC qui s'intéresse aux modèles de la perception visuelle sous tous ses aspects (traitement rétinien, vision active et mouvements oculaires, vision proactive, traitement émotionnel).

Dans ce contexte, ce travail de thèse s'est attaché à déterminer dans quelle mesure et de quelle manière les connaissances que nous avons sur notre environnement, et les attentes et prédictions qui en découlent, peuvent influencer la perception subjective de stimuli visuels complexes tels que les scènes et les objets.

Chapitre 1 - Contexte théorique : Influence des mécanismes prédictifs sur la perception visuelle

1. Influence des connaissances et des attentes sur la perception visuelle

1.1. Les connaissances issues de l'information visuelle

Notre environnement visuel est extrêmement riche et complexe. La simple perception visuelle d'objets dans une scène est soumise à la variabilité infinie de leur position dans l'espace, mais aussi des points de vue possibles de l'observateur. A cela s'ajoute le masquage des objets entre eux, la projection de leurs ombres et les conditions de luminosité elles aussi variables. Il en résulte la possibilité d'une multitude d'enchevêtrements complexes de formes, de textures et de couleurs pour un seul objet. Malgré cela, notre environnement visuel comporte des caractéristiques qui se répètent fréquemment. On parle de régularités statistiques. Ces régularités sont liées d'une part aux contraintes physiques stables de l'environnement, comme le fait que la lumière vient généralement du haut, que les objets illuminés projettent donc des ombres le plus souvent vers le bas ou que les orientations cardinales (i.e., horizontales et verticales) sont plus fréquentes que les orientations obliques. D'autre part, il existe des régularités contextuelles, spécifiques à une situation ou un environnement donné. Par exemple, les environnements ouverts (e.g., plages, champs) sont caractérisés par la prédominance d'orientations horizontales tandis que les environnements clos (e.g., chambre, bureau) sont caractérisés par la prédominance d'orientations horizontales et verticales (Oliva & Torralba, 2001; Torralba & Oliva, 1999, 2003). Le contexte de la scène rend également plus probable la présence de certains objets ainsi que leur localisation. Il est par exemple plus fréquent de trouver un pot de confiture dans une cuisine que dans un champ, et sur une table que sur le sol (Bar, 2004; Henderson, 2017; Oliva & Torralba, 2007). Il existe également des régularités entre les informations visuelles et celles provenant d'autres modalités sensorielles ou de nos propres actions (de Lange et al., 2018; Seriès & Seitz, 2013). Un flux de parole est généralement associé à des lèvres en mouvement, le son d'une sirène à la présence d'un camion de pompier et le fait de lâcher un objet à la vision du mouvement de sa chute.

Chapitre 1 - Contexte théorique : Influence des mécanismes prédictifs sur la perception visuelle

L'extraction automatique et implicite des régularités statistiques de l'environnement à travers les différents sens, et en particulier la vision, permettrait à l'être humain de construire une représentation abstraite du monde qui l'entoure et de son organisation, mais aussi d'acquérir des connaissances physiques (e.g., savoir qu'un objet qui n'est pas maintenu par une surface solide tombera vers le bas) et sémantiques (e.g., savoir qu'il est plus probable de trouver un réfrigérateur dans une cuisine que dans une chambre). Celles-ci sont acquises au fur et à mesure des expériences sensorielles propres à chaque individu, mais elles pourraient aussi s'ancrer plus largement dans la phylogenèse du système visuel humain. Par exemple, le cortex visuel humain contient plus de neurones sélectifs aux orientations cardinales, prédominantes dans l'environnement par rapport aux orientations obliques (Girshick et al., 2011). L'élaboration de connaissances sur la base de régularités statistiques peut aussi se produire de manière très rapide et temporaire (Chalk et al., 2010; Denison et al., 2016; Richter et al., 2018). Par exemple, l'exposition relativement courte à la répétition d'une même succession d'images de différentes catégories sémantiques entraîne implicitement l'acquisition de connaissances sur leur ordre d'apparition.

Ces connaissances issues des régularités statistiques de l'environnement, nous permettent en retour de l'appréhender de manière proactive en générant des attentes, ou prédictions, sur ses caractéristiques les plus probables. Si nous sommes dans une rue, nous nous attendons par exemple à trouver des piétons sur le trottoir et des voitures sur la route. De nombreux travaux suggèrent que les attentes ou prédictions générées sur la base des connaissances acquises à travers les expériences sensorielles dans l'environnement orientent également sa perception (Bar, 2003; Clark, 2013; de Lange et al., 2018; Press et al., 2020; Summerfield & Egner, 2009).

1.2. La perception comme processus proactif

L'idée que les connaissances, et les attentes qui en résultent, guident la perception visuelle est loin d'être nouvelle. En 1867, von Helmholtz proposait par exemple que la perception, quelle que soit l'entrée sensorielle, est un processus « d'inférence inconsciente » qui consiste à inférer en permanence les causes les plus probables des entrées sensorielles, sur la base des expériences passées et des régularités apprises dans l'environnement. La plupart des modèles actuels de la perception visuelle considèrent également la perception comme un processus proactif, où le traitement de l'information visuelle serait moins lié aux

Chapitre 1 - Contexte théorique : Influence des mécanismes prédictifs sur la perception visuelle

caractéristiques extrinsèques des stimuli visuels, qu'au fait qu'elles correspondent plus ou moins à des attentes ou prédictions que nous formons à leur égard (Bar, 2003; Bar et al., 2006; Clark, 2013; de Lange et al., 2018; Press et al., 2020; Summerfield & Egner, 2009). Le recours à des connaissances a priori pour orienter le traitement de l'information visuelle présente plusieurs avantages. Tout d'abord, cela permettrait de résoudre les situations où les signaux visuels sont bruités ou ambigus. Par exemple, lorsque nous sommes dans le brouillard ou sous une pluie battante, l'environnement paraît plus flou, les contours des objets sont moins définis et leur reconnaissance est donc incertaine. Bien souvent également, les objets sont partiellement cachés par d'autres objets. Dans ce cas, comment savoir que ce bout de fourrure qui dépasse du canapé est en fait la queue de votre chat si vous ne faites pas appel à vos connaissances qui indiquent que votre chat a l'habitude d'y faire la sieste ? Le recours aux connaissances permettrait donc de reconstituer les informations manquantes ou de donner un sens à un stimulus ambigu selon ce qui est le plus probable. Cela s'illustre bien avec l'exemple suivant.

Que voyez-vous dans l'image de la Figure 1.1 ci-dessous ? A première vue et sans connaissances préalables, on ne voit qu'un enchevêtrement de patchs noirs et blancs sans signification. Regardez à présent l'image originale à partir de laquelle l'image noir et blanc a été construite (cf. Figure 1.2) puis revenez à cette image. Elle prend à présent tout son sens et l'on peut aisément y distinguer les objets, et même les traits qui la composent. Cet exemple illustre bien comment la perception d'un même stimulus visuel ambigu peut varier en fonction des connaissances et donc des attentes que nous avons a priori.

Figure 1.1

Exemple d'Image Mooney.



Note. Figure adaptée de Teufel et al. (2018).

Chapitre 1 - Contexte théorique : Influence des mécanismes prédictifs sur la perception visuelle

Lorsque les signaux visuels sont clairs et non ambigus, le recours aux connaissances pourrait aussi permettre d'anticiper le traitement des stimuli les plus probables dans un contexte donné et ainsi en faciliter la reconnaissance lorsqu'ils y sont présents. D'un point de vue computationnel, le recours à nos connaissances serait d'ailleurs plus économique en termes de coûts de traitement dans la mesure où cela permet d'orienter le traitement des informations sensorielles sur la base de ce que nous connaissons plutôt que de traiter toute information comme nouvelle et ce, au risque de les traiter de manière erronée (Kok & de Lange, 2015; Millidge et al., 2021).

Si l'idée que nous nous appuyons sur nos connaissances pour faciliter le traitement de l'information visuelle est relativement admise dans le domaine de la cognition visuelle, la question de savoir dans quelle mesure « ce que nous savons » influence « ce que nous voyons » continue néanmoins de faire l'objet de nombreux débats. Cette question s'inscrit plus largement dans le contexte du débat de longue date concernant la « pénétrabilité cognitive » de la perception (Firestone & Scholl, 2016; Lupyan, 2015; O'Callaghan et al., 2017; Pylyshyn, 1999; Teufel & Nanay, 2017). La perception est-elle un processus isolé et hermétique à l'influence de processus cognitifs (e.g., connaissances, états mentaux, émotionnels ou motivationnels) ou bien peut-elle y être directement soumise ? Ces débats se heurtent à plusieurs problèmes allant de la définition même de la perception et de sa distinction avec la cognition, en passant par la question de savoir si la perception doit être étudiée à partir de mesures objectives lors de tâches perceptives ou sur la base de reports subjectifs (e.g., Firestone & Scholl, 2016; Lupyan, 2015; O'Callaghan et al., 2017; Pylyshyn, 1999; Teufel & Nanay, 2017). Dans les parties suivantes, nous présenterons les arguments expérimentaux en faveur d'une influence des connaissances sur la perception visuelle – telle qu'elle est envisagée par les différents auteurs de ces études – ainsi que leurs limites et les critiques qui ont pu y être apportées dans le cadre de ces débats.

1.3. Arguments comportementaux en faveur de l'influence des connaissances sur la perception visuelle

1.3.1. Influence des connaissances sur la reconnaissance de stimuli visuels complexes

Une grande partie des arguments expérimentaux en faveur d'une influence des connaissances sur la perception visuelle provient d'études portant sur la perception de stimuli visuels complexes tels que les scènes et les objets. Dans ces études, la perception visuelle est généralement définie comme l'ensemble des processus par lesquels le système visuel

Chapitre 1 - Contexte théorique : Influence des mécanismes prédictifs sur la perception visuelle

transforme rapidement les signaux visuels en une représentation qui puisse être reliée à des connaissances et interprétée. La perception est donc plutôt étudiée sous l'angle de la reconnaissance visuelle à travers des mesures objectives quantitatives telles que le temps nécessaire pour reconnaître un stimulus visuel présenté très rapidement (de l'ordre d'une centaine de millisecondes) et/ou l'exactitude des réponses. Ces études se sont particulièrement intéressées au rôle des connaissances acquises sur la base de régularités dans les relations entre les scènes et les objets qu'elles contiennent, ou associations contextuelles, et sur la façon dont elles influencent leur reconnaissance (e.g., Biederman et al., 1982; Brandman & Peelen, 2017; Davenport, 2007; Davenport & Potter, 2004; Greene et al., 2015; Joubert et al., 2007, 2008; Lai et al., 2020; Palmer, 1975; Rémy et al., 2013; Roux-Sibilon et al., 2019, voir pour revues Oliva & Torralba, 2007; Panichello et al., 2013). Par exemple, Davenport et Potter (2004) ont présenté à des participants des images en couleur de scènes contenant un arrière-plan et un objet au premier plan qui pouvaient être fréquemment associés dans l'environnement (congruents, e.g., un prêtre dans une église) ou rarement associés (incongruents, e.g., un prêtre sur un terrain de football) pendant 80 ms. La tâche de reconnaissance visuelle qui était demandée aux participants consistait soit à nommer l'objet présenté au premier plan dans une première expérience, soit à nommer la scène présentée à l'arrière-plan dans une deuxième expérience. Les résultats ont montré qu'un objet était mieux reconnu en présence d'une scène congruente qu'incongruente. De même, une scène était mieux reconnue en présence d'un objet congruent qu'incongruent. De façon similaire, Joubert et al. (2007, 2008) ont montré que des scènes et des objets présentés simultanément pendant seulement 26 ms étaient catégorisés plus rapidement lorsqu'ils étaient congruents qu'incongruents sémantiquement. Ces études suggèrent donc que les connaissances issues des associations contextuelles rencontrées dans l'environnement (condition congruente) facilitent la reconnaissance visuelle. Plusieurs études suggèrent également que ces connaissances facilitent la perception de stimuli bruités ou ambigus (Brandman & Peelen, 2017; Greene et al., 2015; Lai et al., 2020; Roux-Sibilon et al., 2019). Par exemple, Brandman et Peelen (2017) ont montré que les performances d'identification d'objets dégradés (e.g., identifier un objet pixelisé comme étant un mouton) et de catégorisation d'objets dégradés (e.g., catégoriser un mouton pixelisé comme étant animé ou inanimé) présentés pendant seulement 50 ms étaient meilleures lorsque ceux-ci étaient présentés avec un contexte informatif et congruent avec l'objet (un champ) que sans contexte. Des effets similaires ont été observés dans des études manipulant les connaissances a priori sur les objets en présentant en amont des indices (e.g., un mot) sémantiquement congruents ou non avec

Chapitre 1 - Contexte théorique : Influence des mécanismes prédictifs sur la perception visuelle

l'objet (Eger et al., 2007; Esterman & Yantis, 2010; Pinto et al., 2015; Stein & Peelen, 2015). Dans l'ensemble, les résultats de ces études suggèrent que les connaissances que nous avons sur les associations contextuelles entre les scènes et les objets, et les attentes qui en découlent, influencent la rapidité et l'exactitude de la reconnaissance visuelle en faveur des stimuli conformes à ces connaissances, et ainsi la facilité ou l'efficacité avec laquelle nous les voyons.

Cependant, pour certains auteurs, ces études ne seraient pas révélatrices de l'influence des connaissances sur la perception visuelle car ces auteurs considèrent que la reconnaissance visuelle est un processus cognitif distinct de la perception visuelle (Firestone & Scholl, 2015; Panichello et al., 2013; Spaak et al., 2022). Dès lors, les temps de réponse ou l'exactitude des réponses lors de ces tâches de reconnaissance ne seraient pas des mesures fiables de la perception puisqu'elles impliqueraient en plus des processus qu'ils considèrent comme non perceptifs, tels que des processus sémantiques ou mnésiques. Les meilleures performances observées pour les stimuli congruents qu'incongruents dans ces tâches de reconnaissance visuelle pourraient alors s'expliquer par une influence des attentes sur ces processus sémantiques (e.g., en facilitant l'accès au nom ou l'accès à la catégorie) ou mnésiques (e.g., en facilitant l'accès aux représentations d'objets en mémoire) plutôt que sur la perception. Pour ces auteurs, la perception devrait donc plutôt être étudiée sur la base de reports subjectifs concernant l'apparence des stimuli visuels qui serait un reflet plus fidèle de la phénoménologie de la perception (Firestone & Scholl, 2016; Teufel & Nanay, 2017).

1.3.2. Influence des connaissances sur le contenu et l'apparence de la perception visuelle subjective

Quelques études se sont intéressées à l'influence des connaissances sur la perception visuelle définie d'un point de vue phénoménologique, c'est à dire en tant qu'expérience consciente subjective de l'information visuelle. Ces études s'appuient donc sur des tâches où l'on demande à des participants de rapporter soit le contenu de leur perception subjective (« ce que je vois »), soit son apparence (« comment je le vois »), sans que cela ne nécessite de reconnaître ou de catégoriser les stimuli.

Par exemple, Denison et al. (2016) se sont intéressés à l'influence des connaissances sur le contenu de la perception subjective (« ce que je vois »). Ils ont pour cela mené une expérience en utilisant un paradigme de rivalité binoculaire au cours duquel deux images différentes étaient présentées à chaque œil des participants. La rivalité binoculaire induisait ainsi la perception en alternance de chaque image de façon non contrôlable volontairement par les participants. Les

Chapitre 1 - Contexte théorique : Influence des mécanismes prédictifs sur la perception visuelle

participants réalisaient d'abord une phase d'apprentissage sans rivalité binoculaire, où ils voyaient des séquences composées de trois images de catégories différentes (visage, objet et scène). Ces images étaient présentées suivant deux ordres distincts : (1) visage féminin, puis scène d'extérieur, puis objet animé ou (2) objet inanimé, puis visage masculin, puis scène d'intérieur. Cette phase permettait aux participants d'apprendre implicitement l'ordre de succession des images pour chaque séquence. Les participants réalisaient ensuite la phase test : les deux premières images d'une séquence étaient présentées aux participants et la dernière image était présentée en rivalité binoculaire. La dernière image de la séquence apprise (image attendue) était présentée à un œil tandis qu'une image ne correspondant pas à la dernière image de la séquence apprise (image inattendue) était présentée à l'autre œil. La tâche des participants consistait alors à reporter quelle image (ou contenu) avait été perçue de façon dominante en rivalité binoculaire. Les résultats ont montré que les participants percevaient de façon prédominante les images inattendues plutôt que les images attendues. Ces résultats suggèrent donc que les attentes générées sur la base d'un apprentissage statistique déterminent le contenu de la perception subjective consciente (« ce que l'on voit »), en faveur de ce qui est inattendu. Même si ce résultat va à l'encontre de ce qui avait été observé dans les études sur la reconnaissance de scènes et d'objets, à savoir une perception biaisée en faveur de ce qui est conforme aux attentes, il démontre clairement que nos attentes influencent notre perception subjective.

D'autres études se sont plus focalisées sur la façon dont les connaissances influencent l'apparence de la perception subjective (« comment je vois »). Dans ces études, la perception subjective peut être mesurée via le seuil de sensibilité ou de discrimination de différentes caractéristiques dites « de bas niveau » telles que l'orientation ou le contraste de stimuli visuels simples (Cheadle et al., 2015; Christensen et al., 2015; Kok, Jehee, et al., 2012; Teufel et al., 2018; Wyart et al., 2012). Par exemple, Kok, Jehee, et al. (2012) présentaient à des participants un son court qui pouvait être grave ou aigu et qui était suivi de deux patches de Gabor (stimuli de forme circulaire correspondant à une alternance de bandes sombres et claires) présentés l'un après l'autre. Les participants devaient indiquer si le second patch de Gabor était orienté dans le sens horaire ou anti-horaire relativement au premier patch. Les auteurs manipulaient les régularités statistiques entre le son et l'orientation des deux patches de Gabor. Par exemple, un son grave (aigu) était suivi dans 75% des essais de deux patches orientés dans le sens horaire (vs. anti-horaire). Cela conduisait les participants à créer implicitement des connaissances sur les associations entre les sons et l'orientation des patches de Gabor et ainsi, à générer des attentes

Chapitre 1 - Contexte théorique : Influence des mécanismes prédictifs sur la perception visuelle

ou prédictions sur l'orientation de ceux-ci en fonction de l'indice sonore. Les auteurs ont observé que les participants arrivaient mieux à discriminer l'orientation du second patch par rapport au premier lorsque les deux patches étaient précédés d'un son prédisant correctement leur orientation. Ces résultats suggèrent que les connaissances sur l'orientation des patches améliorent la perception de différences fines entre les deux. De façon similaire, Teufel et al. (2018) ont demandé à des participants de discriminer l'orientation de deux patches de Gabor (dont l'orientation différaient de quelques degrés) présentés successivement et intégrés dans des images en noir et blanc sans signification (Figure 1.2 ; comme celle présentée dans la Figure 1.1 en début d'Introduction). Ces images en noir et blanc étaient en fait construites à partir de photographies réelles de scènes visuelles (Figure 1.2) et l'orientation du premier patch était soit compatible (aligné) avec les contours des objets présents dans ces photographies (qui n'étaient pas présents dans les images en noir et blanc), soit orthogonale à ces contours. Les auteurs ont ensuite montré aux participants les photographies originales qui avaient été utilisées pour construire les images en noir et blanc, puis ils ont demandé aux participants de réaliser à nouveau la tâche de discrimination. Les auteurs ont observé que les performances de discrimination sur les patches étaient meilleures après la présentation des photographies originales lorsque l'orientation des patches de Gabor correspondait à celle des contours des objets des images originales que lorsqu'elle était orthogonale. Ces résultats suggèrent que les connaissances sur les contours de la scène acquises après la présentation des photographies originales a permis aux participants de générer des attentes sur les contours invisibles dans les images en noir et blanc, facilitant alors la perception des patches de Gabor dont l'orientation était conforme aux attentes. Cette fois-ci, c'est donc la perception de ce qui est conforme aux attentes qui est réhaussée.

Figure 1.2

Exemple de Stimuli Utilisés par Teufel et al. (2018).



Chapitre 1 - Contexte théorique : Influence des mécanismes prédictifs sur la perception visuelle

Note. A gauche, l'image originale présentée aux participants avant la seconde réalisation de la tâche. A droite, l'image en noir et blanc présentée lors de la première et de la seconde réalisation de la tâche. Les ellipses rouges correspondent à l'emplacement de contours d'objets qui étaient présents dans l'image originale mais pas dans l'image en noir et blanc et le long desquels pouvaient se situer les patches de Gabor. Figure adaptée de Teufel et al. (2018).

D'autres études rapportent une influence des connaissances sur la perception subjective du mouvement (Chalk et al., 2010), du contraste (Han & VanRullen, 2016) ou encore de la netteté de stimuli visuels (Lupyan, 2017; Perez et al., 2020). Ces études s'appuient sur des jugements comparatifs subjectifs portant sur ces différentes caractéristiques. Par exemple, Han et VanRullen (2016) ont réalisé une expérience basée sur le jugement de contraste. A chaque essai, ils présentaient côte à côte deux disques gris contenant chacun un dessin au trait noir et ils demandaient aux participants de juger quel dessin était le plus contrasté par rapport au fond gris du disque (Figure 1.3.A). Les auteurs manipulaient la possibilité de générer des attentes concernant les dessins au trait. Ainsi, un de ces deux dessins au trait représentait un objet 3D pouvant facilement être associé à un objet connu tandis que l'autre dessin représentait les mêmes traits mais arrangés de manière aléatoire, de sorte qu'aucune association à un objet connu n'était vraiment possible. Les résultats ont montré que lorsque le contraste entre ces deux dessins était objectivement égal, les dessins d'objets 3D étaient subjectivement perçus comme plus contrastés que les figures de traits aléatoires. La possibilité de générer des attentes sur les stimuli augmenterait donc leur contraste perçu.

Dans la même logique, Perez et al. (2020) se sont intéressés à la perception du flou d'objets. Ils ont demandé à des participants d'effectuer des jugements comparatifs ou d'égalité sur le flou perçu des contours de silhouettes d'objets familiers (e.g., une lampe) et d'objets nouveaux/inconnus (Figure 1.3.B). Les objets nouveaux étaient construits à partir des mêmes formes de base que celles qui composaient les objets familiers, mais assemblées dans une configuration différente. Les auteurs ont constaté qu'à niveau égal de flou, les participants jugeaient les contours des objets familiers comme moins flous que ceux des nouveaux objets. Ainsi, la possibilité de générer des attentes sur un stimulus visuel améliore la perception subjective de sa netteté.

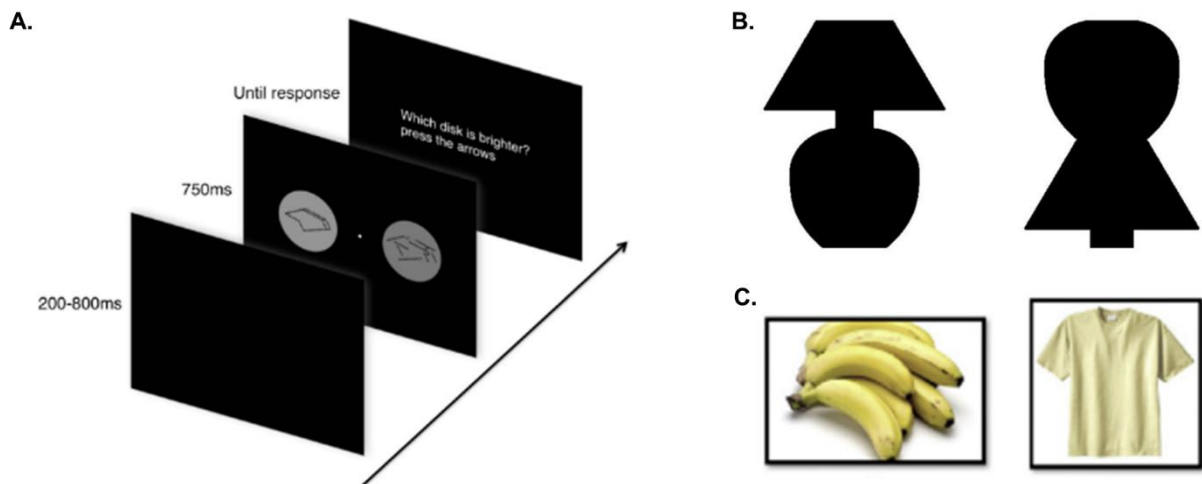
Certains travaux suggèrent également une influence des connaissances sur la perception de la couleur en utilisant cette fois-ci des tâches d'ajustement perceptif. Par exemple, Hansen

Chapitre 1 - Contexte théorique : Influence des mécanismes prédictifs sur la perception visuelle

et al. (2006 ; voir également Olkkonen et al., 2008; Witzel et al., 2011) ont mené une étude basée sur l'ajustement de la couleur de stimuli visuels. Ces stimuli étaient soit des objets fortement associés à une couleur (e.g., des bananes jaunes), soit des disques colorés non sémantiquement associés à une couleur. Les participants devaient ajuster la couleur des stimuli à l'aide d'autres couleurs jusqu'à ce qu'ils les perçoivent subjectivement comme gris. Les auteurs ont observé que les participants ajustaient les objets colorés dans le sens de leur couleur complémentaire pour les percevoir comme gris (e.g., la couleur d'une banane jaune était légèrement ajustée dans le sens du bleu). Cela suggère que lorsque les objets étaient objectivement gris, ils étaient en fait toujours perçus comme dans leur couleur d'origine. Au contraire, les disques colorés non associés à une couleur spécifique étaient ajustés avec précision à un niveau de gris objectif. En utilisant une tâche similaire, Lupyan (2015) a également observé que les objets associés à une couleur typique (e.g., des bananes jaunes) étaient perçus comme ayant une couleur plus vive que des objets de la même couleur, mais non associés à une couleur typique (e.g., un t-shirt jaune ; Figure 1.3.C).

Figure 1.3

Exemples de Stimuli Utilisés Pour Étudier l'Influence des Connaissances sur l'Apparence des Stimuli.



Note. A. Stimuli utilisés par Han et VanRullen (2016) pour étudier l'influence des connaissances sur la perception subjective du contraste. B. Stimuli utilisés par Perez et al. (2020) pour étudier l'influence des connaissances sur la perception subjective du flou. C. Stimuli utilisés par Lupyan (2015) pour étudier l'influence des connaissances sur la perception subjective de la couleur.

Chapitre 1 - Contexte théorique : Influence des mécanismes prédictifs sur la perception visuelle

Dans l'ensemble, ces études soutiennent l'hypothèse selon laquelle les attentes fondées sur des connaissances préalables influencent qualitativement la perception visuelle en modifiant l'apparence de différentes caractéristiques des stimuli visuels (orientation, mouvement, couleur, contraste, netteté), c'est à dire leur perception subjective. De plus, la plupart de ces études suggèrent un « rehaussement » perceptif des caractéristiques des stimuli qui sont conformes aux attentes, conduisant à les voir par exemple comme plus colorés, ou plus nets qu'ils ne le sont.

1.3.3. Critiques et controverses théoriques

Les conclusions des études citées ci-dessus ont néanmoins été remises en question par plusieurs auteurs (Bang & Rahnev, 2017; Firestone & Scholl, 2014, 2015, 2016; Rungratsameetaweemana et al., 2018; Rungratsameetaweemana & Serences, 2019) qui affirment que ce type d'étude ne permet pas de s'assurer que les effets observés sont bien liés à l'influence des connaissances sur la perception subjective et non pas sur d'autres processus non-perceptifs.

Tout d'abord, les tâches impliquant des jugements perceptifs comparatifs ne permettraient pas de dissocier les effets perceptifs d'effets non perceptifs, liés par exemple à la prise de décision. Par exemple, dans la tâche de discrimination d'orientation utilisée par Kok, Jehee, et al. (2012), les participants pourraient s'appuyer sur l'indice sonore pour deviner l'orientation des patchs de Gabor et faciliter leur comparaison sans que cela ne soit associé à une meilleure perception de ceux-ci. Une telle critique est également valable pour les études mentionnées précédemment sur la reconnaissance d'objets en contexte. En effet, les meilleures performances généralement observées pour les stimuli congruents qu'incongruents dans ces tâches de reconnaissance pourraient s'expliquer par un biais de réponse. Par exemple, dans le cas d'une tâche de catégorisation d'objets en contexte, les participants pourraient baser leurs réponses sur les informations contextuelles uniquement, ce qui entraînerait nécessairement de meilleures performances lorsque l'objet est congruent qu'incongruent avec le contexte sans que cela implique une meilleure perception de l'objet (Henderson & Hollingworth, 1999; Spaak et al., 2022 ; voir toutefois Roux-Sibilon et al., 2019 pour un contrôle de ce biais). En ce sens, Bang et Rahnev (2017) ont mené une étude dans le but de déterminer si les prédictions influençaient plutôt les processus perceptifs ou décisionnels. Pour ce faire, ils présentaient un patch de Gabor à des participants. Ceux-ci devaient indiquer si le patch de Gabor était orienté

Chapitre 1 - Contexte théorique : Influence des mécanismes prédictifs sur la perception visuelle

dans le sens horaire ou anti-horaire. Le patch de Gabor était soit précédé d'un indice informatif sur son orientation (des flèches vers la droite ou vers la gauche pour indiquer une orientation dans le sens horaire ou anti-horaire, respectivement), soit suivi de cet indice informatif. Lorsque l'indice précédait le patch de Gabor, il pouvait influencer à la fois le traitement perceptif et les processus décisionnels, et lorsque l'indice suivait le patch de Gabor, il ne pouvait influencer que les processus décisionnels. Les performances des participants étaient meilleures lorsque l'indice était présenté après le patch de Gabor, suggérant que les attentes affectent plutôt les processus décisionnels que perceptifs.

Firestone et Scholl (2015) distinguent également ce qui relève de la perception de ce qui relève de jugements non-perceptifs. Pour ces auteurs, lorsque nous voyons un objet, nous percevons certaines de ses propriétés (e.g., sa taille), mais nous pouvons également faire des jugements non-perceptifs à son sujet (e.g., concernant son prix). Cependant, certaines propriétés (e.g., la couleur) peuvent à la fois être perçues visuellement et faire l'objet de jugements non-perceptifs. Par exemple, la couleur d'une banane peut être perçue comme jaune mais, grâce à nos connaissances et nos attentes, elle pourrait également être jugée comme jaune même dans l'obscurité où l'on distinguerait à peine sa couleur. Dans ce contexte, les effets des attentes observés dans les études portant sur l'apparence (colorée, contrastée, nette, etc.) pourraient être en fait imputables à des biais de réponse ou des jugements de haut niveau. Ainsi, le fait qu'une banane soit associée à la couleur jaune pourrait induire un biais de réponse des participants qui reviendrait à toujours juger une image de banane comme jaune, indépendamment de ce qui leur est présenté, et à ajuster sa couleur pour correspondre à ce jugement sans que cela ne soit associé à une modification réelle de la perception subjective de la couleur. De la même façon, les stimuli pouvant plus aisément être associés à un objet dans les études de Han et VanRullen (2016) et de Perez et al. (2020) pourraient être plus facilement jugés comme visibles et donc « contrastés » ou « nets » qu'un objet sans signification, conduisant à biaiser les reports perceptifs dans le sens de ces jugements. L'utilisation de tâches où l'on demande aux participants de répondre explicitement sur ces caractéristiques (e.g., indiquer quel objet est le plus contrasté, quel objet est le plus flou, etc.) serait d'autant plus propice à la présence de tels biais (Firestone & Scholl, 2016).

Par ailleurs, dans la plupart de ces études impliquant des jugements comparatifs entre deux stimuli (Han & VanRullen, 2016; Kok, Jehee, et al., 2012; Perez et al., 2020; Teufel et al., 2018), les stimuli partagent des caractéristiques visuelles de bas niveau globalement similaires, mais leur signal est objectivement différent. Par exemple, le réarrangement spatial de traits noirs

Chapitre 1 - Contexte théorique : Influence des mécanismes prédictifs sur la perception visuelle

dans l'étude de Han et VanRullen (2016) ou celui de formes dans l'étude de Perez et al. (2020) dans le but de limiter l'accès à des connaissances et la génération d'attentes pourrait en contrepartie induire des différences objectives dans le signal. Il est donc possible que ces différences visuelles en terme de signal expliquent les différences de perception entre les stimuli.

Enfin, Firestone et Scholl (2016) soulignent que la plupart des études sur des comparaisons de stimuli sont uniquement confirmatoires et ne comportent pas de condition pouvant permettre de mettre à l'épreuve leurs conclusions quant au fait que les effets observés sont bien de nature perceptive et non de nature non-perceptive (notamment décisionnelle ou liées à des jugements de haut niveau). En effet, si les connaissances influencent effectivement la perception, il faudrait non seulement montrer que la perception de deux stimuli diffère en fonction de leur conformité avec ces connaissances, mais également s'assurer qu'une telle différence de perception n'est pas observée dans des conditions où elle ne devrait pas l'être, à savoir dans le cas de deux stimuli parfaitement identiques. Par exemple, si les connaissances influencent effectivement la perception, alors cela devrait toujours s'appliquer de la même manière quel que soit le contexte. Ainsi, une banane grise devrait toujours être subjectivement perçue comme jaunâtre (Firestone & Scholl, 2016). De plus, une seconde banane grise identique présentée en même temps que la première devrait également bénéficier de ce biais perceptif et devrait donc paraître tout aussi jaune. Ainsi, il ne devrait pas y avoir de différence dans les reports subjectifs concernant deux stimuli identiques soumis aux mêmes biais perceptifs. En revanche, si les reports concernant deux stimuli identiques sont différents, cela indiquerait plutôt la présence d'un biais non-perceptif comme par exemple un biais de réponse. Ce raisonnement est généralement qualifié par le terme de « *El Greco fallacy* » (Firestone & Scholl, 2014 ; mais voir Dubova & Goldstone, 2022 pour une critique de ce raisonnement). Il fait référence à une explication erronée des peintures distordues d'El Greco (caractérisées par des formes exagérément allongées), selon laquelle ces distorsions viendraient du fait que l'artiste était astigmatique et percevait le monde de façon distordue, ce qu'il reproduisait ensuite dans ses toiles. Cependant, s'il percevait le monde de façon distordue, il aurait également dû percevoir ses toiles comme distordues. In fine les effets se seraient annulés et ses toiles auraient été peintes non distordues. Firestone et Scholl (2014) font référence à cet exemple pour indiquer que lorsqu'il y a un biais perceptif (e.g., l'astigmatisme dans le cas d'El Greco), il doit s'appliquer de manière similaire partout et notamment lors de la comparaison de deux stimuli identiques, ce qui entraîne l'annulation des effets du biais. Cependant, les études menées

Chapitre 1 - Contexte théorique : Influence des mécanismes prédictifs sur la perception visuelle

jusqu'alors n'incluaient pas ce type de contrôle où deux stimuli identiques seraient comparés afin de vérifier ce raisonnement.

En résumé, bien que de nombreuses études démontrent que la perception visuelle est influencée par les connaissances que nous avons a priori sur notre environnement visuel, et par les attentes qui en découlent, ces études se voient fréquemment opposer l'idée que les effets observés reflètent plutôt une influence des connaissances sur des processus qui ne relèveraient pas de la perception. De plus, la diversité des paradigmes et des mesures utilisées dans ces études, mais aussi les critiques émises à leur égard révèlent que la définition de ce qu'est la perception est loin d'être consensuelle. Alors qu'elle est considérée par certains comme un processus objectivable par des mesures telles que le temps de réponse (Biederman et al., 1982; Greene et al., 2015; Joubert et al., 2007, 2008; Lai et al., 2020; Roux-Sibilon et al., 2019), d'autres l'envisagent au contraire d'un point de vue phénoménologique et donc comme nécessairement subjective (Chalk et al., 2010; Denison et al., 2016; Han & VanRullen, 2016; Hansen et al., 2006; Lupyan, 2015, 2017; Perez et al., 2020). De plus, alors que la perception englobe pour certains des processus allant jusqu'à la reconnaissance (Biederman et al., 1982; Brandman & Peelen, 2017; Davenport, 2007; Davenport & Potter, 2004; Greene et al., 2015; Joubert et al., 2007, 2008; Lai et al., 2020; Palmer, 1975; Rémy et al., 2013; Roux-Sibilon et al., 2019), d'autres distinguent la perception d'autres processus qui seraient non-perceptifs (Bang & Rahnev, 2017; Firestone & Scholl, 2016; Rungratsameetaweemana et al., 2018; Rungratsameetaweemana & Serences, 2019). Cette distinction suppose qu'il existerait alors une perception « pure » qui serait isolée de ces processus non-perceptifs et qu'il serait en outre possible de la mesurer. Si l'on peut rester sceptique quant à cette possibilité, on peut toutefois se donner l'objectif d'étudier l'influence des connaissances et des attentes sur la perception visuelle en minimisant au maximum l'influence d'autres facteurs supposément non-perceptifs (e.g., décisionnels, mnésiques, sémantiques...).

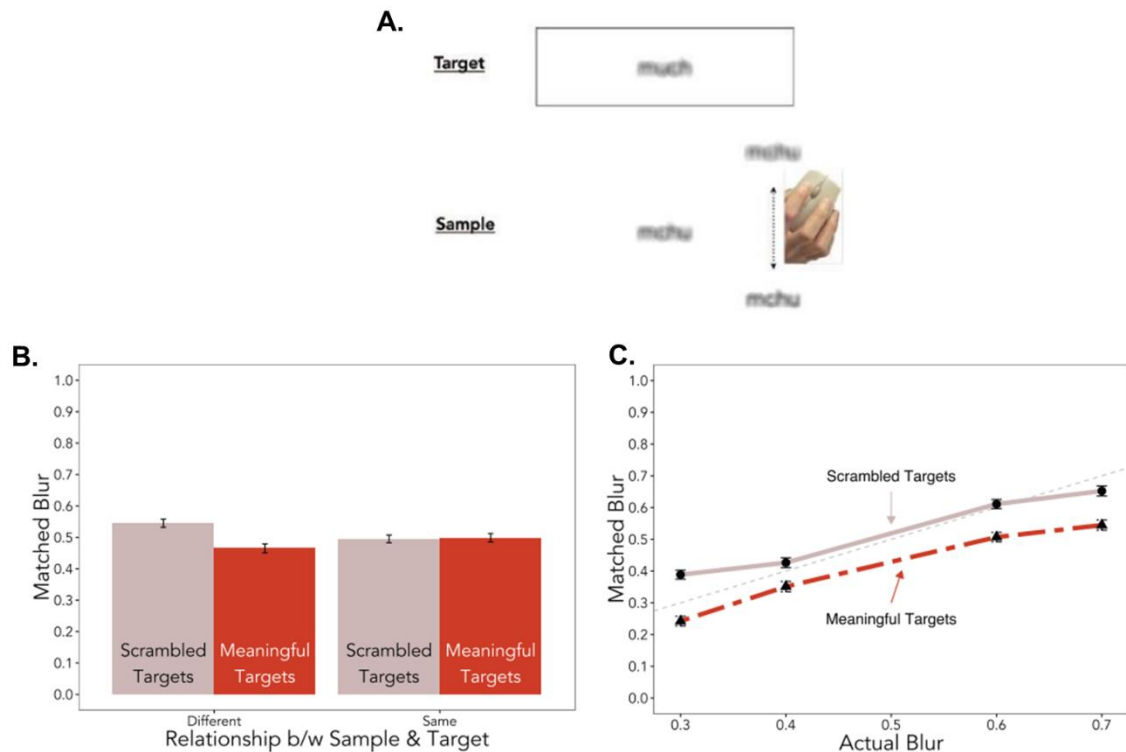
C'est par exemple l'approche qui a été adoptée par Lupyan (2017) dans une étude visant à déterminer si les connaissances sémantiques influencent l'apparence subjective de mots écrits. Dans son étude, l'auteur présentait à des participants deux suites de lettres floues l'une en dessous de l'autre, l'une étant appelée *target* et l'autre *sample*. Les participants devaient effectuer une tâche d'ajustement perceptif entre les deux suites de lettres. Ils devaient ajuster le niveau de flou du stimulus sample afin qu'il corresponde au niveau de flou du stimulus target en utilisant la molette d'une souris d'ordinateur (Figure 1.4). Cette tâche d'ajustement perceptif présente l'avantage de mesurer les différences de perception entre les stimuli sample et target

Chapitre 1 - Contexte théorique : Influence des mécanismes prédictifs sur la perception visuelle

de manière plus implicite que des tâches de discrimination ou de jugement comparatif (e.g., dire quel stimulus est le plus flou), ces dernières pouvant faire intervenir dans une plus grande mesure des processus décisionnels ou biais de réponse. Afin d'étudier le rôle des connaissances sémantiques sur la perception de la netteté des mots, l'auteur manipulait la présence d'une information sémantique dans les suites de lettres présentées en target et en sample. Pour cela, une des séries de lettres était un mot ayant une signification (i.e., associé à des connaissances sémantiques antérieures ; e.g., « much ») et l'autre série de lettres était un pseudo-mot sans signification créé en mélangeant l'ordre des lettres du mot (i.e., sans connaissances sémantiques antérieures ; e.g., « mchu »). L'expérience comprenait deux conditions expérimentales d'intérêt : (1) le stimulus en target pouvait être un mot et le stimulus en sample un pseudo-mot ou (2) le stimulus en target pouvait être un pseudo-mot et le stimulus en sample un mot. Les résultats ont montré que les participants ajoutaient plus de flou à un mot en sample apparié à un pseudo-mot en target qu'à un pseudo-mot en sample apparié à un mot en target. Cela signifie qu'à un niveau de flou objectivement égal entre les stimuli target et sample, les participants attribuaient tout de même plus de flou aux mots (associés aux connaissances) et donc, que les mots étaient subjectivement perçus comme étant plus nets que les pseudo-mots. Dans cette étude, l'auteur a également inclus une condition contrôle dans laquelle les stimuli en target et en sample étaient strictement identiques (i.e., deux mots ou deux pseudo-mots). En réponse aux critiques soulevées par Firestone et Scholl (2014, 2015), ce contrôle avait pour but de s'assurer que les effets observés étaient de nature perceptive. En effet, si les connaissances sémantiques influencent la perception subjective lors de la lecture, lorsque les stimuli en target et en sample sont identiques, les effets perceptifs des connaissances devraient affecter de la même manière les deux stimuli, ce qui se traduirait par l'absence de différence de flou perçue entre eux (cf. Raisonnement du El Greco fallacy, Firestone & Scholl, 2014). Au contraire, une différence d'ajustement entre une condition avec deux mots et une condition avec deux pseudo-mots suggérerait un biais de réponse consistant à systématiquement attribuer plus de flou aux mots qu'aux pseudo-mots, indépendamment de ce à quoi ils sont comparés. Les résultats de cette condition se sont révélés compatibles avec la première hypothèse. La différence de flou observée dans les deux conditions expérimentales d'intérêt (avec un mot et un pseudo-mot) disparaissait lorsque les stimuli étaient identiques (deux mots ou deux pseudo-mots). Ces résultats confortent donc l'idée que les attentes fondées sur des connaissances sémantiques antérieures modifient l'apparence subjective des mots flous en les faisant paraître plus nets et donc, que les attentes affectent également « comment nous voyons ».

Figure 1.4

Paradigme d'Ajustement Perceptif de Lupyan (2017).



Note. A. Procédure expérimentale. La tâche consistait à ajuster le niveau de flou du stimulus sample jusqu'à ce qu'il corresponde au niveau de flou du stimulus target. B. Niveau de flou attribué au stimulus sample pour correspondre au stimulus target (*matched blur*) pour les différentes conditions. Les stimuli target avaient un niveau de flou moyen correspondant à 0.5. Lorsque les stimuli target et sample étaient différents, les participants ajoutaient plus de flou que nécessaire pour ajuster un mot en sample apparié à un pseudo-mot en target (*scrambled targets*) et pas assez de flou pour ajuster un pseudo-mot en sample apparié à un mot en target (*meaningful targets*). Cet effet disparaissait lorsque les stimuli en sample et en target étaient identiques. C. Niveau de flou attribué au stimulus sample pour correspondre au stimulus target (*matched blur*) en fonction du niveau de flou du stimulus target pour les conditions où les deux stimuli étaient différents. La ligne en pointillés correspond à un ajustement parfait du niveau de flou du stimulus sample au niveau de flou du stimulus en target. Figure adaptée de Lupyan (2017).

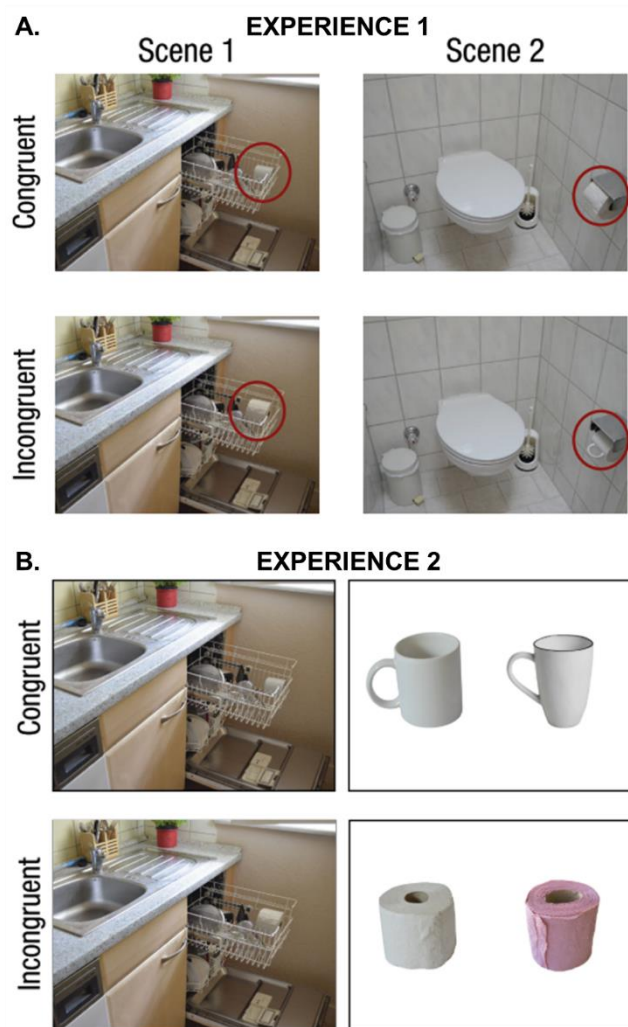
Plus récemment, Spaak et al. (2022) ont développé un paradigme expérimental particulièrement original pour évaluer l'influence des connaissances fondées sur les associations entre les objets et leur contexte sur la perception des objets. Pour cela, les auteurs

Chapitre 1 - Contexte théorique : Influence des mécanismes prédictifs sur la perception visuelle

ont mené deux expériences comportementales en utilisant des photographies (Figure 1.5) de scènes d'intérieur contenant un objet attendu (e.g., une tasse dans un lave-vaisselle) ou un objet inattendu (e.g., un rouleau de papier toilette dans un lave-vaisselle). Contrairement à la plupart des études que nous avons présentées jusqu'ici qui utilisaient des stimuli assez simples et peu écologiques (e.g., patches de Gabor, dessins au trait ou silhouettes d'objets), les scènes utilisées ici sont plus proches des stimuli pouvant être rencontrés au quotidien. Dans la première expérience, les participants devaient détecter les changements survenant dans les scènes (disparition d'un objet) qui pouvaient concerner soit les objets attendus, soit les objets inattendus (Figure 1.5). Dans la deuxième expérience, les participants exploraient librement les scènes (e.g., une cuisine). Ils voyaient ensuite deux exemplaires d'un même objet (e.g., deux tasses). La tâche des participants consistait à identifier lequel des deux objets proposés avait été présenté précédemment dans la scène, l'autre étant un distracteur non présent dans la scène (Figure 1.5). Dans cette seconde expérience, les participants ne pouvaient pas s'appuyer sur le contexte pour effectuer la tâche puisqu'ils devaient toujours choisir entre deux objets congruents ou deux objets incongruents avec le contexte. Ainsi, ils pouvaient seulement s'appuyer sur l'apparence des objets pour répondre. Les performances des participants étaient donc directement liées à la manière dont ils percevaient subjectivement les stimuli et ne pouvaient pas être expliquées par un biais de réponse consistant à baser leur réponse uniquement sur le contexte (potentiellement présent dans la première expérience). Les résultats des deux expériences ont montré que les performances de détection de changement et d'identification étaient plus élevées pour les objets inattendus que pour les objets attendus dans les scènes. Ces résultats suggèrent donc que même lorsque le contexte n'est pas pertinent pour la résolution de la tâche, il peut influencer la perception d'objets et que les objets inattendus sont mieux perçus que les objets attendus. Toutefois, contrairement aux travaux de Lupyan (2017) qui permettaient de quantifier les différences dans l'apparence perçue des différents stimuli (en termes de niveau de flou perçu), cette étude ne permet pas d'avoir directement accès à la perception qu'ont les participants des stimuli, en termes d'apparence subjective.

Figure 1.5

Paradigme de Spaak et al. (2022).



Note. A. Expérience 1. Exemple de stimuli attendus et inattendus sur la base du contexte de scène congruent ou incongruent utilisés dans le paradigme de détection de changement. B. Expérience 2. Exemple de stimuli cible et de distracteurs attendus et inattendus utilisés dans le paradigme de choix forcé. Figure adaptée de Spaak et al. (2022).

On peut cependant remarquer que les résultats de cette étude vont dans le sens opposé des effets observés dans l'étude de Lupyan (2017) et la plupart des études que nous avons précédemment rapportées. En effet, ces études suggèrent que les connaissances favorisent l'efficacité du traitement ou entraînent un rehaussement perceptif de ce qui est attendu ou conforme à ces connaissances (Biederman et al., 1982; Brandman & Peelen, 2017; Chalk et al., 2010; Davenport, 2007; Davenport & Potter, 2004; Greene et al., 2015; Han & VanRullen,

Chapitre 1 - Contexte théorique : Influence des mécanismes prédictifs sur la perception visuelle

2016; Hansen et al., 2006; Joubert et al., 2007, 2008; Lai et al., 2020; Lupyan, 2015, 2017; Perez et al., 2020; Roux-Sibilon et al., 2019). Au contraire, les résultats de Spaak et al. (2022, voir aussi Denison et al., 2016; Hollingworth & Henderson, 2000; LaPointe et al., 2013; Mack et al., 2017) suggèrent que la perception est biaisée en faveur de ce qui est inattendu ou non conforme aux connaissances. La plupart des études présentées ne fournissent d'ailleurs pas forcément d'explication sur le sens de l'effet observé, leur objectif étant plutôt de démontrer une influence des connaissances sur la perception, indépendamment du sens de l'effet. La plupart de ces études ne détaillent pas non plus les mécanismes neurocognitifs sous-jacents qui expliqueraient la meilleure perception des stimuli attendus ou inattendus. Elles ne sont d'ailleurs pas toutes menées dans le même contexte théorique. Ces études diffèrent aussi les unes des autres par la tâche demandée ou encore le type de stimuli utilisés. Par conséquent, il est difficile de les comparer pour en extraire des principes communs pouvant expliquer ces résultats contradictoires. Des pistes intéressantes concernant la manière dont les connaissances et les attentes sont susceptibles de moduler la perception de stimuli attendus ou inattendus peuvent néanmoins être trouvées dans les modèles actuels dits « proactifs » du fonctionnement neurocognitif.

2. Par quels mécanismes les attentes basées sur les connaissances a priori sont-elles susceptibles d'influencer la perception ?

Les modèles actuels dits « proactifs » de la perception visuelle s'appuient sur les propriétés neurophysiologiques des voies visuelles pour proposer des mécanismes biologiquement plausibles par lesquels les connaissances a priori et les attentes sont susceptibles d'influencer la perception. Afin d'en faciliter la compréhension, nous rappellerons dans un premier temps les principales caractéristiques de l'organisation anatomique et fonctionnelle des voies visuelles.

2.1. Rappels sur l'organisation anatomique et fonctionnelle des voies visuelles

2.1.1. De la rétine au cortex visuel

D'un point de vue anatomique et fonctionnel, le système visuel est caractérisé par une organisation hiérarchique. Les signaux en provenance de la rétine transitent par une succession d'aires visuelles, où chaque aire intègre les traitements effectués par les aires précédentes. Chaque aire est spécialisée dans le traitement de caractéristiques visuelles de plus en plus complexes et élaborées (Felleman & Van Essen, 1991; Marr & Vaina, 1982; Van Essen & Deyoe, 1995; Van Essen & Maunsell, 1983). Ainsi, le traitement de l'information visuelle se fait de manière ascendante, des premiers niveaux (dont la rétine) vers les derniers niveaux de la hiérarchie.

Dès la rétine, on observe différents types de cellules ganglionnaires aux propriétés distinctes (Curcio et al., 1990; Enroth-cugell & Robson, 1966, 1984; Kaplan, 2004; Kaplan & Shapley, 1986). En particulier, les cellules ganglionnaires de type *midget* sont sensibles aux informations chromatiques mais peu sensibles à la luminance. Elles présentent un champ récepteur de petite taille ainsi qu'un axone fin permettant de véhiculer des informations fines et détaillées mais de manière assez lente. Elles sont donc sensibles aux hautes fréquences spatiales (HFS) et aux basses fréquences temporelles. A l'inverse, les cellules ganglionnaires de type *parasol* sont sensibles aux contrastes de luminance mais insensibles aux informations chromatiques. Elles présentent un large champ récepteur ne leur permettant de répondre qu'à de grandes formes visuelles. Elles sont donc sensibles aux basses fréquences spatiales (BFS). Par contre, le large diamètre de leur axone et la myélinisation de celui-ci leur permettent une vitesse de conduction très rapide, les rendant donc sensibles aux hautes fréquences temporelles (Curcio et al., 1990; Enroth-cugell & Robson, 1966, 1984; Kaplan, 2004; Kaplan & Shapley, 1986). Il existe également des cellules ganglionnaires koniocellulaires mais nous n'aborderons pas en détails leurs propriétés puisque leurs fonctions sont d'un intérêt moindre pour la suite de ce travail.

En sortie de la rétine, l'information visuelle est projetée depuis les cellules ganglionnaires rétiniennes vers le corps genouillé latéral du thalamus. Les cellules ganglionnaires parasol et midget se projettent sur des couches distinctes du corps genouillé latéral où elles font synapse avec des cellules qui se projettent à leur tour sur des couches distinctes du cortex visuel primaire (V1), puis secondaire (V2) donnant ainsi naissance aux

Chapitre 1 - Contexte théorique : Influence des mécanismes prédictifs sur la perception visuelle

voies magnocellulaire et parvocellulaire, respectivement. Les aires visuelles de bas niveau telles que V1 et V2 analysent le signal visuel de façon locale, rétinotopique, selon des composantes d'orientation et de fréquences spatiales (Daniel & Whitteridge, 1961; De Valois et al., 1982; Hubel & Wiesel, 1962, 1968, 1977). Au niveau de V1, une extraction de caractéristiques élémentaires à lieu en chaque point du champ visuel par des neurones ayant un champ récepteur de petite taille. Ces champs récepteurs ont la forme de barres orientées ayant différentes fréquences spatiales. Les traitements effectués par les cellules de V1 sont intégrés par les cellules de V2 dont les champs récepteurs sont donc plus grands et qui comportent des représentations plus globales (extraction de contours).

2.1.2. Au-delà du cortex strié : Les voies ventrale et dorsale

Après leur passage dans V1 et V2, les voies magnocellulaire et parvocellulaire se séparent lors de leur projection dans les aires visuelles extrastriées pour former les voies visuelles ventrale et dorsale (Figure 1.6 ; Bullier, 2001; Goodale & Milner, 1992). La voie dorsale véhicule l'information visuelle vers le cortex pariétal postérieur. Elle contient des projections exclusivement magnocellulaires. Elle est donc sensible aux BFS. Elle comporte des aires qui encodent notamment les relations spatiales entre objets et qui sont sensibles au mouvement. Elle serait plus spécifiquement impliquée dans les traitements visuo-spatiaux. La voie ventrale véhicule l'information jusqu'au cortex inférotemporal. Elle contient majoritairement des afférences parvocellulaires, mais aussi magnocellulaires (Merigan & Maunsell, 1993). Les aires le long de cette voie sont sensibles aux BFS, mais surtout aux HFS et répondent principalement à des informations chromatiques ainsi qu'aux formes complexes. Cette voie serait tout particulièrement impliquée dans le traitement des informations permettant la reconnaissance visuelle de stimuli complexes, tels que les objets, les visages ou les scènes visuelles (Goodale & Milner, 1992). Ces deux voies, ventrale et dorsale, entretiendraient cependant des connexions et elles ne fonctionneraient pas de manière totalement indépendante (Merigan & Maunsell, 1993).

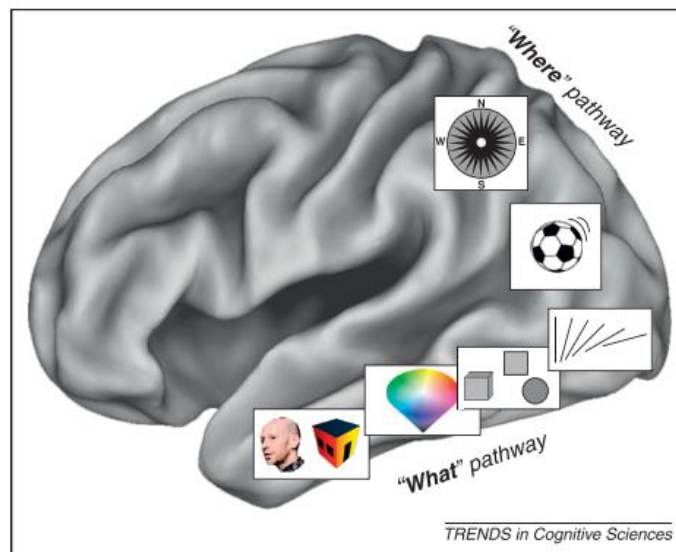
Le long de ces deux voies sont situées des aires cérébrales dont les cellules ont des champs récepteurs de plus en plus grands et qui sont spécialisées dans l'extraction de caractéristiques de plus en plus élaborées. Par exemple, le long de la voie ventrale, on retrouve l'aire V4 spécialisée dans la représentation des couleurs et des formes complexes, (Gallant et al., 1996) ou encore des aires spécialisées dans la reconnaissance de catégories spécifiques

Chapitre 1 - Contexte théorique : Influence des mécanismes prédictifs sur la perception visuelle

(scènes, objets, visages). On peut, par exemple, citer un ensemble de régions spécifiques à l'analyse de scènes, telles que le cortex rétro-splénial (Aggleton & Vann, 2004), l'occipital place area (Dilks et al., 2013) et la parahippocampal place area (Epstein & Kanwisher, 1998). Ces régions sont dites spécifiques aux scènes dans le sens où elles ne répondent pas à d'autres stimuli comme les objets isolés ou les visages. A contrario, la fusiform face area répond sélectivement aux visages (Kanwisher et al., 1997), tandis que le complexe occipital latéral (LOC) est sélectif aux objets (Grill-spector, 2003; Malach et al., 1995). Ainsi, les aires de haut niveau dans la hiérarchie corticale comporteraient la représentation la plus élaborée des stimuli visuels, invariante au point de vue, aux changements de luminance ou de couleurs qui permettrait d'aboutir à la reconnaissance.

Figure 1.6

Représentation Schématique des Caractéristiques Extraites Dans les Aires Cérébrales le Long des Voies Ventrale et Dorsale.



Note. Figure adaptée de de Haan & Cowey (2011).

2.1.3. Rôle des connexions descendantes

Plusieurs études en neurophysiologie, réalisés notamment chez le primate non-humain, ont mis en évidence la présence de connexions bidirectionnelles le long des voies ventrale et dorsale (Bullier, 2001; Ergenzinger et al., 1998) remettant en question l'idée que le traitement

Chapitre 1 - Contexte théorique : Influence des mécanismes prédictifs sur la perception visuelle

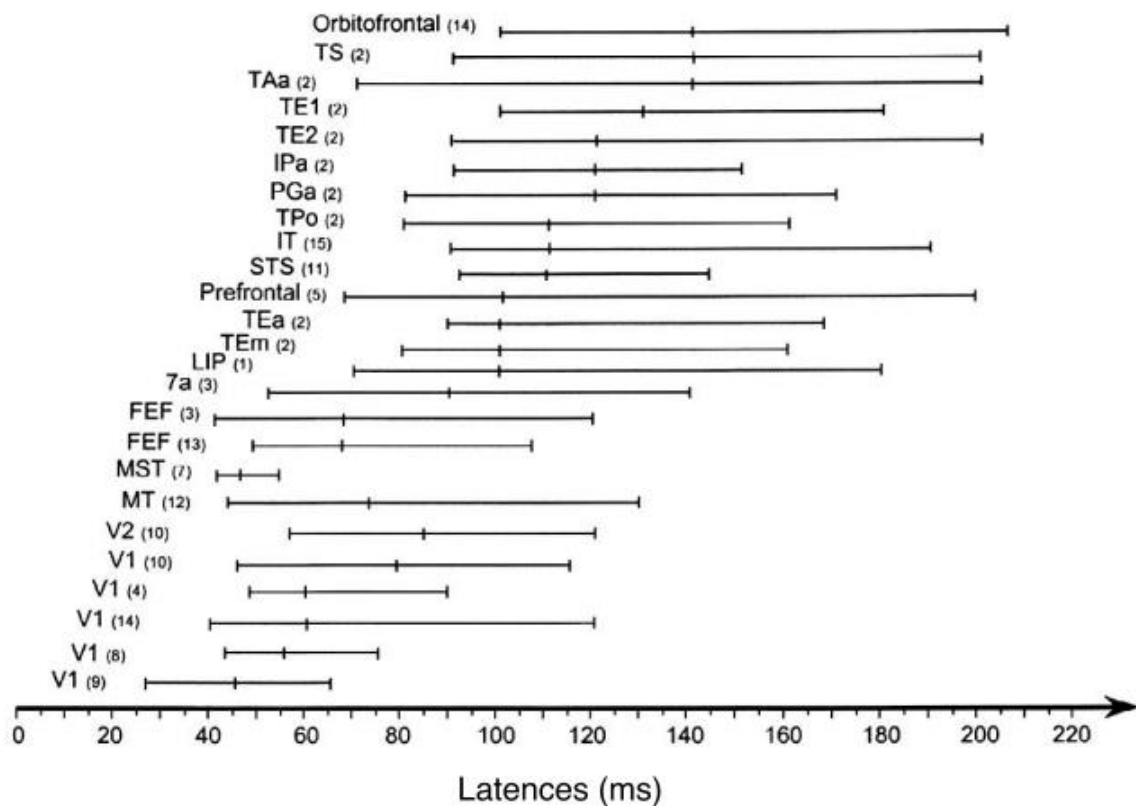
de l'information visuelle se fait de façon purement ascendante. Les connexions descendantes entre ces aires seraient d'ailleurs plus nombreuses que les connexions ascendantes. En effet, les projections ascendantes en provenance des aires de bas niveau auraient tendance à se projeter dans un nombre restreint d'aires de plus haut niveau tandis que les projections descendantes en provenance des aires de haut niveau se projetteraient dans un nombre important d'aires de bas niveau et auraient un vaste réseau de connexions qui leur permettrait d'orienter les traitements au niveau local au sein des aires de plus bas niveau (Angelucci et al., 2002). On parle d'influences descendantes. Elles constitueraient le substrat neuroanatomique permettant à des représentations de haut niveau (e.g., connaissances a priori) d'influencer le traitement de l'information visuelle dès les plus bas niveaux.

Des recherches en neurosciences suggèrent en effet que l'activité des aires visuelles de bas niveau n'est pas toujours expliquée par le seul traitement ascendant de l'information. Nous pouvons par exemple citer la célèbre illusion du triangle de Kanizsa où l'arrangement de trois « pacmen » noirs sur un fond blanc donne l'illusion d'un triangle blanc superposé à trois petits ronds noirs à chacune de ses extrémités. Lors de la perception de cette figure, les neurones de V1 dont les petits champs récepteurs se situent sur le contour illusoire du triangle s'activent alors qu'aucun contour n'est présent dans le signal visuel (Lee & Nguyen, 2001). Cela pourrait suggérer que l'activité des neurones de V1 est modulée de façon descendante par des aires de plus haut niveau, dont les champs récepteurs plus grands permettent une représentation plus globale de la figure.

Par ailleurs, des études en neurophysiologie chez le macaque ont mis en évidence des latences de réponse équivalentes voire plus courtes dans certaines aires de haut niveau dans la hiérarchie corticale que dans les aires de bas niveau (Bullier, 2001; Lamme et al., 1998). Par exemple, l'aire MT située bien après V1 le long de la voie dorsale a pourtant une latence d'activation comparable à celle de V1. On peut également noter que les aires de la voie dorsale ont en général des latences plus courtes que celles de la voie ventrale (Figure 1.7). Cela peut s'expliquer par la prédominance de cellules parasol à conduction rapide le long de la voie dorsale tandis que la voie ventrale compte majoritairement des cellules midget à conduction plus lente. Ces aires de haut niveau aux latences de réponse très courtes pourraient donc être activées assez rapidement pour pouvoir ensuite moduler la réponse d'aires de plus bas niveau telles que V1.

Figure 1.7

Latences de Réponse de Neurones Dans Différentes Aires Corticales.



Note. Pour chaque aire, la barre verticale centrale représente la latence médiane et les barres verticales de chaque côté représentent les 10^{ème} et 90^{ème} centiles. Figure adaptée de Bullier (2001).

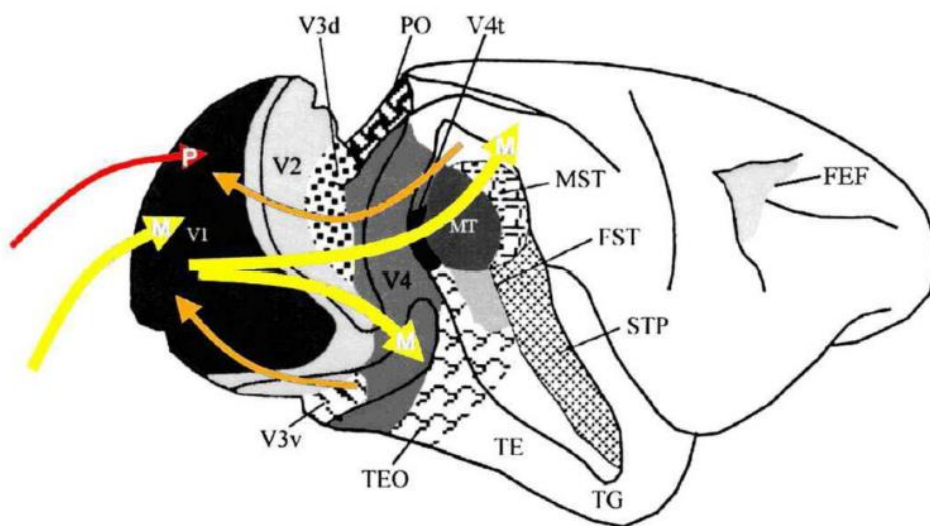
Ces données ont conduit Bullier (2001) à proposer un modèle intégratif du traitement de l'information visuelle qui repose sur l'hypothèse d'une rétro-injection du résultat des traitements réalisés dans les aires cérébrales de haut niveau sur les traitements réalisés dans les aires de plus bas niveau. Selon ce modèle, le traitement de l'information visuelle impliquerait trois vagues d'activation basées sur les propriétés des voies visuelles magnocellulaire et parvocellulaire (Figure 1.8). Dans une première vague, l'information visuelle de la voie magnocellulaire (qui véhicule, par exemple, les informations en BFS) se projetterait rapidement dans les aires corticales de bas niveau, puis dans les aires extra-striées de haut niveau via les voies dorsale et également ventrale. Dans une deuxième vague, le résultat des traitements effectués par les aires de haut niveau seraient rétro-injectés grâce à des connexions descendantes dans les aires de plus bas niveau où ils seraient intégrés au traitement des informations ascendantes de la voie parvocellulaire (qui véhicule, par exemple, les informations

Chapitre 1 - Contexte théorique : Influence des mécanismes prédictifs sur la perception visuelle

en HFS) en provenance de la rétine véhiculées plus lentement, constituant la troisième vague d'activation. Ce mécanisme de rétro-injection permettrait ainsi de contraindre, sur la base d'un premier traitement basé sur les informations magnocellulaires, le traitement des informations issues des afférences parvocellulaires dès leur entrée au niveau du cortex visuel primaire. Le modèle de Bullier (2001) est l'un des premiers modèles neurobiologiques de la perception visuelle à tenir compte de la réalité physiologique du système visuel en intégrant les influences descendantes qui contraignent le traitement de l'information visuelle ascendante.

Figure 1.8

Modèle Intégré du Traitement de l'Information Visuelle de Bullier (2001).



Note. Les flèches jaunes représentent la première vague d'activation véhiculée par la voie magnocellulaire (M), propagée des aires visuelles de bas niveau vers des aires visuelles de haut niveau. Les flèches orange représentent la vague d'activation descendante (rétro-injection), propagée des aires de haut niveau vers des aires de plus bas niveau. Cette rétro-injection a lieu au moment de l'arrivée plus tardive de l'information véhiculée par la voie parvocellulaire (P, flèche rouge) sur le cortex visuel primaire, qui constitue la troisième vague d'activation. Figure adaptée de Bullier (2001).

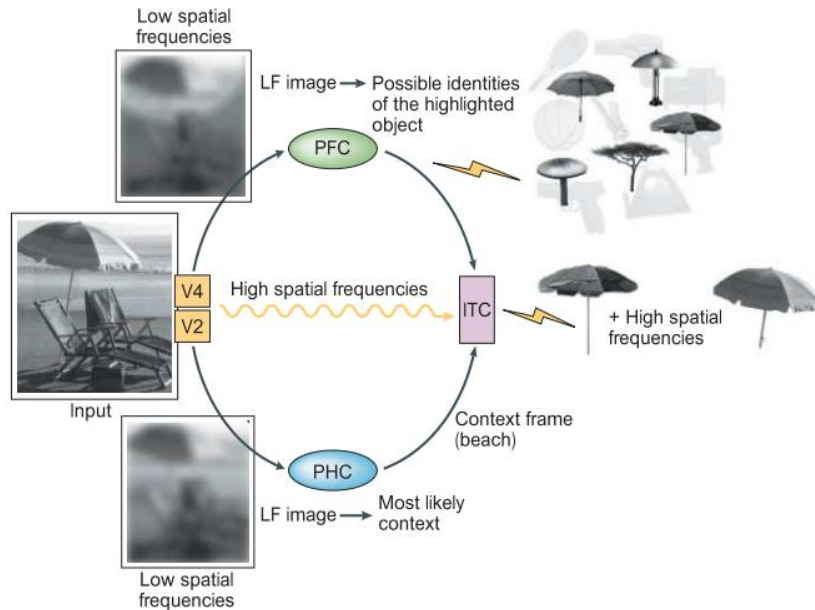
Grâce aux connaissances accumulées sur l'organisation et les propriétés physiques du système visuel, et en s'appuyant notamment sur le modèle de Bullier (2001), plusieurs chercheurs ont essayé de modéliser les étapes de traitement de l'information visuelle menant à la perception et la reconnaissance de stimuli complexes tels que les scènes et les objets.

2.2. Modèles neurocognitifs proactifs de la perception visuelle de scènes et d'objets

Dans le contexte de la perception visuelle de scènes et d'objets, le modèle le plus influant tenant compte explicitement de l'influence des connaissances a priori est sans doute le modèle proactif de la reconnaissance visuelle proposé par Moshe Bar et collaborateurs (Bar, 2003 ; voir aussi Bar, 2004, 2007, 2009; Bar et al., 2006; Trapp & Bar, 2015). Initialement développé dans le contexte de la reconnaissance d'objets, ce modèle s'appuie sur les mécanismes proposés par Bullier (2001) et suppose que la reconnaissance visuelle s'opère selon trois étapes (Figure 1.9). Lors d'une première étape, une information grossière en BFS de l'objet (e.g., un parasol) serait extraite de l'input visuel et véhiculée très rapidement via les fibres magnocellulaires jusqu'aux cortex orbitofrontal. En parallèle, l'information en BFS du contexte de la scène (e.g., une plage) serait également projetée rapidement vers le cortex parahippocampique. Ces projections en BFS permettraient aux cortex orbitofrontal et parahippocampique de générer des hypothèses ou prédictions concernant l'identité possible de l'objet et du contexte, respectivement. Lors d'une deuxième étape, ces prédictions seraient alors transmises, via des connexions descendantes, vers les aires inférotemporales de la voie ventrale impliquées dans la reconnaissance d'objet telles que le gyrus fusiforme où elles permettraient de limiter les interprétations possibles du stimulus. L'intégration de ces influences descendantes à l'arrivée des informations ascendantes en HFS véhiculées plus lentement le long de la voie ventrale permettrait dans un troisième temps d'aboutir à la reconnaissance.

Figure 1.9

Schéma du Modèle de Bar.



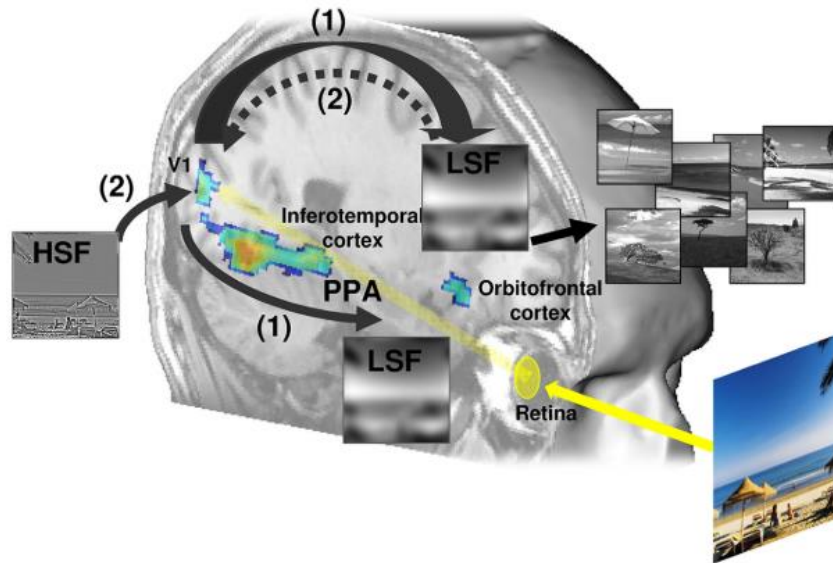
Note. Les informations grossières de basses fréquences spatiales (*Low spatial frequencies*) extraites de l'objet permettent de générer des prédictions sur les candidats potentiels de l'objet. En parallèle, les informations grossières de basses fréquences spatiales extraites du contexte permettent de restreindre les interprétations possibles de l'objet. Ces deux informations de basses fréquences spatiales permettent ensuite de guider le traitement des informations de hautes fréquences spatiales (*High spatial frequencies*) véhiculées plus lentement afin d'aboutir à la reconnaissance de l'objet. Figure adaptée de Bar (2004).

Ce modèle a par la suite été étendu à la reconnaissance de scènes dans le contexte des recherches menées par Carole Peyrin et collaborateurs (Kauffmann et al., 2014; Kauffmann, Bourgin, et al., 2015; Kauffmann, Chauvin, et al., 2015; Peyrin et al., 2010). Leurs études en neuroimagerie ont notamment permis de montrer que l'information en BFS issue des scènes visuelles active rapidement le cortex frontal inférieur et que celui-ci influence ensuite de manière descendante l'activité des aires inférotemporales impliquées dans la reconnaissance de scènes telles que le cortex parahippocampique, mais aussi l'activité du cortex visuel primaire (Kauffmann, Chauvin, et al., 2015; Peyrin et al., 2010). Pour ces chercheurs, le cortex frontal inférieur générerait des attentes sur la base des BFS qui lui arrivent rapidement. Ces attentes pourraient alors influencer de façon descendante le traitement plus tardif de l'information en

HFS et ce, dès leurs premières étapes corticales de traitement (Figure 1.10), en accord avec le mécanisme de rétro-injection proposé par Bullier (2001).

Figure 1.10

Illustration du Modèle Cortical de la Catégorisation de Scènes Basé sur les Fréquences Spatiales.



Note. (1) Les informations en basses fréquences spatiales (LSF) atteignent rapidement les aires de haut niveau de la voie dorsale (cortex frontal inférieur), permettant une première analyse grossière de la scène visuelle permettant de générer des prédictions sur la catégorie de la scène, avant de se propager le long de la voie ventrale (cortex inféro-temporal). (2) Cette analyse initiale rudimentaire pourrait ensuite influencer de manière descendante l'activité des aires de niveau inférieur (y compris le cortex visuel primaire, V1) pour guider l'analyse plus lente des informations en hautes fréquences spatiales (HSF) le long de la voie ventrale et sélectionner les détails plus fins nécessaires à la reconnaissance et à l'identification. Figure adaptée de Kauffmann et al. (2014).

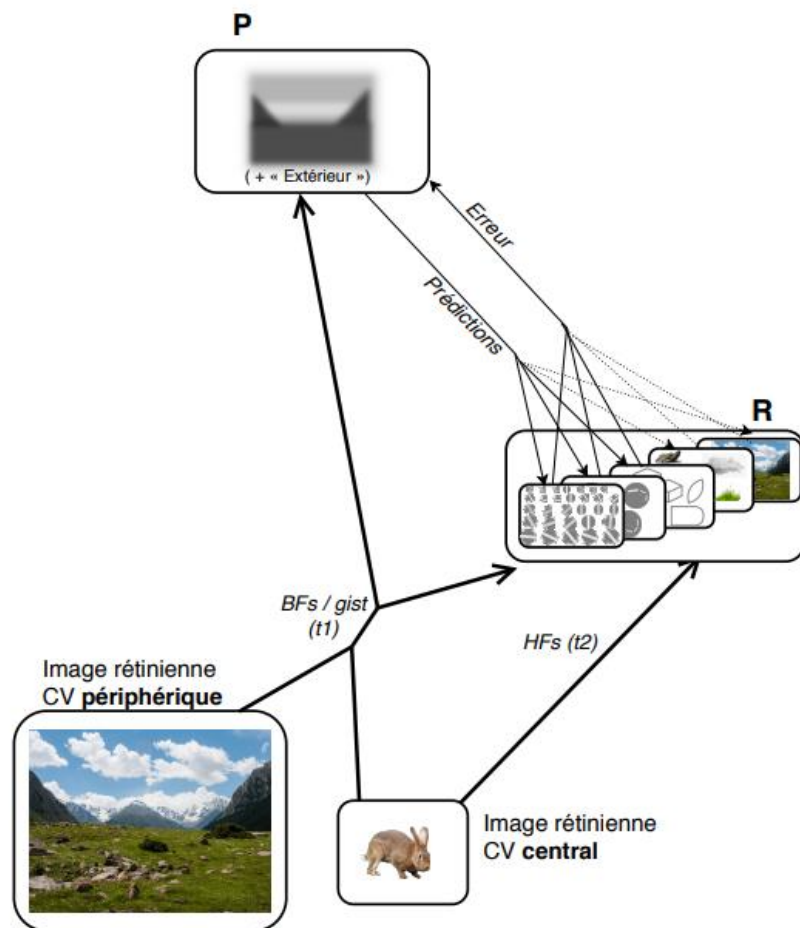
Une extension de ce modèle a plus récemment été proposée dans le cadre des travaux de thèse d'Alexia Roux-Sibilon (Peyrin et al., 2021; Roux-sibilon, 2020) pour tenir compte du traitement non homogène des fréquences spatiales sur la rétine (Figure 1.11). En effet, la répartition des cellules ganglionnaires sur la rétine est telle que l'information en HFS (les détails) n'est extraite qu'en vision fovéale et parafovéale, donc en vision centrale, tandis que l'information en BFS (les formes globales) est extraite sur l'ensemble du champ visuel et

Chapitre 1 - Contexte théorique : Influence des mécanismes prédictifs sur la perception visuelle

majoritairement en vision périphérique. Les BFS, en provenance des champs visuels central et périphérique, seraient rapidement véhiculées vers un système de reconnaissance, ainsi que vers un système de gestion des prédictions qui enverrait donc le résultat de ces prédictions au système de reconnaissance. Dans le même temps, les HFS extraites au niveau du champ visuel central seraient véhiculées plus lentement jusqu'au système de reconnaissance et elles pourraient alors intégrer les prédictions pour aboutir à la reconnaissance de la scène.

Figure 1.11

Illustration du Modèle Fréquentiel de la Catégorisation de Scènes.



Note. Les basses fréquences spatiales (BFS) en provenance des champs visuels (CV) central et périphérique seraient rapidement véhiculées vers un système de reconnaissance (R) mais aussi vers un système de gestion des prédictions (P) qui enverrait le résultat de ses prédictions au système de reconnaissance. Dans le même temps, les hautes fréquences spatiales (HFS) extraites au niveau du champ visuel central seraient véhiculées plus lentement jusqu'au système de reconnaissance et elles

Chapitre 1 - Contexte théorique : Influence des mécanismes prédictifs sur la perception visuelle

pourraient alors intégrer les prédictions pour aboutir à la reconnaissance de la scène. Figure adaptée de Roux-sibilon (2020).

Dans l'ensemble, en tenant compte des propriétés neurophysiologiques du système visuel, ces modèles proactifs proposent des mécanismes par lesquels les connaissances (sur les catégories de scènes, d'objets et les associations contextuelles entre eux) influencent le traitement de l'information visuelle dès les aires de bas niveau telles que V1. Selon ces modèles, c'est donc le traitement rapide d'une information grossière en BFS qui permettrait de faire le lien avec les connaissances et ainsi de générer des prédictions qui guideraient le traitement ultérieur de l'information plus détaillée en HFS. Ces modèles permettent donc d'expliquer comment le recours aux connaissances influence la rapidité et l'exactitude de la reconnaissance, en facilitant le traitement de l'information visuelle. Cependant, ces modèles ne précisent pas les mécanismes par lesquels les prédictions pourraient influencer le traitement de l'information dans les aires de bas niveau. Cela limite donc la portée de ces modèles et notamment leur capacité à expliquer comment les prédictions pourraient influencer l'apparence subjective des stimuli visuels. En incluant la notion de prédiction, ces modèles se rapprochent cependant d'un contexte théorique plus large qui est celui du codage prédictif. Il s'agit de modèles non spécifiques à la vision ou à la perception de scènes et d'objets mais qui s'intéressent plus largement au fonctionnement neurocognitif et dont les mécanismes sous-jacents sont plus détaillés. Les modèles du codage prédictif sont donc intéressants pour étudier les effets perceptifs des attentes (ou prédictions, qui est le terme privilégié dans ce courant théorique).

2.3. Modèles du codage prédictif

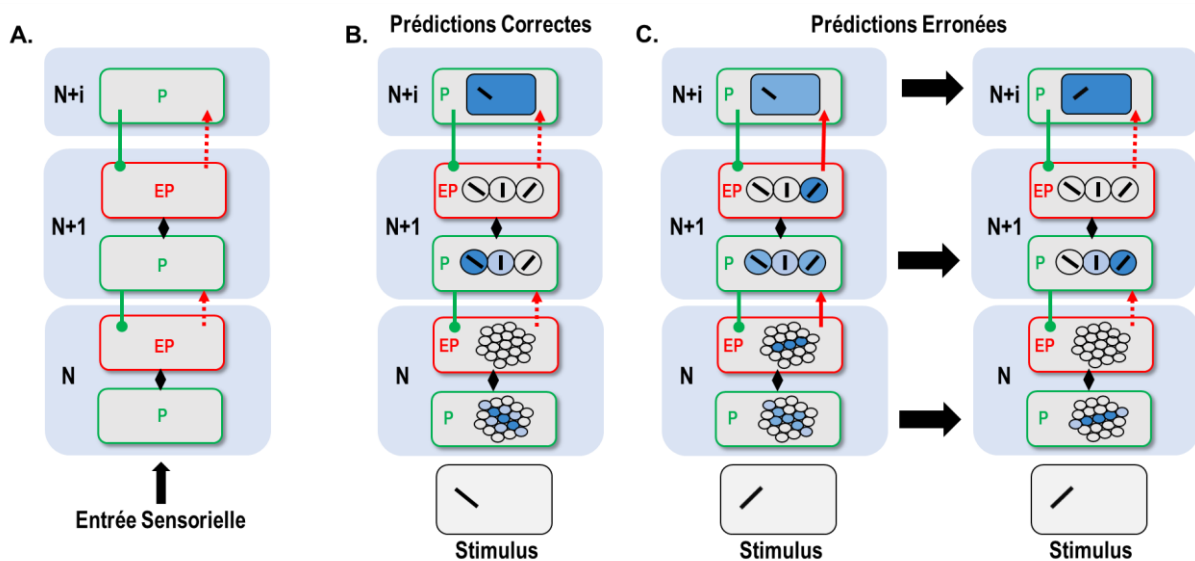
2.3.1. Modèle computationnel du codage prédictif

La théorie du codage prédictif est probablement l'une des théories qui a eu le plus d'influence dans le domaine des sciences cognitives au cours des 20 dernières années. Initialement proposé par Rao et Ballard (1999 ; voir aussi Mumford, 1992) comme modèle computationnel du traitement de l'information dans les aires visuelles corticales, ce modèle a été étendu à la plupart des processus perceptifs et cognitifs pour devenir une théorie plus générale et unifiée du fonctionnement neurocognitif (Clark, 2013; Feldman & Friston, 2010; Friston, 2005; Jiang & Rao, 2022; Millidge et al., 2021). Bien que plusieurs formalisations de

ce modèle aient été proposées, le postulat général repose sur l'idée que le cerveau construit des modèles internes du monde extérieur qui sont utilisés pour prédire et expliquer les signaux qui lui parviennent. Dans leur formalisation computationnelle, ces modèles s'appuient sur l'organisation hiérarchique des aires cérébrales et supposent que chaque niveau de traitement comporte deux populations de neurones aux propriétés distinctes : les unités de prédictions (P) et les unités d'erreur de prédiction (EP). Les unités de prédictions représentent les hypothèses ou prédictions concernant l'entrée reçue du niveau inférieur. Elles permettent donc en quelque sorte de pré-expliquer le signal reçu et elles inhibent le traitement des caractéristiques prédites du signal au niveau inférieur via des connexions descendantes. Les unités d'erreur de prédiction représentent l'écart entre les signaux de prédictions provenant des niveaux supérieurs et le signal provenant des niveaux inférieurs. L'erreur de prédiction est transmise via des connexions ascendantes vers le niveau supérieur qui les intègre afin de mettre à jour les prédictions pour qu'elles correspondent mieux aux signaux entrants. Ces échanges continus de signaux de prédictions et d'erreur de prédiction entre niveaux adjacents de la hiérarchie auraient pour résultat de réduire itérativement l'erreur de prédiction jusqu'à ce qu'il n'y en ait plus aucune et que l'activité des unités de prédictions à chaque niveau représente le plus fidèlement le stimulus. La représentation finale du stimulus résulterait donc de l'interaction entre les prédictions et l'erreur de prédiction (cf. Figure 1.12.A).

Figure 1.12

Illustration du Fonctionnement des Modèles du Codage Prédictif.



Note. A. Échange continu entre signaux provenant des unités de prédiction (P) et d'erreur de prédiction

Chapitre 1 - Contexte théorique : Influence des mécanismes prédictifs sur la perception visuelle

(EP) le long de la hiérarchie des aires visuelles lors de la perception visuelle. B. Représentation schématique du contenu des unités P et EP le long de la hiérarchie des aires visuelles dans le cas de prédictions correctes à propos du stimulus. Le stimulus correspondant parfaitement aux prédictions, la réponse des cellules des unités P est sélective aux caractéristiques prédites (illustré par la couleur bleu foncé) et l'activité cellulaire des unités EP est faible puisqu'il n'y a pas d'erreur de prédiction. C. Représentation schématique du contenu des unités P et EP le long de la hiérarchie des aires visuelles dans le cas de prédictions erronées à propos du stimulus. Le stimulus n'étant pas correctement prédit, l'erreur de prédiction est importante et sa représentation dans les unités P est moins univoque. Cela entraîne une réponse moins sélective des cellules des unités P et une augmentation de l'activité cellulaire des unités EP (partie de gauche). Dans un deuxième temps, le traitement de l'erreur de prédiction permet de mettre à jour la représentation du stimulus dans les unités P jusqu'à ce qu'il n'y ait plus d'erreur de prédiction (partie de droite).

Contrairement aux modèles proactifs de la reconnaissance cités plus haut (Bar, 2003; Bar et al., 2006; Kauffmann et al., 2014), les modèles du codage prédictif supposent donc que des prédictions sont générées à tous les niveaux de la hiérarchie des aires corticales (et pas seulement dans une région spécifique comme le cortex préfrontal). De plus, les prédictions générées à chaque niveau concernent des représentations de différente nature selon le niveau. Par exemple, dans le contexte de la perception visuelle d'objet, les prédictions générées aux niveaux les plus élevés concerneraient des propriétés abstraites sur le stimulus (e.g., catégorie sémantique) tandis que les prédictions générées aux niveaux les plus bas concerneraient des propriétés plus élémentaires (e.g., présence de contours à certains points du champ visuel).

Ce principe de codage prédictif comporte plusieurs avantages. Tout d'abord, il est efficace en termes de coûts de traitement, puisque l'inhibition du traitement des caractéristiques prédites par les unités de prédiction permet d'éviter des redondances dans le traitement de l'information à chaque niveau de la hiérarchie. De plus, le principe d'inférence hiérarchique sous-jacent à ce modèle permet d'aboutir rapidement à des représentations robustes et fiables du stimulus. Lorsque les caractéristiques de bas niveau sont ambiguës (e.g., alternance de patches noirs et blancs ; cf. Figure 1.1), leur traitement bénéficie de représentations de plus haut niveau permettant de résoudre l'ambiguïté (e.g., une femme embrassant un cheval). À l'inverse, les représentations de haut niveau (e.g., un visage de femme) peuvent également être affinées sur la base de représentations plus détaillées de bas niveau (e.g., traits spécifiques à l'identité du visage ; Kok & de Lange, 2015). Enfin, bien que ce modèle ne vise pas explicitement à

Chapitre 1 - Contexte théorique : Influence des mécanismes prédictifs sur la perception visuelle

expliquer comment les connaissances ou prédictions sont susceptibles d'influencer la perception, il permet de faire des hypothèses sur les mécanismes par lesquels les prédictions peuvent moduler le traitement des entrées sensorielles (Clark, 2013; Kok & de Lange, 2015; Walsh et al., 2020). Effectivement, lorsqu'un stimulus est conforme aux prédictions, l'erreur de prédiction est relativement faible voire absente, ce qui entrainera une activité cellulaire globalement moindre dans ces unités. De plus, le stimulus devrait être correctement représenté dans les unités de prédiction (cf. Figure 1.12.B). En revanche, lorsqu'un stimulus ne correspond pas aux prédictions, l'erreur de prédiction est plus importante, ce qui entrainera une activité cellulaire accrue dans les régions concernées. Par ailleurs, la représentation du stimulus dans les unités de prédictions devrait être moins stable et univoque. Le traitement de l'erreur de prédiction permettrait dans un deuxième temps de mettre à jour les prédictions jusqu'à ce qu'elles soient plus conformes au stimulus (cf. Figure 1.12.C).

Cependant, l'implémentation de ce modèle au niveau neurobiologique et les prédictions qu'il permet de faire concernant la représentation de stimuli conformes/non conformes avec les attentes au sein des aires cérébrales et la manière dont cela façonne in fine la perception restent débattus (Blank & Davis, 2016; de Lange et al., 2018).

2.3.2. Corrélats cérébraux du codage prédictif : Mécanismes de Sharpening vs. Dampening

La principale difficulté rencontrée dans les études visant à fournir des arguments empiriques en faveur du codage prédictif comme principe de traitement de l'information au niveau cérébral provient du fait que la plupart de ces études utilisent des méthodes de neuroimagerie (e.g., IRMf, MEG, EEG) qui permettent de mesurer indirectement l'activité globale d'une grande population de neurones mais n'ont pas la précision nécessaire pour dissocier par exemple les populations de neurones encodant les prédictions vs. l'erreur de prédiction – si tant est qu'une telle distinction existe au niveau neuronal. La question se pose alors de savoir si l'activité des aires visuelles – et in fine le contenu de la perception – devrait plutôt refléter les représentations des unités de prédiction ou d'erreur de prédiction (de Lange et al., 2018; Walsh et al., 2020).

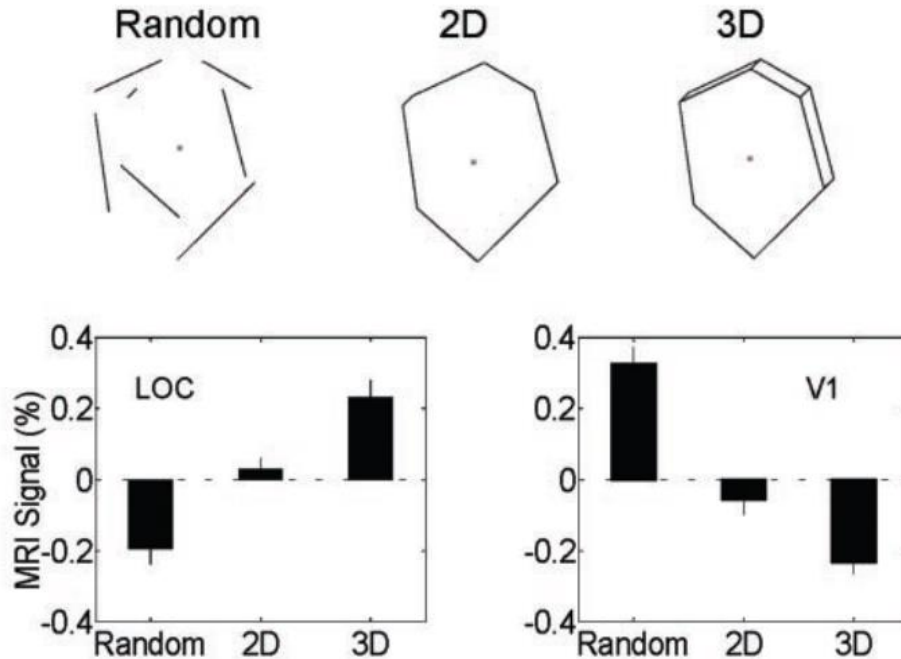
La plupart des études comparant l'activité cérébrale en réponse à des stimuli conformes aux prédictions vs. non-prédits ou non conformes aux prédictions rapportent que les stimuli prédits entraînent une activité moindre que les stimuli non-prédits ou non conformes avec les prédictions (phénomène de « suppression par les prédictions » ou « *expectation suppression* »

Chapitre 1 - Contexte théorique : Influence des mécanismes prédictifs sur la perception visuelle

; Alink et al., 2010; Kok, Jehee, et al., 2012; Kok, Rahnev, et al., 2012; Kok & de Lange, 2014; Kumar et al., 2017; Murray et al., 2002; Richter et al., 2018 ; voir pour revues de Lange et al., 2018; Walsh et al., 2020). Une telle observation semble compatible avec les modèles du codage prédictif : un stimulus conforme aux prédictions devrait entraîner moins d'erreur de prédiction et donc moins d'activité dans les aires de bas niveau. Dans une étude en IRMf, Murray et al. (2002) ont par exemple présenté aux participants des dessins représentant des objets en 3D, en 2D ou bien des figures dans lesquelles les traits composant les objets étaient arrangés de manière aléatoire (cf. Figure 1.13). Ces auteurs ont observé que plus le stimulus était élaboré (i.e., objet 3D), plus l'activité moyenne au sein du cortex visuel primaire diminuait. L'effet inverse était observé dans une région de plus haut niveau sélective aux objets, le complexe occipital latéral (LOC). Ces résultats ont été interprétés comme reflétant le fait que les objets 3D aient pu activer des représentations de haut niveau au sein du LOC, qui aurait à son tour pu générer des prédictions inhibitrices sur les caractéristiques de bas niveau du stimulus, résultant en une activité réduite au sein de V1. Au contraire, l'arrangement aléatoire des traits n'aurait pas permis d'activer des représentations de haut niveau au sein du LOC pouvant contraindre leur analyse au sein de V1, résultant en une augmentation de l'erreur de prédiction et donc une activité plus importante dans cette région. Selon cette interprétation, l'effet de suppression par les prédictions résulte donc de l'inhibition du traitement des caractéristiques prédites au sein de V1 dont l'activité reflèterait alors seulement l'erreur de prédiction concernant le stimulus.

Figure 1.13

Stimuli et Résultats de l'Expérience de Murray et al. (2002).



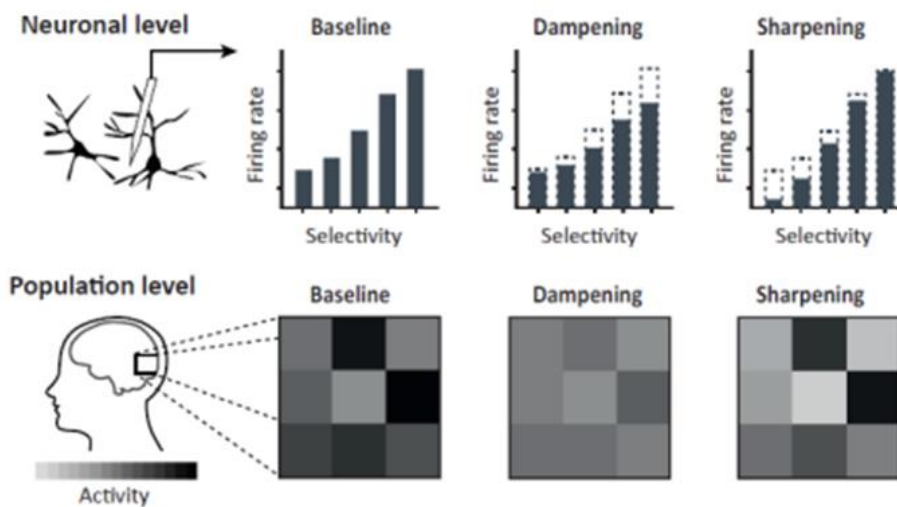
Note. Les stimuli utilisés étaient des dessins représentant des objets en 3D, en 2D ou des figures dans lesquelles les traits composant les objets étaient arrangés de manière aléatoire. Les résultats IRM ont montré que plus le stimulus était élaboré (i.e., objet 3D), plus l'activité moyenne au sein du cortex visuel primaire (V1) diminuait. L'effet inverse était observé dans une région de plus haut niveau sélective aux objets, le complexe occipital latéral (LOC). Figure adaptée de Murray et al. (2002).

Certains auteurs remettent néanmoins en question cette interprétation (e.g., Clark, 2013; Jiang & Rao, 2022; Kok, Jehee, et al., 2012; Kok & de Lange, 2015). Pour ces auteurs, l'effet de suppression par les prédictions au sein de V1 ne résulte pas nécessairement de l'inhibition du traitement des caractéristique prédites du stimulus et donc d'une diminution de l'erreur de prédiction mais pourrait aussi traduire la meilleure représentation du stimulus dans les unités de prédictions qui se trouvent tout le long de la hiérarchie des aires corticales. En effet, le modèle du codage prédictif implique que plus un stimulus est conforme aux prédictions (e.g., objet 3D), plus sa représentation sera précise dans les unités de prédiction, via l'activité sélective des neurones codant les caractéristiques prédites du stimulus dans ces unités. Au contraire, un stimulus non prédit devrait entraîner une représentation moins stable et univoque du stimulus dans les unités de prédictions, résultant en une réponse moins sélective des neurones et une réponse globalement plus élevée dans ces unités.

Ces deux interprétations alternatives de l'effet de suppression par les prédictions ont été désignées dans la littérature sous les termes de mécanisme de *dampening* (i.e., atténuation du traitement des caractéristiques prédites relativement aux caractéristiques non prédites) vs. de *sharpening* (i.e., rehaussement du traitement des caractéristiques prédites relativement aux caractéristiques non prédites, de Lange et al., 2018, cf. Figure 1.14).

Figure 1.14

Illustration des Mécanismes de Dampening et de Sharpening au Niveau Cérébral.



Note. Le mécanisme de dampening correspond à une diminution de l'activité des neurones codant les caractéristiques attendues des stimuli au profit d'une augmentation de la sélectivité des neurones codant les caractéristiques inattendues, résultant en une représentation corticale moins saillante des caractéristiques attendues. A l'inverse, le mécanisme de sharpening correspond à une diminution de l'activité des neurones codant les caractéristiques inattendues des stimuli au profit d'une augmentation de la sélectivité des neurones codant les caractéristiques attendues, résultant en une représentation corticale plus saillante des caractéristiques attendues. Dans les deux cas, un stimulus prédit entraîne une activité globalement moindre qu'un stimulus non prédit (suppression par les prédictions). Figure adaptée de de Lange et al. (2018).

Afin de départager les deux interprétations alternatives concernant l'effet de suppression par les prédictions, Kok, Jehee, et al. (2012) ont mené, en plus de l'étude comportementale présentée précédemment, une étude IRMf dans laquelle ils présentaient aux participants des patches de Gabor faiblement contrastés et dont l'orientation pouvait être prédite sur la base d'un indice sonore précédant leur apparition. Dans une partie des essais, l'orientation des patches

Chapitre 1 - Contexte théorique : Influence des mécanismes prédictifs sur la perception visuelle

n'était cependant pas conforme aux prédictions. Les participants devaient effectuer une tâche de discrimination d'orientation entre deux patchs (décider si le second patch était orienté de façon horaire ou anti-horaire par rapport au premier). Pour répondre à leur objectif de recherche principal, les auteurs se sont intéressés non seulement à l'activité moyenne induite par la présentation des patchs de Gabor dont l'orientation pouvait être attendue ou inattendue au sein du cortex visuel primaire, mais également aux patterns spatiaux d'activation dans cette région (analyse de patterns multivariée - MVPA). Le raisonnement était le suivant : si l'activité du cortex visuel reflète essentiellement l'erreur de prédiction (mécanisme de dampening), les patterns spatiaux d'activité devraient contenir des informations plus représentatives de l'orientation du stimulus lorsque celle-ci est non conforme que conforme aux prédictions. Au contraire, si l'activité du cortex visuel reflète essentiellement la réponse plus sélective aux caractéristiques prédites dans les unités de prédictions (mécanisme de sharpening), les patterns spatiaux d'activité devraient contenir des informations plus représentatives de l'orientation du stimulus lorsque celle-ci est conforme que non conforme aux prédictions. Pour tester ces hypothèses, les auteurs ont entraîné un algorithme (classifieur) à classer les patterns d'activité cérébrale induits par la présentation de patchs de Gabor selon l'orientation de ceux-ci. Ils ont ensuite testé les capacités du classifieur à décoder l'orientation de patchs de Gabor en lui fournissant des patterns d'activation non utilisés pour l'entraînement. Les résultats ont tout d'abord montré un effet de suppression par les prédictions : l'activité moyenne dans le cortex visuel primaire était globalement plus faible lorsque l'orientation du réseau était conforme que non conforme aux prédictions. Plus important, les analyses de classification ont montré que les performances du classifieur pour décoder l'orientation des patchs de Gabor étaient meilleures lorsque l'orientation était conforme que non conforme aux prédictions, ce qui est en accord avec le mécanisme de sharpening. De plus, les performances de classification étaient corrélées positivement avec les performances comportementales des participants, suggérant que l'amélioration de la représentation des stimuli prédits au niveau cortical se traduit également par une meilleure perception de ces stimuli au niveau comportemental. Ces résultats ont été confortés par plusieurs études utilisant une méthode similaire d'analyse de patterns d'activité des signaux cérébraux (en IRMf ou en MEG comme par exemple : Aitken et al., 2020; Brandman & Peelen, 2017; Heilbron et al., 2020; Kok & de Lange, 2014; Yon et al., 2018). Dans une étude combinant des enregistrements en MEG et IRMf, Brandman et Peelen (2017) ont par exemple montré que la catégorie d'objets dégradés (pixelisés) était décodée avec de meilleures performances au sein des aires visuelles impliquées dans la reconnaissance d'objets

Chapitre 1 - Contexte théorique : Influence des mécanismes prédictifs sur la perception visuelle

lorsque ces objets pouvaient être prédits sur la base de leur contexte que lorsqu'ils étaient présentés de façon isolée. Cela était associé à de meilleures performances comportementales lors d'une tâche de catégorisation de ces mêmes objets. De façon similaire, Heilbron et al. (2020) ont montré que la nature de lettres (« u » ou « n ») présentées visuellement au milieu d'autres lettres et mêlées à du bruit était décodée avec de meilleures performances sur la base des patterns d'activité IRMf du cortex visuel primaire lorsque les chaînes de lettres formaient un mot (permettant de générer des prédictions sur la lettre centrale) qu'un pseudo-mot (ne permettant pas de générer des prédictions sur la lettre centrale).

Toutefois, un certain nombre de travaux utilisant pourtant des méthodes d'analyse cérébrale similaires rapportent des résultats opposés (e.g., Kumar et al., 2017; Richter et al., 2018; Yan et al., 2023). Dans une étude en IRMf menée chez le macaque, Kumar et al. (2017) ont par exemple examiné les patterns d'activité de neurones du cortex inférotemporal en réponse à la présentation d'objets attendus ou inattendus. Dans un premier temps, les macaques étaient familiarisés avec des séquences d'images d'objets dont l'ordre était fixé et ils apprenaient implicitement l'ordre d'apparition des objets dans ces séquences (apprentissage de régularités statistiques). Dans un deuxième temps, on leur présentait les objets vus pendant l'entraînement soit selon la même séquence (l'apparition d'un objet pouvait donc prédire l'objet suivant), soit dans un ordre aléatoire (l'apparition d'un objet prédisait alors de manière erronée l'objet suivant). Les résultats ont montré que l'identité de l'objet était décodée avec de meilleures performances au sein du cortex inférotemporal lorsque celui-ci était prédit de manière erronée par l'objet précédent que de manière correcte. Cela suggère donc une meilleure représentation des stimuli au niveau cortical lorsqu'ils ne sont pas correctement prédits. Dans une étude en IRMf conduite chez l'humain, Richter et al. (2018), ont employé un paradigme similaire d'apprentissage statistique dans lequel des séquences de deux images d'objets étaient présentées. Les résultats ont tout d'abord mis en évidence une réduction de l'activité neuronale dans V1 et dans des régions de la voie ventrale sélectives au traitement des objets pour les paires attendues par rapport aux paires inattendues (effet de suppression par les prédictions). Les auteurs ont également examiné l'effet de suppression par les prédictions en fonction de la sélectivité des voxels des aires visuelles aux différentes images d'objets, mesurée lors de la phase d'apprentissage. Ces analyses ont montré que les voxels du LOC les plus sélectifs à une image donnée (i.e., les voxels dont l'activité différait le plus pour cette image relativement à une autre) étaient aussi ceux pour lesquels l'effet de suppression par les prédictions pour cette image était le plus important (i.e., dont la réponse était la plus réduite lorsque cette image était

Chapitre 1 - Contexte théorique : Influence des mécanismes prédictifs sur la perception visuelle

attendue et effectivement présentée). Ces résultats ont donc été interprétés comme suggérant une atténuation sélective des populations de neurones codant les caractéristiques prédites du stimulus, en accord avec le mécanisme de dampening.

Dans l'ensemble, les études portant sur la représentation des stimuli prédits/attendus ou non prédits/inattendus au niveau cérébral rapportent donc des résultats contradictoires. On peut noter que les résultats de toutes ces études sont néanmoins tous interprétés en faveur de la théorie du codage prédictif, comme reflétant soit l'activité des unités d'erreur de prédiction (atténuation du traitement des caractéristiques prédites - dampening), soit l'activité des unités de prédictions (représentation plus précise des caractéristiques prédites - sharpening). Cela souligne d'ailleurs la principale limite de ce modèle : dans la mesure où des résultats opposés peuvent être compatibles avec le modèle, celui-ci est difficilement falsifiable. Ces effets contradictoires au niveau cérébral font par ailleurs écho aux différences observées dans les études comportementales rapportées dans la première partie de cette thèse qui se sont intéressées à l'effet des connaissances a priori sur la perception visuelle. Bien que celles-ci n'ont pas nécessairement été réalisées dans le contexte théorique du codage prédictif, les études suggérant un rehaussement perceptif des stimuli pouvant être prédits ou conformes avec les connaissances (e.g., Chalk et al., 2010; Han & VanRullen, 2016; Hansen et al., 2006; Lupyan, 2015, 2017; Perez et al., 2020) pourraient refléter les conséquences perceptives du mécanisme de sharpening, tandis que les études allant dans le sens d'une atténuation perceptuelle au profit des stimuli non prédits ou non conformes avec les connaissances (e.g., Denison et al., 2016; Spaak et al., 2022) pourraient au contraire refléter les conséquences perceptives du mécanisme de dampening. Dans la suite de cette thèse, nous utiliserons les termes de sharpening et dampening pour faire référence respectivement au rehaussement ou à l'atténuation perceptuelle du traitement au niveau cérébral des caractéristiques prédites relativement aux caractéristiques non-prédites.

Plusieurs explications ont été proposées pour expliquer les effets contradictoires des prédictions qui ont été observés dans les études susmentionnées (e.g., de Lange et al., 2018; Richter et al., 2018). Tout d'abord, il est généralement suggéré que les mécanismes de sharpening et dampening coexistent mais que les corrélats neurophysiologiques des unités de prédiction (dans lesquelles se produirait le mécanisme de sharpening) et d'erreur de prédiction (dans lesquelles se produirait le mécanisme de dampening) sont localisés dans des couches séparées d'une aire visuelle. Cependant, dans les études citées ci-dessus, l'activité d'une aire est mesurée au niveau global, toutes couches comprises. Or, rien ne permet d'expliquer

Chapitre 1 - Contexte théorique : Influence des mécanismes prédictifs sur la perception visuelle

pourquoi l'activité observée au niveau global reflèterait de façon prédominante le mécanisme de sharpening ou de dampening (et donc l'activité de l'une ou l'autre de ces couches). Il a également été proposé que les effets des prédictions au niveau cérébral et comportemental pourraient dépendre de la manière dont ces prédictions ont été générées. En effet, les résultats en faveur du mécanisme de dampening se retrouveraient plutôt dans les études où les prédictions sont générées par un apprentissage statistique réalisé au début de l'expérience (e.g., Denison et al., 2016; Richter et al., 2018) tandis que ceux en faveur du sharpening sont observés dans des études où les prédictions sont basées sur les connaissances a priori et/ou issues du stimulus (e.g., Brandman & Peelen, 2017; Heilbron et al., 2020; Lupyan, 2017; Perez et al., 2020). Cependant, les raisons pour lesquelles ces différents types d'induction des prédictions conduiraient à des résultats opposés ne sont pas spécifiées. Une autre piste serait que les études en faveur d'un mécanisme de sharpening pourraient s'expliquer par le fait que l'effet des prédictions dans ces études reflète en réalité un effet de l'attention, connue pour entraîner également un rehaussement perceptif et neuronal (e.g., Carrasco et al., 2004; Serences et al., 2009). Par exemple, dans l'étude de Kok, Jehee, et al. (2012), l'orientation des patchs de Gabor était à la fois la caractéristique sur laquelle pouvait porter les prédictions et la caractéristique sur laquelle portait la tâche de discrimination – et donc l'attention – des participants. Cependant là encore, les mécanismes exacts par lesquels l'attention portée aux stimuli aurait pu inverser les effets des prédictions n'ont pas été spécifiés. De plus, des effets en faveur du mécanisme de sharpening ont été observés dans des études où la tâche des participants ne portait pas explicitement sur les caractéristiques prédites/non prédites des stimuli (e.g., Brandman & Peelen, 2017; Lupyan, 2017).

Dans l'ensemble, les explications évoquées ci-dessus ne sont donc pas satisfaisantes pour rendre compte des différents résultats observés dans ces études. Les formulations plus récentes du modèle du codage prédictif, ainsi qu'une théorie proposée récemment, l'*opposing process theory* (Press et al., 2020), offrent cependant des pistes intéressantes pour résoudre ces questions.

2.3.3. Modulation du poids relatif des prédictions/de l'erreur de prédiction et opposing process theory

Les formulations les plus récentes du modèle du codage prédictif (Feldman & Friston, 2010; Friston, 2005) intègrent un aspect qui était absent des premières formulations (e.g., Rao & Ballard, 1999), celui d'une modulation du poids relatif des signaux de prédiction et d'erreur

Chapitre 1 - Contexte théorique : Influence des mécanismes prédictifs sur la perception visuelle

de prédiction. En effet, ces nouveaux modèles supposent que pour que le codage prédictif soit efficace, les signaux de prédiction et d'erreur de prédiction ne doivent pas être traités de manière équivalente dans toutes les conditions de vision. Par exemple, lorsque les signaux sensoriels ne sont pas conformes aux prédictions, cela pourrait résulter du fait que les prédictions sont erronées mais aussi du fait que les signaux sensoriels sont tout simplement bruités et donc peu fiables. Dans ce cas, accorder un poids trop important aux signaux d'erreur de prédiction ne serait pas efficace pour traiter le stimulus. Par contre, lorsque le signal sensoriel est fiable et non ambigu, accorder un poids trop important aux prédictions serait en revanche contre-productif. Ainsi, les modèles plus récents du codage prédictif tiennent compte de cela en proposant que le poids relatif des prédictions/erreur de prédiction puisse être modulé par l'estimation de leur fiabilité (ou « précision ») respective. Ainsi, les signaux d'erreur de prédiction auraient plus de poids lorsque le signal est fiable et non conforme aux prédictions (mécanisme de dampening), permettant ainsi de les mettre à jour, tandis qu'un poids plus important serait accordé aux signaux de prédiction si les prédictions sont robustes et que le signal sensoriel est bruité (mécanisme de sharpening).

Ce principe de pondération des signaux de prédictions et d'erreur de prédiction selon leur précision pourrait en partie expliquer les effets contradictoires observés dans les études précédemment citées. En effet, la plupart des études mettant en évidence une meilleure représentation cérébrale ou une meilleure perception de stimuli prédits que non prédits utilisaient généralement des stimuli bruités ou ambigus (e.g., des stimuli flous ; Cheadle et al., 2015; Lupyan, 2017; Perez et al., 2020; des stimuli faiblement contrastés ; Kok, Jehee, et al., 2012; Stein & Peelen, 2015; Wyart et al., 2012; des stimuli pixelisés ; Brandman & Peelen, 2017). Ces résultats pourraient donc refléter le poids plus important attribué aux prédictions qu'à l'erreur de prédiction, résultant en une représentation cérébrale et une perception dominée par ce qui est conforme aux prédictions. Au contraire, les études ayant observé l'atténuation au niveau cérébral ou perceptif du traitement des caractéristiques prédites au profit des caractéristiques non prédites utilisaient des stimuli fiables et non ambigus (e.g., des images nettes de scènes et d'objets ; Denison et al., 2016; Kumar et al., 2017; Richter et al., 2018; Spaak et al., 2022). Les résultats de ces études pourraient alors refléter le poids plus important attribué à l'erreur de prédiction, résultant en une représentation cérébrale et une perception dominée par ce qui est inattendu.

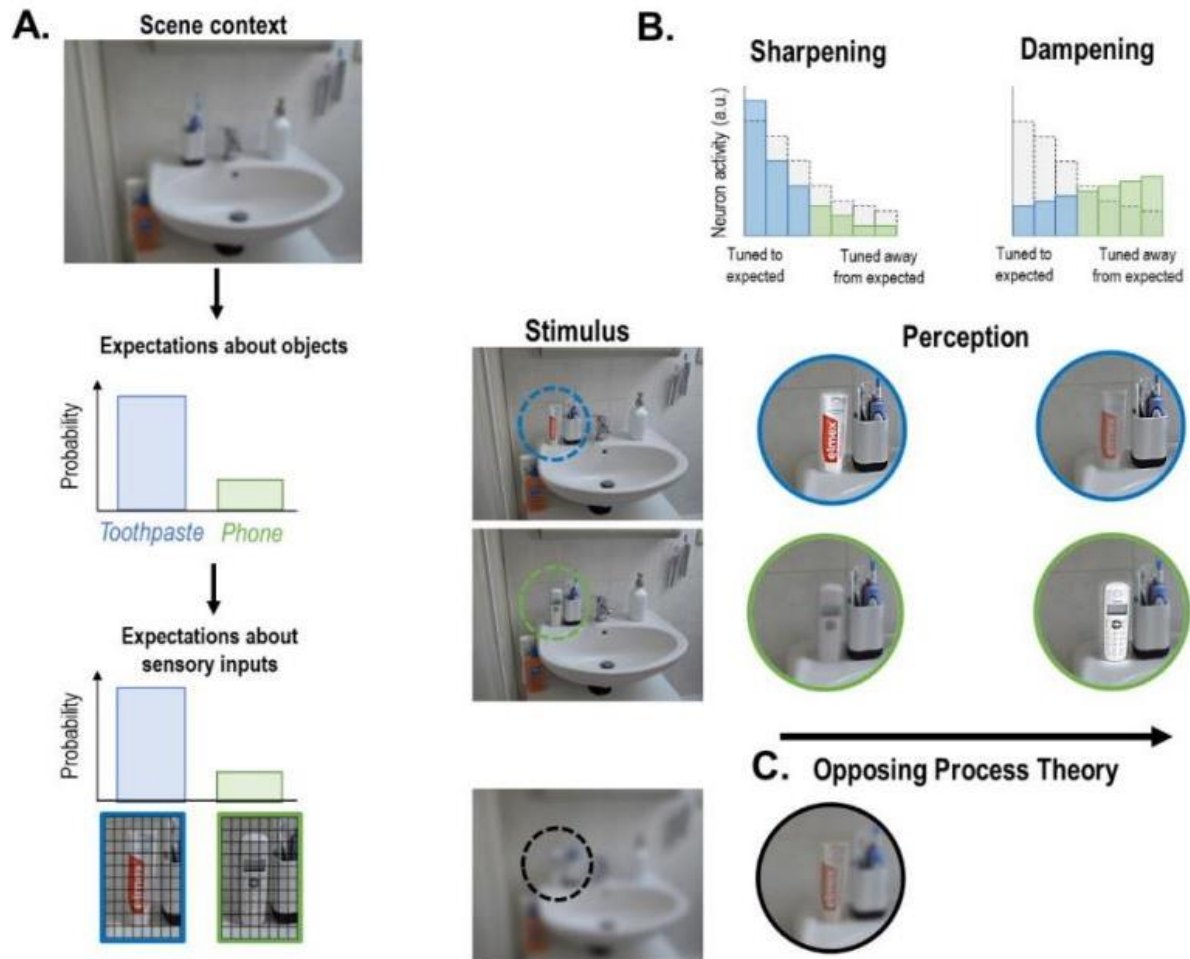
Cette idée que le poids relatif attribué aux prédictions et au signal entrant (i.e., erreur de prédiction) est modulé en fonction de leur fiabilité respective est également sous-jacente à une

Chapitre 1 - Contexte théorique : Influence des mécanismes prédictifs sur la perception visuelle

théorie développée plus récemment, l'opposing process theory (Press et al., 2020), qui vise à réconcilier les résultats opposés qu'on observe dans les travaux antérieurs, notamment dans les études comportementales. Cette théorie suppose que les mécanismes de sharpening et de dampening coexistent et qu'ils peuvent intervenir alternativement en fonction du décours temporel du traitement de l'information visuelle et des contraintes du signal (Figure 1.15). Selon cette conception, la perception serait initialement biaisée en faveur de ce qui est attendu (mécanisme de sharpening), permettant ainsi de confirmer les attentes valides. Cela permettrait, dans un premier temps, d'extraire du signal ce qui est conforme aux prédictions. Si le stimulus est tout à fait conforme aux prédictions, aucun autre mécanisme n'intervient. Ce principe de traitement s'avèrerait utile et efficace la plupart du temps du fait des régularités statistiques de l'environnement. Cependant, si certaines caractéristiques du stimulus sont très éloignées des prédictions, le poids accordé au traitement de celles-ci serait alors favorisé au détriment des caractéristiques prédites (mécanisme de dampening). Cela permettrait donc de privilégier, dans un deuxième temps, le traitement de ce qui n'est pas prédit, et donc potentiellement informatif pour mettre à jour les connaissances. Ce postulat est d'ailleurs compatible avec les modèles du codage prédictif (cf. Figure 1.12). Cette séquence ne serait cependant valable que si les signaux sensoriels sont fiables et non ambigus. Lorsque les entrées sensorielles sont ambiguës ou bruitées, alors toute caractéristique non conforme aux prédictions pourrait être attribuable à du bruit. Dans ce cas, il serait plus pertinent de privilégier le traitement de ce qui est conforme aux prédictions (mécanisme de sharpening) qui se révélera véridique dans la plupart des cas. Ce principe de traitement permettrait donc d'optimiser le traitement de l'information visuelle de façon à aboutir à la génération d'un percept qui soit à la fois véridique et informatif (i.e., qui soit le plus proche possible de l'état objectif de l'environnement mais qui nous apporte des informations sur notre environnement que nous ne savons pas déjà). Bien qu'elle s'appuie essentiellement sur des études comportementales et ne se situe pas explicitement dans le contexte du modèle du codage prédictif, cette théorie en partage néanmoins la plupart des aspects. On peut cependant noter que contrairement au modèle du codage prédictif, cette théorie ne suppose pas nécessairement que le rehaussement perceptif des caractéristiques inattendues lorsque le signal est net résulte de l'inhibition des caractéristiques prédites mais qu'il pourrait aussi simplement résulter d'un poids plus important accordé aux caractéristiques non-prédites bénéficiant d'une allocation attentionnelle plus importante (sans pour autant inhiber le traitement des caractéristiques prédites).

Figure 1.15

Illustration de l'Influence des Prédications au Niveau Perceptif.



Note. A. Exemple de prédictions basées sur les régularités statistiques de l'environnement visuel. B. Illustration des mécanismes de dampening et de sharpening au niveau cérébral et perceptif. C. Illustration de l'opposing process theory. Lorsque le signal visuel est fiable, le mécanisme de sharpening précède le mécanisme de dampening. Lorsque le signal visuel est ambigu, le mécanisme de sharpening prédomine indépendamment du décours temporel.

L'opposing process theory a notamment été développée sur la base des résultats d'expériences de psychophysique (Yon & Press, 2017). Dans cette étude, les participants effectuaient des mouvements d'abduction de leur index ou de leur majeur tout en observant des images de luminosité variable montrant un mouvement d'abduction de l'index ou du majeur. Les mouvements de doigts des participants étaient donc soit congruents (i.e., mouvement du même doigt), soit incongruents (i.e., mouvement d'un doigt différent) avec les images

Chapitre 1 - Contexte théorique : Influence des mécanismes prédictifs sur la perception visuelle

présentées. La tâche consistait à juger la luminosité des images présentées à différents délais après l'action des participants (50 ms, 200 ms et 350 ms). Les résultats ont montré que les stimuli congruents étaient perçus comme étant plus lumineux à un délai de 50 ms, mais cet effet s'inversait en faveur d'une perception des stimuli incongruents comme plus lumineux à un délai de 200 ms, puis aucun effet n'était plus observé pour un délai de 350 ms. Cette étude mettant en évidence une alternance des mécanismes de sharpening et de dampening en fonction du décours temporel du traitement visuel a été le point de départ de l'élaboration de l'opposing process theory.

Cependant, les hypothèses formulées par l'opposing process theory concernant le rôle que joue la fiabilité du signal (son ambiguïté) sur la façon dont les prédictions influencent la perception subjective n'ont, à notre connaissance, jamais été systématiquement abordées au sein d'une même étude. En effet, les études susmentionnées portant sur les effets perceptifs des prédictions ont soit utilisé des stimuli ambigus (e.g., flous ; Lupyan, 2017; Perez et al., 2020), soit des stimuli non ambigus (e.g., Spaak et al., 2022), mais elles n'ont pas directement comparé les deux conditions. De plus, les études utilisant des stimuli ambigus ne manipulaient généralement que la présence ou l'absence de connaissances préalables ou d'informations pertinentes permettant de former des prédictions. Cependant, elles n'ont pas abordé l'effet de la validité des attentes en incluant une condition dans laquelle le signal visuel pouvait être surprenant et aller à l'encontre des attentes (i.e., un stimulus ayant du sens mais étant incohérent/surprenant).

L'étude de ces mécanismes est pourtant importante pour mieux comprendre les déterminants de la perception subjective des stimuli visuels attendus/inattendus sous diverses contraintes et comment cela peut à son tour influencer le comportement. Par exemple, dans des situations quotidiennes telles que la conduite automobile, où les attentes et l'anticipation sont cruciales et où l'information dans le champ visuel devient plus bruitée avec la vitesse, un stimulus inattendu dans la périphérie serait-il atténué et donc ignoré, non détecté et/ou mal compris ? Répondre à ces questions a également des implications cliniques car cela permettrait par exemple de mieux identifier dans quelle mesure les déficits sensoriels liés aux pathologies visuelles ou au vieillissement peuvent être compensés ou au contraire amplifiés par ces mécanismes prédictifs lors du traitement de stimuli attendus et inattendus, ou de mieux comprendre les conséquences perceptives de leurs dysfonctionnements comme suggéré dans divers troubles (e.g., troubles du spectre autistique, psychose ; Sterzer et al., 2018; Van de Cruys et al., 2014).

3. Objectif et problématique de la thèse

La plupart des modèles actuels de la perception visuelle s'accordent sur l'idée que la perception est un processus proactif. Nos connaissances acquises sur l'environnement visuel nous conduiraient à générer en permanence des attentes ou prédictions sur ses caractéristiques les plus probables, afin d'en faciliter le traitement. Cependant, dans quelle mesure et par quels mécanismes ces connaissances et ces attentes influencent et façonnent in fine la perception reste mal compris.

Si de nombreux travaux suggèrent une influence des connaissances a priori sur le traitement et la reconnaissance de stimuli complexes tels que les scènes et les objets, ces études se sont principalement concentrées sur des mesures telles que les temps de réaction ou les taux d'erreurs lors de tâches de reconnaissance (Brandman & Peelen, 2017; Davenport, 2007; Davenport & Potter, 2004; Greene et al., 2015; Joubert et al., 2008; Lai et al., 2020; Palmer, 1975; Roux-Sibilon et al., 2019) qui ne permettent pas de rendre compte de la phénoménologie de la perception. La question de savoir si les prédictions affectent l'apparence subjective du stimulus (comment nous voyons) a été traitée par un nombre limité d'études et les stimuli utilisés dans ces travaux étaient assez peu écologiques. Effectivement, il s'agissait de patches de Gabor (Teufel et al., 2018), de dessins d'objets (Perez et al., 2020) ou encore de mots (Lupyan, 2017). Les résultats de ces études ne permettent donc pas de savoir si les mécanismes prédictifs influencent la perception subjective des stimuli plus complexes et écologiques.

Par ailleurs, la plupart des études sur l'influence des connaissances sur la perception subjective rapportent des résultats contradictoires. Alors que certaines études suggèrent une amélioration perceptive de ce qui est conforme aux prédictions (i.e., sharpening), d'autres observent au contraire leur atténuation perceptive, au profit de ce qui n'est pas conforme aux prédictions (i.e., dampening). Des développements récents des modèles du codage prédictif apportent toutefois des pistes intéressantes pour réconcilier ces résultats opposés en postulant que le poids relatif attribué aux prédictions et au signal entrant est modulé en fonction de leur fiabilité respective. Cette idée a également été reprise et étendue par une théorie très récente, l'opposing process theory (Press et al., 2020), qui suppose que les mécanismes de sharpening et de dampening coexistent et qu'ils peuvent intervenir alternativement en fonction du décalage temporel du traitement de l'information visuelle et de la fiabilité du signal.

Malgré cela, plusieurs questions restent en suspens quant à l'impact des prédictions sur l'apparence subjective des stimuli. Tout d'abord, la question de savoir si les mécanismes

Chapitre 1 - Contexte théorique : Influence des mécanismes prédictifs sur la perception visuelle

prédictifs influencent la perception de stimuli visuels complexes et écologiques que nous pouvons rencontrer au quotidien n'a, à notre connaissance, pas été abordée. Il s'agit d'une question importante permettant de déterminer si mécanismes de sharpening et de dampening sont effectivement des principes généraux de la perception visuelle. De plus, les études qui se sont intéressées à l'impact des prédictions sur l'apparence subjective des stimuli ont été assez critiquées de par le fait que les méthodes utilisées ne permettaient pas de s'assurer que les effets observés étaient bien des effets perceptifs et non pas des effets non-perceptifs tels que des effets décisionnels par exemple. Même s'il n'existe peut-être pas de perception pure que l'on peut mesurer, on peut toutefois se donner l'objectif d'étudier l'influence des connaissances et des attentes sur la perception visuelle en minimisant au maximum l'influence d'autres facteurs supposément non-perceptifs. Ensuite, les mécanismes par lesquels les prédictions influencent la perception restent assez méconnus. En effet, peu d'études se sont intéressées à la modulation de effets des prédictions sur la perception selon la fiabilité du signal, telle que proposée par les théories du codage prédictif et l'opposing process theory. D'ailleurs, les hypothèses formulées récemment par l'opposing process theory n'ont pas été testées systématiquement au sein d'une même étude. Il est donc nécessaire d'y remédier afin de déterminer si les effets contradictoires des prédictions observés dans les travaux antérieurs peuvent être réconciliés en tenant compte de la fiabilité du signal visuel et des contraintes de temps de traitement.

Le premier objectif de ce travail de thèse a donc été de déterminer dans quelle mesure les attentes ou prédictions basées sur nos connaissances et expériences antérieures avec notre environnement visuel peuvent affecter l'apparence subjective des stimuli visuels complexes tels que les scènes et les objets. **Le second objectif de cette thèse** était de comprendre quels sont les mécanismes sous-jacents aux effets des prédictions sur la perception visuelle, en nous basant sur les hypothèses proposées par les théories du codage prédictif et l'opposing process theory.

Pour tester notre premier objectif, nous avons tout d'abord mené trois expériences comportementales afin de déterminer si la possibilité de générer des prédictions grâce aux informations contextuelles ou à notre expérience antérieure influençait la perception subjective de la netteté d'objets (Expérience 1) et de scènes (Expériences 2 et 3), et ce en contrôlant que les effets observés n'étaient pas dus à des biais non-perceptifs. Nous nous sommes ensuite concentrées plus particulièrement sur le second objectif de cette thèse. Pour ce faire, nous avons tout d'abord mené une quatrième expérience comportementale (Expérience 4) afin de préciser dans quelle mesure l'effet des prédictions sur la netteté perçue varie en fonction de la fiabilité (ou ambiguïté) du signal visuel, conformément aux théories du codage prédictif et de l'opposing

Chapitre 1 - Contexte théorique : Influence des mécanismes prédictifs sur la perception visuelle

process theory postulant que l'influence des prédictions sur la perception visuelle augmenterait à mesure que la fiabilité du signal diminue.

Les quatre premières expériences de cette thèse (Expériences 1, 2, 3 et 4) manipulaient la possibilité de générer ou non des prédictions et ne permettaient donc pas d'arbitrer entre les mécanismes de sharpening (i.e., plus grande sensibilité aux stimuli attendus) et de dampening (i.e., plus grande sensibilité aux stimuli inattendus/surprenants) dans la mesure où elles n'incluaient pas des stimuli inattendus/surprenants pouvant aller à l'encontre des attentes. Nous avons donc construit de tels stimuli dans deux dernières expériences comportementales (Expériences 5 et 6) dans le but de tester les conditions d'application des mécanismes de sharpening et de dampening et de tester les hypothèses proposées par l'opposing process theory (Press et al., 2020). Pour rappel, cette théorie suggère que ces deux mécanismes puissent intervenir alternativement en fonction du décours temporel et de la fiabilité du signal visuel. Le mécanisme de sharpening prédominerait lorsque le signal est bruité mais que nous avons de fortes prédictions à son sujet, tandis que le mécanisme de dampening ne prendrait le dessus que lorsqu'un signal fiable diverge fortement des prédictions (comme dans le cas de stimuli surprenants). Nous avons mis en place l'Expérience 5 afin de tester explicitement cette hypothèse. Par ailleurs, cette séquence serait modulée par le décours temporel. À savoir que la perception serait initialement biaisée vers ce qui est attendu (sharpening), permettant de confirmer les prédictions. Le mécanisme de dampening prendrait ensuite le relais pour stimuler le traitement des autres caractéristiques inattendues. Cette variabilité de l'influence des attentes selon le décours temporel a donc été examinée dans l'Expérience 6.

Chapitre 2 – Contribution empirique

L'ensemble des stimuli, scripts expérimentaux, données recueillies et scripts d'analyses statistiques de chacune des expériences réalisées dans le cadre de cette thèse sont disponibles dans l'open science framework :

https://osf.io/z8vu6/?view_only=6373c08690d64a48bcb3106b5cd1046b.

Cette thèse a été menée en conformité avec le code éthique de l'American Psychological Association (2017) pour protéger les personnes qui se prêtent à des recherches. Tous les participants ont donné leur consentement écrit et l'ensemble de nos expériences ont été approuvées par le comité d'éthique de l'Université Grenoble Alpes (CER-Grenoble Alpes, CERGA-Avis-2020-4).

1. Etude de l'influence des prédictions sur la perception subjective de scènes et d'objets

N.B. Dans cette partie expérimentale, nous reportons les résultats des trois premières expériences publiées dans la revue *Journal of Experimental Psychology : Human Perception and Performance* (Rossel et al., 2022 ; Annexe A).

Objectif des Expériences 1, 2 et 3

L'objectif des trois premières expériences était de déterminer si la possibilité de générer des prédictions sur la base d'informations contextuelles ou de notre expérience antérieure influençait l'apparence visuelle subjective d'objets en contexte (Expérience 1) et de scènes (Expériences 2 et 3).

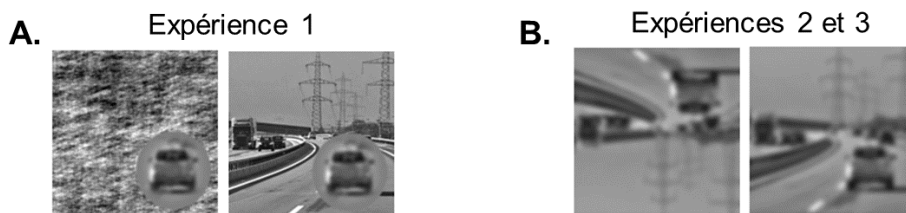
Pour ce faire, nous avons utilisé une tâche d'ajustement perceptif inspirée du paradigme de Lupyan (2017) pour les Expériences 1 et 2. Cette tâche semble particulièrement pertinente pour aborder les conséquences des prédictions sur la perception subjective car elle permet de comparer directement les jugements perceptifs subjectifs des participants à des paramètres objectifs dans le signal visuel. Toute différence qui en résulte est donc plus susceptible de refléter des effets perceptifs qualitatifs, contrairement à des mesures telles que les temps de réaction ou l'exactitude de la reconnaissance. Dans nos expériences, deux images floues, target

Chapitre 2 – Contribution empirique

et sample, étaient affichées côte à côte sur un écran d'ordinateur. À chaque essai, l'image target était présentée à un niveau de flou qui restait constant pendant tout l'essai mais qui variait entre les essais, allant d'une image assez nette et facilement identifiable à une image très floue et difficilement reconnaissable selon les essais. L'image sample avait toujours un niveau de flou moyen au début de l'essai et la tâche des participants consistait à augmenter ou diminuer manuellement son niveau de flou pour qu'il corresponde au niveau de flou de l'image target. Les participants étaient encouragés à se baser sur leur perception subjective pour faire correspondre le niveau de flou des images (i.e., « Ajustez le flou de l'image sample jusqu'à ce qu'il vous paraisse aussi flou que l'image target. »).

Figure 2.1

Stimuli des Expériences 1, 2 et 3.



Note. A. Exemple de stimuli utilisés dans l'Expérience 1. A gauche, un objet entouré d'un contexte de scène non informatif, construit en aléatoirisant le spectre de phase. A droite, un objet entouré d'un contexte de scène informatif. B. Exemple de stimuli utilisés dans les Expériences 2 et 3. A gauche, une scène à l'envers et à droite, une scène à l'endroit.

Comme dans Lupyan (2017), nous avons manipulé la disponibilité d'informations pertinentes pour générer des prédictions sur le contenu de l'image. Dans l'Expérience 1, nous avons étudié l'effet des prédictions pouvant être générées sur la base des informations contextuelles de scènes sur la perception d'objets. Pour cela, nous avons manipulé l'information contextuelle présente dans les images (Figure 2.1). La photographie d'un objet était entourée soit de la scène dans laquelle il a été photographié (contexte informatif), soit de cette même scène mais dont le spectre de phase avait été aléatoirisé (i.e., *scramble* de phase). Cette dernière manipulation nous permettait de supprimer l'information sémantique du contexte (contexte non informatif), mais de conserver ses propriétés de bas niveau (i.e., distribution des orientations et des fréquences spatiales) de la scène. Les images sample et target contenaient toujours le même

Chapitre 2 – Contribution empirique

objet mais dans un contexte différent (paires target-sample différentes). Dans l'Expérience 2, nous avons étudié l'effet des prédictions sur la netteté perçue d'une scène entière en manipulant la conformité de la disposition de la scène avec notre expérience antérieure (Figure 2.1). La scène était présentée soit à l'endroit (disposition prédictible de la scène, c'est-à-dire conforme à l'expérience antérieure), soit à l'envers (disposition non prédictible de la scène, c'est-à-dire non conforme à l'expérience antérieure). Les images sample et target étaient des images de la même scène mais avec une orientation différente (paires target-sample différentes). Sur la base des résultats de Lupyan (2017), nous nous attendions à ce qu'à niveau de flou objectivement égal, les objets (Expérience 1) ou les scènes (Expérience 2) soient subjectivement perçus comme plus nets dans des conditions prédictibles que dans des conditions non prédictibles. Pour compenser cette différence de flou perçue, nous nous attendions à ce que les participants attribuent un niveau de flou plus élevé à une image sample prédictible appariée à une image target non prédictible qu'à une image sample non prédictible appariée à une image target prédictible. Il est néanmoins possible qu'un tel résultat s'explique par un biais de réponse (Firestone & Scholl, 2016) plutôt qu'un biais perceptif en faveur des stimuli attendus. Par exemple, un objet présenté dans un contexte brouillé (Expérience 1) ou une scène à l'envers difficilement identifiable (Expérience 2) pourraient être plus facilement associées à quelque chose de « flou » qu'un objet dans un contexte intact ou une scène à l'endroit sans que cela ne se traduise réellement par une différence de netteté perçue.

Afin d'exclure cette possibilité, les Expériences 1 et 2 comprenaient également une condition contrôle où la prédictibilité des images target et sample était identique (i.e., deux objets en contexte intact ou deux objets en contexte brouillé pour l'Expérience 1 ; deux scènes à l'endroit ou deux scènes à l'envers pour l'Expérience 2). Il s'agit tout simplement de paires de stimuli identiques. Si les prédictions influencent effectivement la perception subjective des objets et scènes, ces effets perceptifs devraient s'appliquer de la même façon à tous les stimuli présents à l'écran (donc aux deux stimuli de la paire). Par exemple, si un objet prédictible dans un contexte donné est perçu comme plus net qu'il ne l'est en réalité, le même objet dans le même contexte présenté à côté devraient également bénéficier de ce rehaussement perceptif, ce qui se traduirait par l'absence de différence de flou perçue entre les deux objets, qu'ils soient prédictibles ou non prédictibles (cf. raisonnement du El Greco fallacy, Firestone & Scholl, 2014). Dans cette condition où les stimuli de la paire sample-target sont identiques, les participants devraient donc ajuster correctement le niveau de flou des images sample prédictibles et non prédictibles aux images target identiques (i.e., à un niveau de flou moyen de

Chapitre 2 – Contribution empirique

15.5), conduisant à une absence de différence entre la condition sample prédictible (target prédictible) et la condition sample non prédictible (target non prédictible). Au contraire, dans le cas où les stimuli de la paire target-sample sont différents (donc lorsque des images sample prédictibles et non prédictibles sont comparées à des images target respectivement non prédictibles et prédictibles) les deux images devraient être perçues comme différentes, se traduisant par une différence plus importante entre le niveau de flou attribué aux images sample prédictibles et non prédictibles. Cependant, si la différence entre le niveau de flou attribué aux images sample prédictibles et non prédictibles est la même pour les paires target-sample identiques et différentes, cela irait plutôt dans le sens d'un biais de réponse.

1.1. Expérience 1 – Ajustement du flou lors de la perception subjective d'objets prédictibles ou non prédictibles sur la base de leur contexte

1.1.1. Méthode

1.1.1.1 Participants et Design

Trente-deux étudiants en licence de psychologie de l'Université Grenoble Alpes (28 femmes, $M_{\text{age}} = 22.38$, $SD_{\text{age}} = 4.84$) ayant une vision normale ou corrigée à la normale ont participé à l'expérience. La taille de l'échantillon a été fixée sur la base d'une taille d'effet estimée de $d_z = 0.62$ à partir de l'Expérience 1B de Lupyan (2017) pour obtenir une puissance (i.e., correspondant à $1 - \beta$, β étant le risque d'erreur de type 2) de 0.8 avec un seuil alpha de .05. Pour estimer cette taille d'effet, nous avons calculé le d_z de Cohen de la statistique t pour l'interaction entre la prédictibilité de la target et le type d'essai. Nous avons ensuite calculé l'intervalle de confiance pour le d_z de Cohen et nous avons sélectionné la borne inférieure afin de ne pas surestimer l'effet.

1.1.1.2 Stimuli

Les stimuli ont été construits à partir de 20 photographies en couleur de scènes (10 scènes d'intérieur et 10 scènes d'extérieur) téléchargées sur le site Pixabay (<https://pixabay.com/fr/>), un site de partage de photos sous licence Creative Commons Zero (CC0). Les stimuli ont été créés à l'aide de MATLAB R2019b (MathWorks, Natick, MA, USA). Les images originales ont été recadrées au format carré de 350×350 pixels. Elles ont été converties en images à 256 niveaux de gris en faisant la moyenne des valeurs des trois canaux

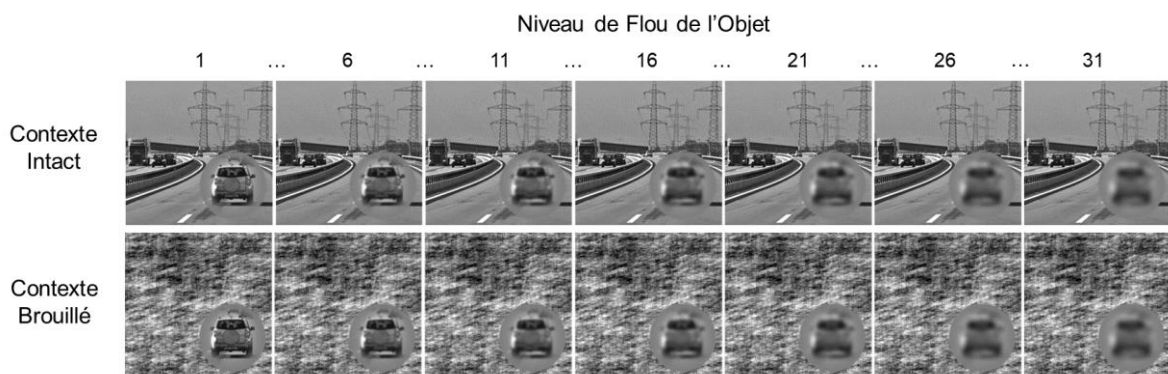
Chapitre 2 – Contribution empirique

de couleur pour chaque pixel. La luminance moyenne de chaque image a ensuite été égalisée à 0.5 (pour des valeurs d'intensité des pixels comprises entre 0 et 1).

Chaque scène contenait un objet sémantiquement lié à la catégorie de la scène (e.g., une voiture pour une scène d'extérieur, une lampe pour une scène d'intérieur) qui était délimité par une ellipse de 176×166 pixels de diamètre vertical et horizontal, respectivement. L'objet pouvait se trouver à différents endroits de l'image selon la scène. Pour chaque scène, nous avons créé deux versions : une dans laquelle le contexte de la scène autour de l'objet était intact (objet prédictible sur la base des informations contextuelles, condition intacte) et une dans laquelle l'information de phase était brouillée et ne contenait donc pas d'informations sémantiques (objet non prédictible d'après les informations contextuelles, condition brouillée). La version de la scène dont l'information de phase était brouillée a été construite en combinant le spectre d'amplitude de la scène avec le spectre de phase d'un bruit blanc dans l'espace de Fourier et l'image résultante a été obtenue par transformation inverse de Fourier.

Figure 2.2

Exemples d'un Objet à Différents Niveaux de Flou.



Note. Exemples d'un objet en conditions intacte et brouillée présenté à différents niveaux de flou entre 1 et 31.

L'ellipse contenant l'objet a ensuite été floutée à l'aide de filtres spatiaux gaussiens bidimensionnels (fonction *imgaussfilt* dans Matlab) de taille linéairement croissante (i.e., avec un écart-type croissant - SD) afin d'obtenir 31 versions floutées de l'objet (Figure 2.2), allant d'un objet net (proche de l'intact ; niveau de flou 1, $SD = 1$) à un objet très flou et difficilement reconnaissable (niveau de flou 31, $SD = 10$). Le floutage des objets n'a eu aucun effet sur la

Chapitre 2 – Contribution empirique

luminance moyenne des stimuli, qui a été fixée à 0.5 (pour des intensités de pixels comprises entre 0 et 1) avant l'application des filtres gaussiens. Il n'a entraîné qu'une légère réduction du contraste Root Mean Square (RMS) des stimuli (de 0.189 au niveau de flou 1 à 0.185 au niveau de flou 31). Pour chaque niveau de flou, l'ellipse contenant l'objet a ensuite été ajoutée sur les contextes de scène intact ou brouillé et dont elle était séparée par un cadre elliptique gris plus grand que celui de l'objet, de luminance moyenne 0.5 (pour les valeurs d'intensité des pixels comprises entre 0 et 1) et de 30 pixels de large. Ce cadre gris nous a permis d'isoler l'objet du contexte et d'éviter les différences de contraste locales entre l'objet et les contextes intacts ou brouillés.

L'intégration de l'objet et du cadre gris sur les contextes intacts et brouillés a été réalisée comme suit. Pour intégrer le cadre elliptique gris au fond de l'image, nous avons d'abord créé pour chaque stimulus un masque binaire du cadre elliptique. Ce masque consistait en une image de 350×350 pixels (même résolution que les scènes) contenant une ellipse de 236×226 pixels de diamètres vertical et horizontal centrée sur l'emplacement de l'objet. Les valeurs des pixels à l'intérieur de l'ellipse ont été fixées à 1 tandis que les valeurs des pixels à l'extérieur de l'ellipse ont été fixées à 0. Ce masque binaire a ensuite été filtré spatialement par une gaussienne avec un écart-type de 2 pour lisser le bord du cadre elliptique. L'image résultante a été multipliée par des valeurs de luminance moyennes de 0.5 pour remplir l'ellipse d'un gris uniforme. Nous avons ensuite calculé l'inverse du masque binaire lissé de l'ellipse pour obtenir une image avec des valeurs de pixel de 0 à l'intérieur de l'ellipse et de 1 à l'extérieur de l'ellipse. Ce masque inverse a été multiplié par la scène intacte ou brouillée, ce qui a permis d'obtenir une image contenant l'arrière-plan de la scène et des valeurs de pixels égales à 0 dans l'ellipse. Enfin, cette image a été ajoutée à l'image contenant le cadre elliptique gris. Une procédure similaire a été utilisée pour fusionner l'ellipse plus petite contenant l'objet avec le fond et le cadre elliptique gris. Nous avons d'abord créé un masque binaire d'une ellipse de 176×166 pixels de diamètres vertical et horizontal centré à l'emplacement de l'objet, qui a ensuite été lissé spatialement par une gaussienne avec un écart-type de 2. L'image résultante a été multipliée avec la scène originale pour obtenir une image contenant uniquement l'objet (les valeurs des pixels à l'extérieur de l'ellipse ont été fixées à 0). Nous avons ensuite créé l'inverse du masque binaire lissé et l'image résultante a été multipliée avec la scène contenant le cadre elliptique gris. L'image résultante a été ajoutée à l'image contenant l'objet elliptique pour obtenir le stimulus final.

1.1.1.3. Procédure

L'expérience a été réalisée dans le contexte des périodes de confinement liées à la pandémie Covid-19 et a donc été menée en ligne. Elle a été construite avec le logiciel E-Prime 3.0 (E-Prime Psychology Software Tools Inc., Pittsburgh, USA) et les stimuli ont été affichés avec le logiciel E-Prime Go 1.0 sur l'ordinateur personnel du participant. Dans le cadre de cette expérience, l'utilisation de l'ordinateur personnel des participants entraînant des angles visuels ou des conditions d'éclairage différents d'un participant à l'autre n'était pas problématique. En effet, l'important était qu'à chaque essai, les deux stimuli en target et en sample soient vus dans les mêmes conditions par chaque participant. Avant de commencer l'expérience, l'expérimentatrice rencontrait chaque participant par vidéoconférence afin d'expliquer la tâche et de s'assurer que les critères informatiques de base étaient remplis pour réaliser l'expérience (e.g., un système d'exploitation supportant E-prime Go, une résolution d'écran de 1920×1080 pixels assurant l'affichage correct des stimuli).

À chaque essai, deux images contenant le même objet étaient affichées à gauche et à droite de l'écran de l'ordinateur du participant sur un fond gris (luminance moyenne de 0.5). L'objet dans l'image de gauche était toujours l'objet target et l'objet dans l'image de droite était toujours l'objet sample. L'objet target était présenté avec un niveau de flou choisi aléatoirement entre les niveaux 6 et 25 et ce niveau de flou restait constant tout au long de l'essai. L'objet sample avait toujours un niveau de flou initial de 16 (niveau de flou médian entre les niveaux 1 et 31) au début de l'essai. Les participants devaient ajuster le niveau de flou de l'objet sample pour qu'il corresponde au niveau de flou de l'objet target (Figure 2.3) en augmentant ou en diminuant le flou de l'objet sample par paliers à l'aide de la touche « P » de leur clavier pour augmenter et de la touche « O » pour diminuer. L'objet sample pouvait être ajusté jusqu'à un flou maximal correspondant au niveau 31 et un flou minimal correspondant au niveau 1. Les participants devaient appuyer sur la touche « Y » pour valider leur réponse une fois qu'ils avaient jugé que les objets sample et target étaient aussi flous l'un que l'autre. Les participants étaient encouragés à se fier à leur jugement subjectif pour faire correspondre le niveau de flou des objets.

Pour chaque essai, le contexte de chaque objet pouvait être soit intact (objet prédictible), soit brouillé (objet non prédictible). De plus, les objets target et sample avaient soit un contexte différent (types de paires target-sample différentes), soit le même contexte (types de paires target-sample identiques). L'expérience comprenait ainsi quatre conditions expérimentales

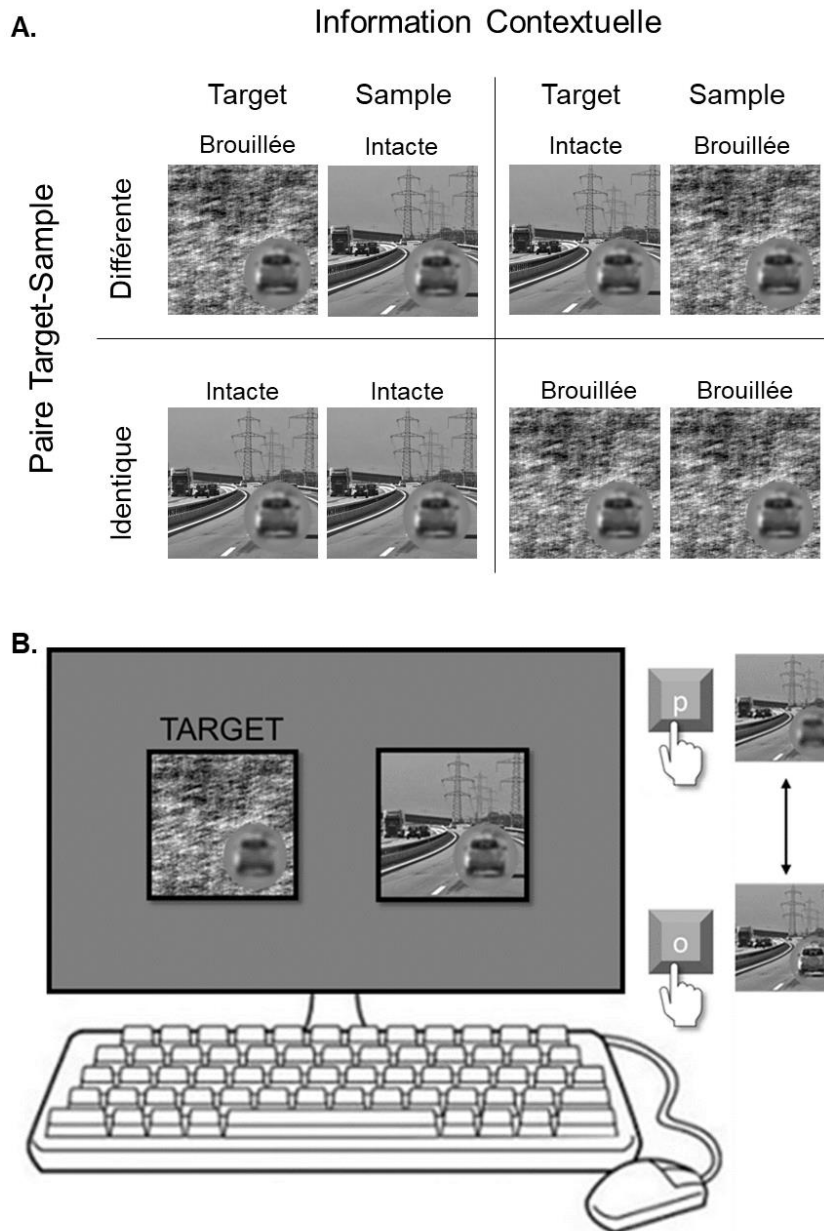
Chapitre 2 – Contribution empirique

(Figure 2.3.A). Les paires target-sample différentes nous ont permis d'évaluer l'effet des informations contextuelles sur la netteté perçue des objets. Les paires target-sample identiques ont été utilisés comme condition contrôle face à la problématique d'El Greco fallacy afin de s'assurer que nos effets étaient bien de nature perceptive et n'étaient pas la conséquence de biais de jugement. Plus spécifiquement, les paires identiques ont permis d'évaluer toute tendance générale à surestimer ou sous-estimer le flou de l'objet sample par rapport à l'objet target, indépendamment de sa prédictibilité, ou pour détecter un biais de réponse plus systématique de sous-estimation du flou des objets intégrés dans des contextes intacts, indépendamment du contexte de l'objet target (i.e., un objet avec un contexte intact ou brouillé). Les conditions prédictible et non prédictible manipulaient la présence et l'absence d'informations sémantiques permettant de générer des prédictions concernant l'objet, mais le contexte (intact ou brouillé) n'était jamais pertinent pour réaliser la tâche qui consistait à ajuster le niveau de flou de l'objet. Cela nous permettait de dissocier les effets des prédictions (i.e., ce qui est susceptible d'être présent dans le signal) des effets de l'attention (i.e., ce qui est pertinent pour la tâche à accomplir) car il a été démontré que l'attention améliore la perception (Rungratsameetaweemana & Serences, 2019; Summerfield & de Lange, 2014).

Une session expérimentale contenait 160 essais apparaissant dans un ordre aléatoire (20 objets présentés deux fois dans chacune des quatre conditions) avec une pause au premier et deuxième tiers de l'expérience. Pour toutes les conditions, chacun des 20 objets target apparaissait avec un niveau de flou différent, variant de 6 à 25 (i.e., un total de 20 niveaux de flou), de sorte que chaque niveau de flou était présenté une fois par condition pour un objet sélectionné au hasard et cet objet était présenté au même niveau de flou dans chacune des quatre conditions expérimentales. L'expérience durait environ 30-40 minutes, en fonction de la vitesse à laquelle chaque participant effectuait son ajustement (durée moyenne \pm SD d'un essai : 1858 \pm 2133 ms). Pour chaque essai, nous avons enregistré le niveau de flou attribué à l'objet sample pour correspondre à celui de l'objet target (*Matched Blur Level*, MBL). Avant la session expérimentale, les participants effectuaient une session d'entraînement sur quatre essais en utilisant des stimuli qui n'ont pas été inclus par la suite dans l'expérience principale.

Figure 2.3

Illustration des Conditions Expérimentales et du Déroulement de l'Expérience 1.



Note. A. Exemple de stimuli présentés dans chaque condition expérimentale. Pour chaque essai, les objets target et sample étaient identiques. En revanche, leur contexte pouvait être soit intact (informations contextuelles prédictibles), soit avec une information de phase brouillée (informations contextuelles non prédictibles). Ils avaient soit un contexte différent (paires target-sample différentes), soit le même contexte (paires target-sample identiques), ce qui donne quatre conditions expérimentales : (1) target et sample étaient intacts (T-intact/S-intact), (2) target et sample étaient brouillés (T-brouillé/S-brouillé), (3) target intact et sample brouillé (T-intact/S-brouillé), et (4) target brouillé et

Chapitre 2 – Contribution empirique

sample intact (T-brouillé/S-intact). B. Illustration d'un essai avec un objet target brouillé et un objet sample intact (paires target-sample différentes). Deux scènes contenant un objet flou apparaissent simultanément, l'image target à gauche et l'image sample à droite. Les participants devaient ajuster le niveau de flou de l'objet sample en appuyant sur les touches « O » et « P » du clavier, jusqu'à ce qu'ils jugent les deux objets comme aussi flous l'un que l'autre.

1.1.2. Analyse des données

Les données ont été analysées à l'aide de R (R Core Team, 2019) et du package *lme4* (Bates, Mächler, et al., 2015) pour effectuer une analyse linéaire à effets mixtes du type de paire target-sample (identique vs. différente) et des informations contextuelles de l'image sample (intactes vs. brouillées) dans un design intra-sujet 2×2 . Contrairement aux analyses de variance classique (ANOVA ; Judd et al., 2012), les analyses à effets mixtes permettent de maximiser la généralisation de nos résultats à d'autres participants, mais également à d'autres stimuli. Nous avons utilisé le Matched Blur Level (MBL) comme variable dépendante. Le type de paire target-sample, les informations contextuelles de l'image sample et leur terme d'interaction ont été introduits dans le modèle comme effets fixes (i.e., comme effets des variables d'intérêt). Les intercepts pour les sujets et les stimuli, ainsi que les pentes aléatoires par sujet et par stimulus pour l'effet du type de paire target-sample, des informations contextuelles de l'image sample et de leur interaction ont été spécifiés comme effets aléatoires (i.e., comme les effets des variables auxquelles nous voulons généraliser nos résultats).

Nous avons effectué le nombre minimum d'analyses requises pour tester nos hypothèses afin de ne pas augmenter le risque d'erreur de type 1. Sur la base des résultats de Lupyan (2017), nous nous attendions à ce qu'à un niveau de flou objectivement égal, les objets intégrés dans des contextes intacts soient subjectivement perçus comme plus nets que les objets intégrés dans des contextes brouillés. Ainsi, pour les paires target-sample différentes, les objets dans les contextes intacts devraient être jugés comme plus flous que les objets dans les contextes brouillés pour être en fait subjectivement perçus comme aussi flous l'un que l'autre. Par conséquent, pour les paires target-sample différentes, le MBL des objets en sample dans la condition T-brouillé/S-intact devrait être plus élevé que le MBL des objets en sample dans la condition T-intact/S-brouillé. Cet effet pourrait également s'expliquer par un biais systématique de sous-estimation du flou des objets intégrés dans des contextes intacts, indépendamment du contexte de l'objet en target auquel il est apparié. Si la différence entre le MBL des objets en

Chapitre 2 – Contribution empirique

sample dans des contextes intacts et brouillés était due à un tel biais systématique, elle devrait être aussi importante quel que soit le type de paire target-sample (différente ou identique). Au contraire, une différence de MBL plus importante pour les types de paires différentes que pour les types de paires identiques conduirait à la conclusion que la différence de MBL pour les types de paires target-sample différentes n'est pas susceptible d'être expliquée par un biais basé sur les informations contextuelles de l'image sample. Afin de vérifier ces hypothèses, nous avons donc testé l'interaction entre l'information contextuelle de l'image sample et le type de paire target-sample, puis l'effet simple de l'information contextuelle de l'objet en sample pour les paires target-sample différentes.

Nous avons utilisé la méthode proposée par Bates, Kliegl, et al. (2015) pour construire des modèles mixtes parcimonieux limitant les problèmes de convergence. L'inspection visuelle des graphiques de résidus n'a pas révélé de déviations évidentes de l'homoscédasticité ou de la normalité. Les p-valeurs ont été obtenues par approximation de Satterthwaite avec le package *lmerTest* (Kuznetsova et al., 2017). Le seuil de significativité a été fixé à .05. Les tailles d'effet ont été estimées à l'aide du d_z de Cohen (Lakens, 2013).

1.1.3. Résultats

Les MBL moyens pour chaque condition expérimentale sont présentés dans la Figure 2.4. Étant donné que l'objet en target pouvait apparaître à 20 niveaux de flou différents allant de 6 à 25, une correspondance parfaite entre les objets en target et en sample se traduirait par un MBL moyen de $(6+25)/2 = 15.5$. Un MBL moyen dont la valeur est supérieure à 15.5 indique donc une surestimation du flou de l'objet sample par rapport à celui de l'objet target, tandis qu'un MBL moyen inférieur à 15.5 indique une sous-estimation du flou de l'objet sample par rapport à celui de l'objet target.

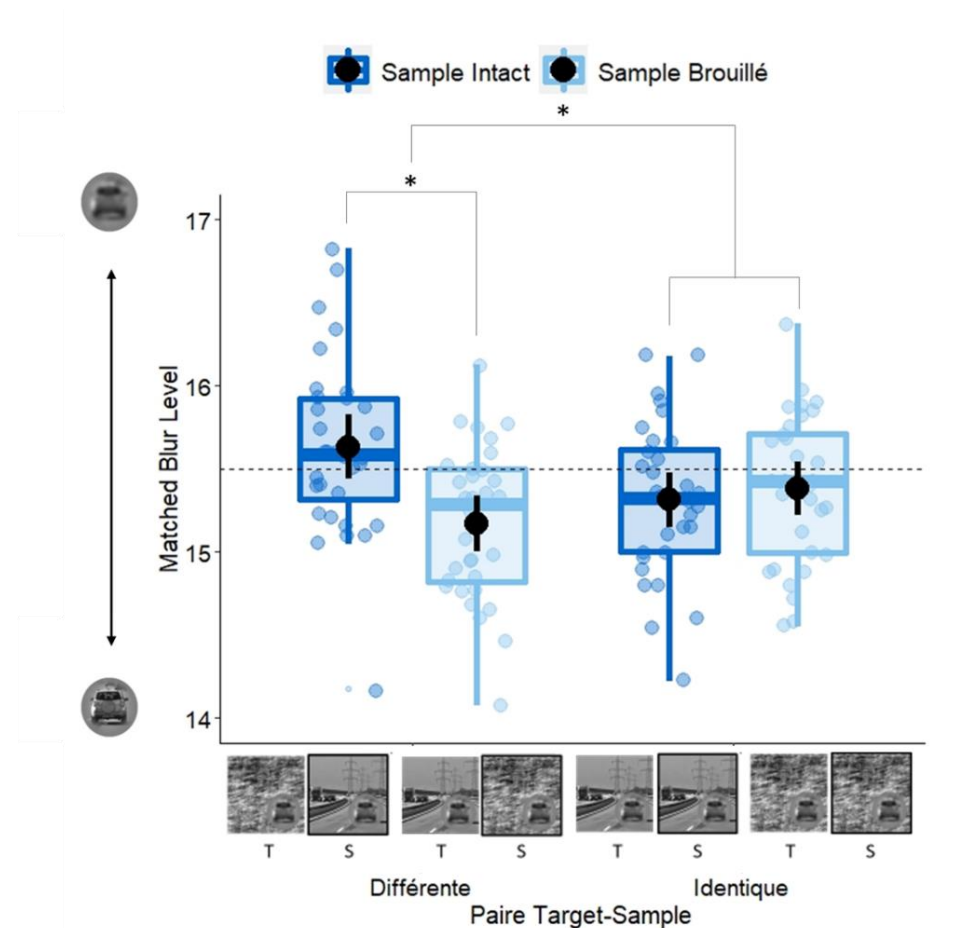
Les résultats n'ont révélé ni effet principal du type de paire target-sample, $t(18.99) = 0.21$, $p = .838$, $d_z = 0.037$ (type de paire différente : $M = 15.40$, $SE = 1.02$; type de paire identique : $M = 15.35$, $SE = 1$), ni effet principal de l'information contextuelle de l'image sample, $t(5078) = 1.53$, $p = .127$, $d_z = 0.270$ (condition intacte : $M = 15.47$, $SE = 1.03$; condition brouillée : $M = 15.28$, $SE = 0.99$). Cependant, l'interaction entre le type de paire target-sample et l'information contextuelle de l'image sample était significative, $t(5078) = 2.06$, $p = .039$, $d_z = 0.364$. Conformément à nos hypothèses, la différence de MBL entre les conditions intacte et brouillée était plus importante lorsque les contextes en target et en sample étaient différents (T-

Chapitre 2 – Contribution empirique

brouillé/S-intact : $M = 15.63$, $SE = 1.05$; T-intact/S-brouillé : $M = 15.17$, $SE = 0.98$) que lorsqu'ils étaient identiques (T-intact/S-intact : $M = 15.31$, $SE = 1$; T-brouillé/S-brouillé : $M = 15.38$, $SE = 1$). Par ailleurs, lorsque les contextes en target et en sample étaient différents, le MBL était significativement plus élevé pour un objet en sample dans la condition intacte que dans la condition brouillée, $t(5079.36) = 2.53$, $p = .011$, $d_z = 0.448$. Dans l'ensemble, ces résultats soutiennent nos hypothèses. Les participants ont ajouté plus de flou à un objet en sample dans un contexte intact apparié à un objet en target dans un contexte brouillé qu'à un objet en sample dans un contexte brouillé apparié à un objet en target dans un contexte intact. Cela suggère qu'à un niveau de flou objectivement égal, les objets dans des contextes intacts étaient subjectivement perçus comme plus nets que les objets dans des contextes brouillés.

Figure 2.4

Matched Blur Level Moyen pour Chaque Condition de l'Expérience 1.



Note. Niveau de flou moyen attribué à l'objet en sample pour correspondre au niveau de flou de l'objet en target (Matched Blur Level) pour les sample en condition intacte et brouillée en fonction du type de paire target-sample. Un niveau de flou élevé correspond à un objet très flou tandis qu'un niveau de flou

Chapitre 2 – Contribution empirique

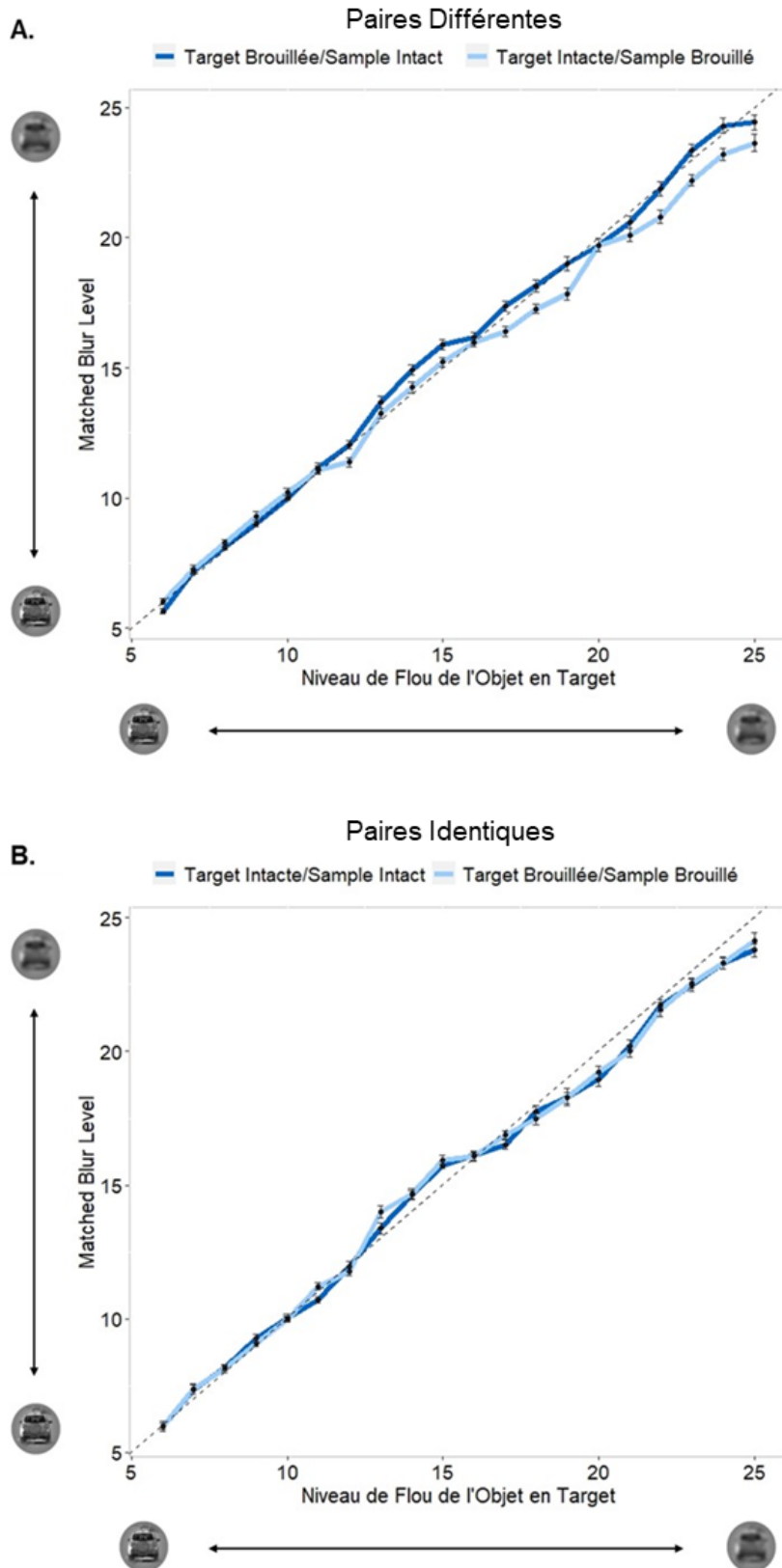
faible correspond à un objet relativement net. La ligne en pointillés représente la correspondance parfaite entre le flou des objets en target (T) et en sample (S). Un MBL moyen dont la valeur est supérieure ou inférieure à 15.5 indique une surestimation ou une sous-estimation du flou de l'objet sample par rapport à l'objet target, respectivement. Les points noirs et les barres d'erreur indiquent la moyenne et l'erreur standard sur tous les individus, respectivement. Les points en couleur correspondent aux observations individuelles.

* $p < .05$.

La Figure 2.5 fournit une autre représentation de nos résultats dans laquelle le MBL de l'objet en sample est représenté en fonction du niveau de flou réel de l'objet en target pour les paires target-sample différentes uniquement. La ligne en pointillés correspond à une correspondance parfaite entre le flou de l'objet en sample et celui de l'objet en target. Les points au-dessus ou en-dessous de cette ligne indiquent une surestimation ou une sous-estimation du flou du sample, respectivement. Il est intéressant de noter que l'inspection visuelle de ce graphique suggère que l'effet des informations contextuelles sur le MBL varie en fonction du niveau de flou de l'objet en target. Cet effet semble être négligeable pour les objets en target dont le niveau de flou est inférieur à 11 et être le plus fort pour les objets en target dont le niveau de flou est supérieur à 16. Comme il n'y avait qu'un nombre limité d'essais par niveau de flou de l'objet en target (i.e., chaque objet en target n'était pas affiché à tous les niveaux de flou pendant toute l'expérience, voir la Méthode), l'effet du niveau de flou de l'objet en target n'a pas été inclus comme effet d'intérêt dans notre modèle. Pour avoir un aperçu de l'importance de cet effet, nous avons regroupé les essais dans lesquels le niveau de flou de l'objet en target était faible (inférieur à 16) et les essais dans lesquels le niveau de flou de l'objet en target était élevé (supérieur à 16) et nous avons effectué une analyse en modèles mixtes post-hoc afin d'examiner si l'effet des informations contextuelles pouvait varier en fonction du niveau de flou faible ou élevé de l'objet en target. Nous avons calculé comme effet fixe l'interaction entre les informations contextuelles du sample (intactes ou brouillées) pour les paires target-sample différentes uniquement et le niveau de flou de l'objet en target (faible ou élevé). L'interaction était significative, $t(2378.01) = 3.15$, $p = .002$, $d_z = 0.557$. Cette interaction suggère que l'effet de l'information contextuelle du sample était plus important pour un niveau de flou élevé de l'objet en target (supérieur à 16 ; T-brouillé/S-intact : $M = 20.98$, $SE = 0.57$; T-intact/S-brouillé : $M = 20.13$, $SE = 0.55$) qu'un niveau de flou faible (inférieur à 16 ; T-brouillé/S-intact : $M = 10.77$, $SE = 0.61$; T-intact/S-brouillé : $M = 10.63$, $SE = 0.56$).

Figure 2.5

Matched Blur Level Moyen pour les Paires Target-Sample Différentes et Identiques en Fonction du Niveau de Flou de l'Objet en Target.



Chapitre 2 – Contribution empirique

Note. A. Niveau de flou moyen attribué à l'objet en sample pour correspondre au niveau de flou de l'objet en target (MBL) pour les paires target-sample différentes (conditions T-brouillé/S-intact et T-intact/S-brouillé) en fonction du niveau de flou de l'objet en target. La ligne en pointillés représente le niveau de flou correct de l'objet en sample pour chaque niveau de flou de l'objet en target. Les points noirs et les barres d'erreur indiquent respectivement la moyenne et l'erreur standard de tous les participants. B. Niveau de flou moyen attribué à l'objet en sample pour correspondre au niveau de flou de l'objet en target (MBL) pour les paires target-sample identiques (conditions T-intact/S-intact et T-brouillé/S-brouillé) en fonction du niveau de flou de l'objet en target. La ligne en pointillés représente le niveau de flou correct de l'objet en sample pour chaque niveau de flou de l'objet en target. Les points noirs et les barres d'erreur indiquent respectivement la moyenne et l'erreur standard de tous les participants.

1.1.4. Discussion de l'Expérience 1

Le but de cette première expérience était de déterminer si les prédictions basées sur les informations contextuelles de la scène améliorent la netteté perçue des objets. Pour ce faire, les participants ont effectué une tâche d'ajustement perceptif dans laquelle ils devaient ajuster le niveau de flou d'un objet (sample) intégré dans un contexte intact prédictible ou dans un contexte brouillé non prédictible afin de correspondre au niveau de flou du même objet (target) intégré dans un contexte intact (objet prédictible) ou dans un contexte brouillé (objet non prédictible).

Conformément à nos hypothèses, les résultats ont montré que lorsque les paires d'objets avaient un contexte différent (types de paires target-sample différentes), les participants faisaient correspondre le niveau de flou des objets en target et en sample en attribuant un niveau de flou plus élevé aux objets en sample ayant un contexte intact (appariés aux objets en target ayant un contexte brouillé) qu'aux objets en sample ayant un contexte brouillé (appariés aux objets en target ayant un contexte intact). Cela suggère qu'à un niveau de flou objectivement égal, les objets dans un contexte intact étaient subjectivement perçus comme plus nets que les objets dans un contexte brouillé. Pour vérifier que cet effet des informations contextuelles dans les paires target-sample avec des contextes différents n'était pas dû à (1) un biais systématique consistant à systématiquement attribuer plus de flou que nécessaire aux objets en sample intégrés dans un contexte intact indépendamment du contexte de l'objet en target ou (2) un biais systématique consistant à attribuer plus de flou à l'objet en sample par rapport à l'objet en target,

Chapitre 2 – Contribution empirique

nous avons également inclus dans notre expérience des types de paires target-sample ayant un contexte identique. Nous avons observé que la différence entre le MBL pour un objet en sample intégré dans un contexte intact et le MBL pour le même objet en sample intégré dans un contexte brouillé était plus grande lorsque l'objet en sample et en target avaient un contexte différent plutôt qu'identique. Cela indique que la tendance à attribuer plus de flou aux objets du sample dans un contexte intact était plus importante lorsqu'ils étaient comparés à des objets target dans un contexte brouillé. Cela plaide donc en défaveur d'une explication de nos résultats par un biais systématique et soutient plutôt l'idée que les prédictions basées sur des informations contextuelles ont pour effet d'augmenter la netteté perçue des objets. On peut également noter que le MBL moyen pour les paires target-sample identiques était inférieur à 15.5, ce qui suggère une tendance générale à attribuer moins de flou à l'objet sample que le niveau de flou requis pour le faire correspondre à l'objet en target lorsqu'ils partagent un contexte similaire, en particulier pour les objets très flous avec un niveau de flou supérieur à 16 (voir la Figure 2.5 pour les MBL en fonction du niveau de flou de l'objet en target dans les types de paires identiques). Cette tendance à ne pas suffisamment augmenter le flou de l'objet en sample dans les paires identiques pourrait s'expliquer par un biais perceptif consistant à toujours percevoir l'objet sample comme moins net que l'objet target ou bien un biais de réponse pendant la tâche, les participants adaptant le niveau de flou de l'objet en sample pour que cet objet soit plus net que nécessaire. La présence d'un tel biais renforce nos conclusions dans les conditions où les paires target-sample étaient différentes. Les participants ont ajouté plus de flou que nécessaire aux objets du sample dans les contextes intacts malgré une tendance générale à rendre l'objet en sample plus net que l'objet en target.

Il est intéressant de noter que les analyses post-hoc suggèrent que l'effet des prédictions basées sur les informations contextuelles (i.e., l'effet des informations contextuelles dans les paires target-sample différentes) était significativement plus important lorsque l'objet en target était très flou (i.e., un niveau de flou supérieur à 16) que lorsqu'il était relativement net (i.e., un niveau de flou inférieur à 16). Ce résultat est également conforme aux théories du codage prédictif qui supposent que l'influence des prédictions sur la perception augmente lorsque la fiabilité du signal diminue (Feldman & Friston, 2010; Friston, 2005).

Dans l'ensemble, les résultats de cette expérience sont cohérents avec les résultats précédents de Lupyan (2017) selon lesquels les prédictions basées sur des connaissances préalables augmentent la netteté perçue du signal visuel. Nos résultats suggèrent même que ce mécanisme ne s'applique pas seulement à la lecture de mots, qui nécessite un traitement des

Chapitre 2 – Contribution empirique

détails, mais aussi à la perception des objets. Par conséquent, nous nous sommes demandées si ces résultats pouvaient être étendus plus largement à la perception de scènes entières qui nécessitent moins de détails pour être identifiées. Cela confirmerait que le rehaussement perceptif des stimuli prédictibles est un principe plus général de la vision. De plus, bien que l'objet soit exactement le même en target et en sample, nous ne pouvons exclure que la perception d'une plus grande netteté pour les objets intégrés dans des contextes intacts plutôt que brouillés résulte d'une perturbation du traitement de l'objet dans un contexte brouillé qu'intact, plutôt que de prédictions basées sur des informations contextuelles. L'Expérience 2 visait donc à confirmer les résultats de l'Expérience 1 et à les étendre à la perception de scènes entières. Nous avons utilisé la même tâche d'ajustement perceptif, mais cette fois-ci, les stimuli target et sample avaient exactement les mêmes propriétés de signal et ne différaient qu'en termes d'orientation : une scène entière prédictible à l'endroit contre une scène non prédictible à l'envers.

1.2. Expérience 2 – Ajustement du flou lors de la perception subjective de scènes prédictibles et non prédictibles

1.2.1. Méthode

1.2.1.1. Participants et Design

Trente étudiants en licence de psychologie de l'Université Grenoble Alpes (25 femmes, $M_{\text{age}} = 21.43$, $SD_{\text{age}} = 4.07$) ayant une vision normale ou corrigée à la normale ont participé à l'expérience. Cette expérience a été menée en parallèle de l'Expérience 1 et la taille d'échantillon a été estimée de la même manière que pour l'Expérience 1.

1.2.1.2. Stimuli

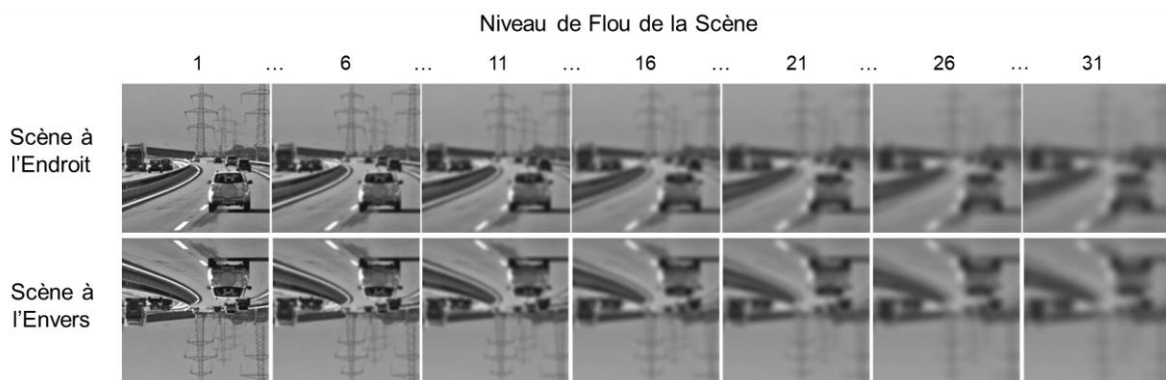
Les stimuli ont été construits à partir des mêmes 20 photographies de scènes d'intérieur et d'extérieur que l'Expérience 1 (350 × 350 pixels, images en niveaux de gris avec une luminance moyenne égalisée à 0.5 en utilisant MATLAB R2019b). Pour chaque scène, nous avons créé deux versions dans lesquelles les scènes étaient exactement les mêmes mais avec une orientation différente : l'une dans laquelle la scène était laissée à l'endroit (disposition prédictible de la scène - condition à l'endroit) et l'autre dans laquelle la scène était inversée

Chapitre 2 – Contribution empirique

autour de l'axe horizontal (disposition non prédictible de la scène - condition à l'envers). Les versions à l'endroit et à l'envers de chaque scène ont ensuite été floutées en utilisant la même procédure que dans l'Expérience 1 pour les objets afin d'obtenir 31 versions de l'image (Figure 2.6) allant d'une image nette proche de l'état intact (niveau de flou 1, $SD = 1$) à une image très floue et difficilement reconnaissable (niveau de flou 31, $SD = 10$).

Figure 2.6

Exemples d'une Scène à Différents Niveaux de Flou.



Note. Scène en conditions d'orientation à l'endroit et à l'envers présentée à différents niveaux de flou de 1 à 31.

1.2.1.3. Procédure

Cette expérience a également été réalisée dans le contexte des périodes de confinement liées à la pandémie Covid-19 et a donc été menée en ligne. Les paramètres expérimentaux étaient similaires à ceux de l'Expérience 1 (i.e., les stimuli étaient affichés à l'aide du logiciel E-Prime Go 1.0 sur l'ordinateur personnel du participant). La procédure était également similaire à celle de l'Expérience 1, à la différence que la tâche consistait à faire correspondre le niveau de flou de scènes entières au lieu d'objets.

Pour chaque essai, la scène en target et la scène en sample pouvait être soit à l'endroit (orientation prédictible, conformément à l'expérience antérieure), soit à l'envers (orientation non prédictible car sans expérience antérieure). En outre, elles avaient soit la même orientation (types de paires target-sample identiques), soit une orientation différente (types de paires target-sample différentes), ce qui donne quatre conditions (voir Figure 2.7). Comme dans l'Expérience 1, l'orientation de la scène n'était pas pertinente pour réaliser la tâche d'ajustement de niveau

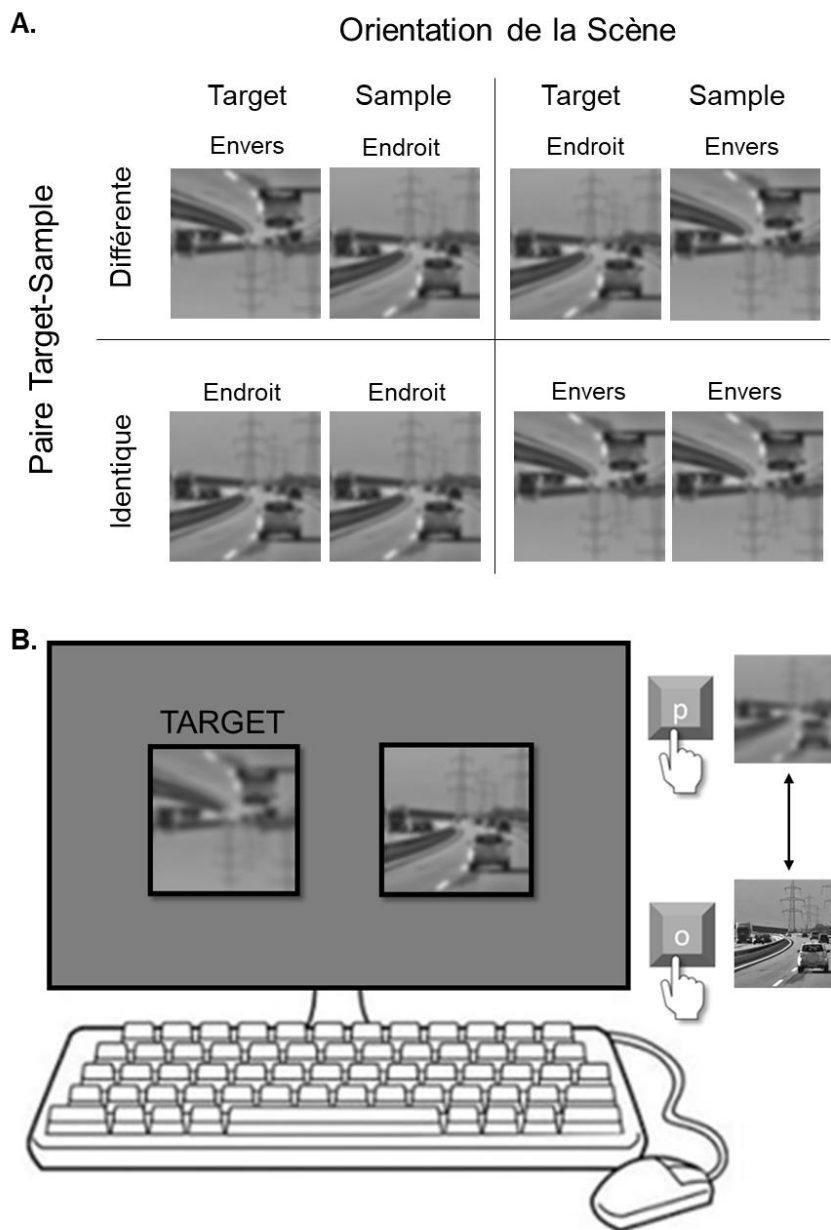
Chapitre 2 – Contribution empirique

de flou. Une fois encore, la manipulation du niveau de flou des stimuli n'a eu aucun effet sur la luminance moyenne des stimuli, qui a été fixée à 0.5 (pour des intensités de pixels comprises entre 0 et 1) avant l'application des filtres gaussiens. Le floutage des stimuli n'a entraîné qu'une légère réduction du contraste RMS des stimuli (de 0.189 au niveau de flou 1 à 0.155 au niveau de flou 31).

Une session expérimentale contenait 160 essais apparaissant dans un ordre aléatoire (20 scènes présentées deux fois dans chacune des quatre conditions) avec une pause aux premier et deuxième tiers de l'expérience. Pour toutes les conditions, chacune des 20 scènes en target apparaissait avec un niveau de flou variant de 6 à 25 (soit un total de 20 niveaux de flou), de sorte que chaque niveau de flou était présenté une fois par condition pour une scène sélectionnée au hasard et chaque scène était présentée au même niveau de flou dans chacune des quatre conditions expérimentales. L'expérience durait environ 30-40 minutes, en fonction de la vitesse à laquelle chaque participant effectuait son ajustement (durée moyenne \pm *SD* d'un essai : 2134 \pm 2192 ms). Pour chaque essai, nous avons enregistré le niveau de flou attribué à la scène en sample pour qu'il corresponde à celui de la scène en target (Matched Blur Level, MBL). Avant la session expérimentale, les participants ont effectué une session d'entraînement sur quatre essais en utilisant des stimuli qui n'ont pas été inclus par la suite dans l'expérience principale.

Figure 2.7

Illustration des Conditions Expérimentales et du Déroulement de l'Expérience 2.



Note. A. Exemple de stimuli présentés dans chaque condition expérimentale. Pour chaque essai, les scènes en target et en sample pouvaient être soit à l'endroit (orientation prédictible), soit à l'envers (orientation non prédictible). Elles avaient soit une orientation différente (paires target-sample différentes), soit la même orientation (paires target-sample identiques), ce qui donne quatre conditions : (1) target et sample à l'endroit (T-endroit/S-endroit), (2) target et sample à l'envers (T-envers/S-envers), (3) target à l'endroit et sample à l'envers (T-endroit/S-envers), et (4) target à l'envers et sample à l'endroit (T-envers/S-endroit). B. Illustration d'un essai avec une scène target à l'envers et une scène sample à l'endroit (paires target-sample différentes). Deux scènes floues apparaissent simultanément, la

Chapitre 2 – Contribution empirique

scène target à gauche et la scène sample à droite. Les participants devaient ajuster le niveau de flou de la scène en sample en appuyant sur les touches « O » et « P » du clavier, jusqu'à ce qu'ils jugent les deux scènes comme aussi floues l'une que l'autre.

1.2.2. Analyse des données

L'analyse des données et le logiciel utilisé étaient similaires à ceux de l'Expérience 1. Nous avons effectué une analyse linéaire à effets mixtes du type de paire target-sample (identique ou différente) et de l'orientation de la scène en sample (à l'endroit ou à l'envers) dans un modèle intra-sujet 2×2 , avec le MBL comme variable dépendante. Comme dans l'Expérience 1, le type de paire target-sample, l'orientation de la scène en sample et leur terme d'interaction ont été introduits dans le modèle comme effets fixes. Les intercepts pour les sujets et les stimuli, ainsi que les pentes aléatoires par sujet et par stimulus pour l'effet du type de paire target-sample, de l'orientation de la scène en sample et de leur interaction ont été spécifiés comme effets aléatoires.

Similairement à l'Expérience 1, nous avons testé l'interaction entre l'orientation de la scène en sample et le type de paire target-sample, ainsi que l'effet simple de l'orientation de la scène en sample pour les paires target-sample différentes. Nous avons utilisé la même méthode que dans l'Expérience 1 pour construire des modèles mixtes parcimonieux limitant les problèmes de convergence et vérifier les déviations d'homoscédasticité et de normalité. Les p-values ont été obtenues par approximation de Satterthwaite avec le package *lmerTest* (Kuznetsova et al., 2017). Les données d'un participant ont été écartées de l'analyse car celui-ci n'a pas réalisé la tâche et se contentait de valider chaque essai sans modifier le niveau de flou de l'image sample.

1.2.3. Résultats

Les MBL moyens pour chaque condition expérimentale sont présentés dans la Figure 2.8. Comme dans l'Expérience 1, un MBL moyen de 15.5 traduit une correspondance parfaite entre les scènes en target et en sample, tandis qu'un MBL moyen avec une valeur supérieure/inférieure à 15.5 indique une sur/sous-estimation du flou de la scène en sample par rapport à celui de la scène en target.

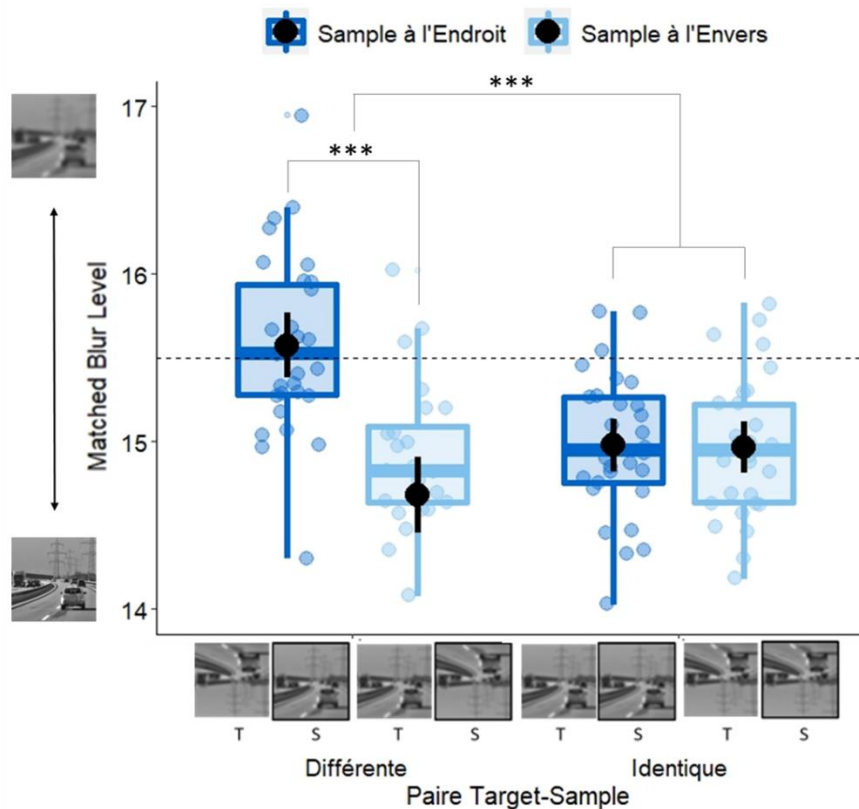
Les résultats n'ont révélé aucun effet principal du type de paire target-sample, $t(18.99) = 0.14$, $p = .892$, $d_z = 0.025$ (type de paire target-sample différente : $M = 15.13$, $SE = 1.05$; type

Chapitre 2 – Contribution empirique

de paire target-sample identique : $M = 14.97$, $SE = 1.01$), mais nous avons observé un effet principal de l'orientation de la scène en sample, $t(4758) = 3.63$, $p < .001$, $d_z = 0.662$, les participants attribuant un MBL plus élevé à la scène en sample lorsqu'elle était à l'endroit ($M = 15.28$, $SE = 1.05$) que lorsqu'elle était à l'envers ($M = 14.83$, $SE = 1$). De plus, l'interaction entre le type de paire target-sample et l'orientation de la scène en sample était significative, $t(4758) = 3.55$, $p < .001$, $d_z = 0.649$. Conformément à nos hypothèses, la différence de MBL entre les scènes en sample à l'endroit et à l'envers était plus importante lorsque les scènes en sample et en target avaient une orientation différente (T-envers/S-endroit : $M = 15.58$, $SE = 1.07$; T-endroit/envers : $M = 14.68$, $SE = 1.01$) qu'une orientation identique (T-endroit/S-endroit : $M = 14.98$, $SE = 1.02$; T-envers/S-envers : $M = 14.97$, $SE = 0.99$). De façon critique, lorsque la target et le sample avaient une orientation différente, le MBL était significativement plus élevé pour un sample à l'endroit qu'à l'envers, $t(4758) = 5.08$, $p < .001$, $d_z = 0.927$. Dans l'ensemble, ces résultats confirment nos hypothèses. Les participants ont ajouté plus de flou à une scène en sample à l'endroit appariée à une scène en target à l'envers qu'à une scène en sample à l'envers appariée à une scène en target à l'endroit.

Figure 2.8

Matched Blur Level Moyen pour Chaque Condition de l'Expérience 2.



Chapitre 2 – Contribution empirique

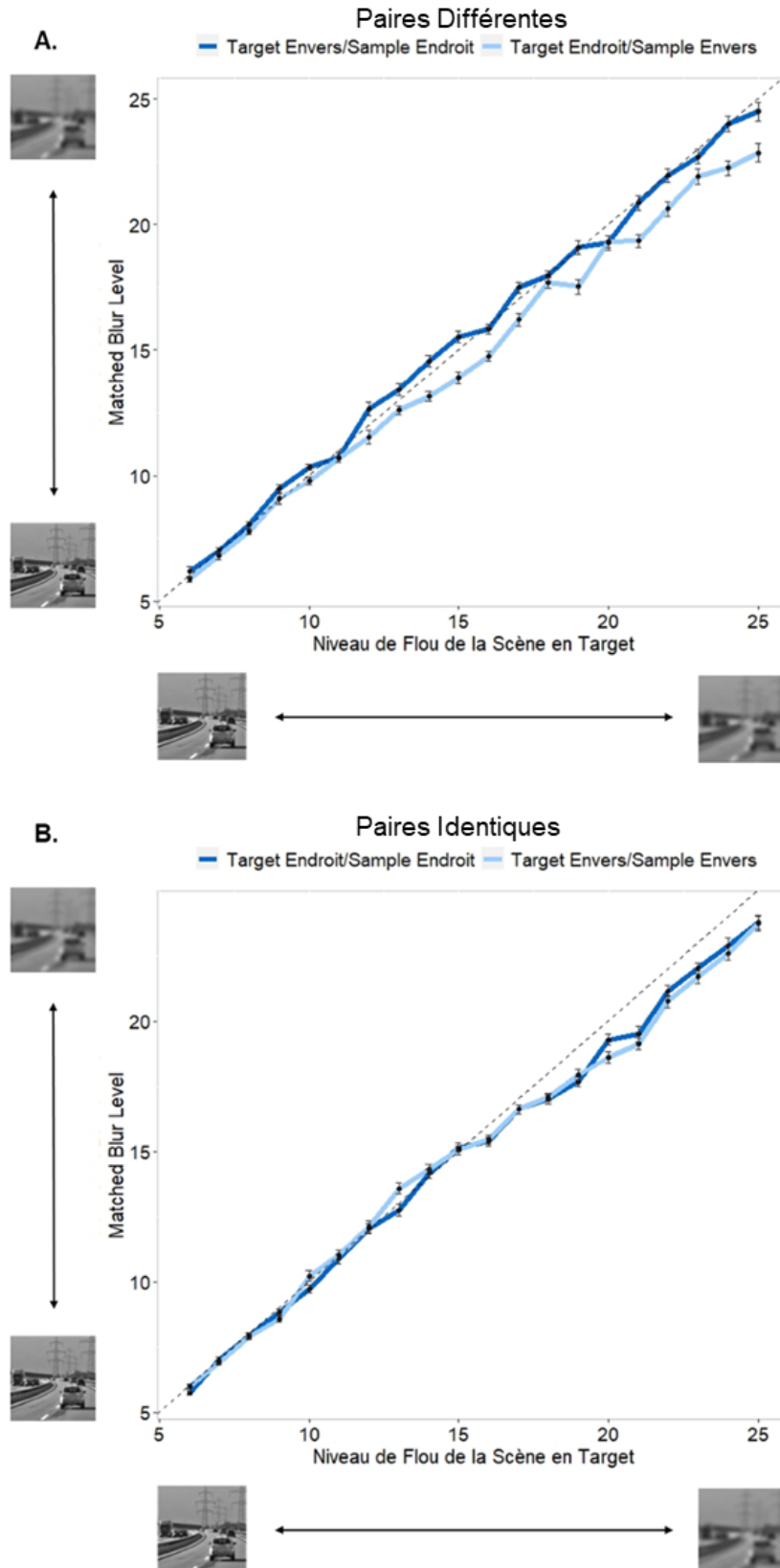
Note. Niveau de flou moyen attribué à la scène en sample pour correspondre au niveau de flou de la scène en target (Matched Blur Level) pour les scènes sample à l'envers et à l'endroit en fonction du type de paire target-sample. Un niveau de flou élevé correspond à une scène très floue tandis qu'un niveau de flou faible correspond à une scène relativement nette. La ligne en pointillés représente la correspondance parfaite entre le flou des scènes en target (T) et en sample (S). Un MBL moyen dont la valeur est supérieure ou inférieure à 15.5 indique une surestimation ou une sous-estimation du flou de la scène en sample par rapport à la scène en target, respectivement. Les points noirs et les barres d'erreur indiquent respectivement la moyenne et l'erreur standard de tous les participants. Les points en couleur correspondent aux observations individuelles.

* $p < .05$.

Comme dans l'Expérience 1, la Figure 2.9 fournit une autre représentation de nos résultats dans laquelle le MBL de la scène en sample est représenté en fonction du niveau de flou de la scène en target pour des paires target-sample différentes uniquement. La ligne en pointillés représente une correspondance parfaite entre le flou de la scène en sample et celui de la scène en target, et les points situés au-dessus ou en-dessous de cette ligne indiquent une surestimation ou une sous-estimation du flou de la scène en sample, respectivement. Une fois de plus, l'inspection visuelle de ce graphique suggère que l'effet de l'orientation de la scène en sample sur le MBL varie en fonction du niveau de flou de la scène en target. Comme dans l'Expérience 1, nous avons calculé dans une analyse post-hoc l'interaction entre l'orientation de la scène en sample pour les paires target-sample différentes uniquement et le niveau de flou de la scène en target (faible vs. élevé). Cependant, l'interaction n'était pas significative, $t(2227.56) = 1.86$, $p = .063$, $d_z = 0.339$, suggérant que l'effet de l'orientation de la scène en sample n'était pas significativement plus important pour un niveau de flou élevé de la scène en target (supérieur à 16 ; T-envers/S-endroit : $M = 20.86$, $SE = 0.59$; T-endroit/S-envers : $M = 19.74$, $SE = 0.56$) que pour un niveau de flou faible (inférieur à 16 ; T-envers/S-endroit : $M = 10.80$, $SE = 0.62$; T-endroit/S-envers : $M = 10.12$, $SE = 0.54$).

Figure 2.9

Matched Blur Level Moyen pour les Paires Target-Sample Différentes et Identiques en Fonction du Niveau de Flou de la Scène en Target.



Chapitre 2 – Contribution empirique

Note. A. Niveau de flou moyen attribué à la scène en sample pour correspondre au niveau de flou de la scène en target (Matched Blur Level) pour les conditions T-envers/S-endroit et T-endroit/S-envers en fonction du niveau de flou de la scène en target. La ligne en pointillés représente le niveau de flou correct de la scène en sample pour chaque niveau de flou de la scène en target. Les points noirs et les barres d'erreur indiquent respectivement la moyenne et l'erreur standard de tous les participants. B. Niveau de flou moyen attribué à la scène en sample pour correspondre au niveau de flou de la scène en target (MBL) pour les conditions T-endroit/S-endroit et T-envers/S-envers en fonction du niveau de flou de la scène en target. La ligne en pointillés représente le niveau de flou correct de la scène en sample pour chaque niveau de flou de la scène en target. Les points noirs et les barres d'erreur indiquent respectivement la moyenne et l'erreur standard de tous les participants.

1.2.4. Discussion de l'Expérience 2

Le but de cette deuxième expérience était de déterminer si les prédictions fondées sur l'expérience antérieure augmentent la netteté perçue de scènes entières. Elle nous a également permis de nous assurer que les résultats de l'Expérience 1 ne pouvaient pas être expliqués par des différences de signaux visuels entre les images en target et en sample. En effet, dans l'Expérience 1, nous avons observé une perception plus nette des objets en contexte intact que brouillé (scramble de phase), mais du fait de la manipulation du contexte des objets, les deux stimuli en target et en sample dans nos essais d'intérêt comportaient des différences dans le signal visuel qui auraient pu en partie expliquer nos résultats. Nous avons utilisé la même tâche d'ajustement perceptif que dans l'Expérience 1, mais ici, les participants devaient ajuster le niveau de flou d'une scène sample à l'endroit prédictible ou à l'envers non prédictible pour correspondre au niveau de flou de la même scène target à l'endroit prédictible ou à l'envers non prédictible. Les images sample et target étaient donc exactement les mêmes en termes de signal visuel.

Nous avons répliqué les résultats de l'Expérience 1 en montrant que lorsque les paires target-sample avaient une orientation différente (types de paires target-sample différentes), les participants faisaient correspondre le niveau de flou des scènes en target et en sample en attribuant à la scène en sample à l'endroit un niveau de flou plus élevé qu'à la scène en sample à l'envers. Cela signifie qu'à un niveau de flou objectivement égal, les scènes à l'endroit seraient subjectivement perçues comme plus nettes que les scènes à l'envers. Comme dans l'Expérience 1, l'interaction significative entre l'orientation de la scène en sample et le type de paire target-

Chapitre 2 – Contribution empirique

sample suggère que l'effet de l'orientation de la scène en sample sur la netteté perçue était plus important dans les paires target-sample différentes que dans les paires target-sample identiques, ce qui permet d'exclure les biais systématiques (tendance à sous-estimer le flou de la scène en sample par rapport à celui de la scène en target ; tendance à sous-estimer le flou des scènes à l'endroit indépendamment de ce à quoi elles sont comparées) qui pourraient expliquer nos résultats lorsque les scènes en target et en sample avaient une orientation différente. En outre, comme dans l'Expérience 1, le MBL moyen pour les paires target-sample identiques était inférieur à 15.5, ce qui suggère une tendance générale à attribuer moins de flou à la scène en sample que le niveau de flou requis pour la faire correspondre à la scène en target lorsqu'elles ont la même orientation, en particulier pour les scènes très floues avec un niveau de flou supérieur à 16 (voir la Figure 2.9 pour les MBL en fonction du niveau de flou de la scène en target pour les paires identiques). Dans l'ensemble, les résultats de cette expérience confirment ceux de Lupyan (2017) sur la perception de mots et ceux de l'Expérience 1 sur la perception d'objets en contexte et nous permettent d'étendre ces conclusions à la perception de scènes entières. Ainsi, le rehaussement perceptif de stimuli prédictibles serait un principe plus général de la vision qui s'applique à de nombreux stimuli, quel que soit le niveau de détails dont ils ont besoin pour être identifiés.

L'analyse post-hoc de l'effet du niveau de flou de la scène en target n'a pas permis de reproduire les résultats de l'Expérience 1. Bien que l'effet de l'orientation de la scène lorsque les scènes en target et en sample étaient différentes ait eu tendance à augmenter avec le niveau de flou de la scène target, cette différence n'était pas significative dans l'Expérience 2. Comme dans l'Expérience 1, ces analyses exploratoires ont été effectuées pour affiner notre interprétation des résultats. Cependant, les Expériences 1 et 2 n'ont pas été conçues pour vérifier cette hypothèse. En particulier, afin de réduire le nombre d'essais et la durée des deux expériences, chacun des 20 stimuli utilisés en target n'a été présenté que deux fois par condition en tant que target, et ils ont été associés à seulement deux niveaux de flou différents. Ainsi, le niveau de flou de la target et l'identité du stimulus en target n'étaient pas indépendants (i.e., un stimulus en target vu avec un niveau de flou élevé n'était pas nécessairement vu avec un niveau de flou faible). Par conséquent, nous ne pouvons pas garantir que les tendances observées au niveau du flou de l'image target sont bien dues à la fiabilité du signal et non à l'identité de l'image target.

La tâche d'ajustement perceptif utilisée dans les Expériences 1 et 2 nous a permis d'évaluer implicitement la netteté perçue relative des stimuli prédictibles et non prédictibles. Le

Chapitre 2 – Contribution empirique

fait que les participants aient rendu les stimuli prédictibles plus flous que les stimuli non prédictibles pour juger les deux stimuli comme étant aussi flous l'un que l'autre nous a conduit à l'interprétation qu'à niveau égal de flou, les stimuli prédictibles sont perçus comme plus nets que les stimuli non prédictibles. Dans l'Expérience 3, nous avons voulu vérifier si des résultats similaires pouvaient être retrouvés en utilisant cette fois-ci une tâche de comparaison explicite de la netteté des stimuli visuels. Dans cette nouvelle expérience, nous avons présenté aux participants les paires de scènes à l'endroit et à l'envers utilisées dans l'Expérience 2, mais cette fois-ci, les participants devaient indiquer s'ils percevaient une différence de flou entre les deux scènes et, le cas échéant, laquelle des deux scènes était perçue comme la plus floue ou la plus nette. Dans la majorité des essais, les deux scènes avaient objectivement le même niveau de flou. Conformément aux résultats des Expériences 1 et 2, nous nous attendions à ce que les participants jugent explicitement les scènes à l'endroit comme plus nettes que les scènes à l'envers dans ces essais. Nous avons également inclus des essais dans lesquels les deux scènes avaient un niveau de flou objectivement différent afin que les participants ne se rendent pas compte que les deux scènes avaient souvent le même niveau de flou, mais aussi pour nous assurer qu'ils effectuaient correctement la tâche lorsque les deux scènes étaient objectivement d'un niveau de flou différent. Dans cette expérience, nous avons également enregistré les mouvements oculaires des participants afin d'évaluer dans quelle mesure les stimuli prédictibles et non prédictibles suscitaient des différences d'attention visuelle, indexée par la durée de fixation. En effet, plusieurs études ont montré que l'attention visuelle portée à un stimulus améliore sa perception, par exemple en augmentant le contraste ou la résolution spatiale perçus (Carrasco et al., 2004; Yeshurun & Carrasco, 1998). Il est donc possible que l'augmentation de la netteté perçue pour les stimuli prédictibles puisse être expliquée par le fait que ces stimuli ont suscité plus d'attention visuelle que les stimuli non prédictibles.

1.3. Expérience 3 – Comparaison explicite du flou lors de la perception subjective de scènes prédictibles et non prédictibles

1.3.1. Méthode

1.3.1.1. Participants et Design

Trente-cinq étudiants en licence de psychologie de l'Université Grenoble Alpes (31 femmes, $M_{\text{age}} = 19.63$, $SD_{\text{age}} = 2.06$) ayant une vision normale ou corrigée à la normale ont

Chapitre 2 – Contribution empirique

participé à l'expérience. La taille de l'échantillon a été fixée sur la base d'une taille d'effet estimée de $d_z = 0.93$ à partir de l'effet observé de la prédictibilité dans la condition de paires target-sample différentes de l'Expérience 2, afin d'obtenir une puissance (i.e., $1 - \beta$, β étant le risque d'erreur de type 2) de 0.8 avec un seuil alpha de .025 après correction de Bonferroni pour les deux tests non orthogonaux prévus.

1.3.1.2. Stimuli et Procédure

Les stimuli étaient exactement les mêmes scènes floues à l'endroit et à l'envers que celles utilisées dans l'Expérience 2. Contrairement aux Expériences 1 et 2, cette expérience a été réalisée en laboratoire, ce qui nous a permis de coupler nos mesures comportementales à des mesures oculométriques. L'expérience a été construite en utilisant la Psychtoolbox (Brainard, 1997; Pelli, 1997) implémentée dans MATLAB R2019b (MathWorks, Natick, MA, USA). Les stimuli étaient affichés sur un écran de 30 pouces (DELL ULTRASHARP) avec une résolution de 1920×1080 pixels. Les participants posaient leur tête sur une mentonnière placée à 70 cm de l'écran afin de garder leur tête fixe pour l'enregistrement des mouvements oculaires. Les mouvements oculaires ont été enregistrés à l'aide d'un eye-tracker Eyelink 1000 (SR Research, Ontario, Canada) avec un taux d'échantillonnage de 1000 Hz et une résolution spatiale nominale de 0.01° d'angle visuel. Pour chaque participant, nous avons enregistré uniquement l'œil gauche en utilisant le mode « réflexion pupillaire-cornéenne ». Le logiciel Eyelink a détecté automatiquement les saccades avec les seuils suivants : vitesse $>30^\circ/s$, accélération $>8000^\circ/s^2$ et déplacement saccadique $>0.15^\circ$. Les fixations étaient détectées lorsque la pupille était visible et qu'aucune saccade n'était en cours. Les clignements étaient détectés pendant l'occlusion partielle ou totale de la pupille. Avant de commencer l'expérience et tous les 50 essais, nous avons mis en place une procédure de calibration dans laquelle les participants devaient orienter leur regard vers cinq points distincts apparaissant séquentiellement au centre et sur chaque bord de l'écran. Une correction de drift était également effectuée tous les 10 essais.

Dans cette expérience, un essai débutait par une croix de fixation apparaissant au centre de l'écran. Ensuite, deux versions de la même scène étaient affichées à gauche et à droite de l'écran sur un fond gris (luminance moyenne de 0.5) à chaque essai. La session expérimentale contenait deux types d'essais : les essais *test* et les essais *catch*. Dans les essais *test* (Figure 2.10.A), les deux scènes avaient objectivement le même niveau de flou, qui pouvait être de 5 (flou faible), 15 (flou moyen) ou 25 (flou élevé). Dans cette expérience, nous n'avons utilisé

Chapitre 2 – Contribution empirique

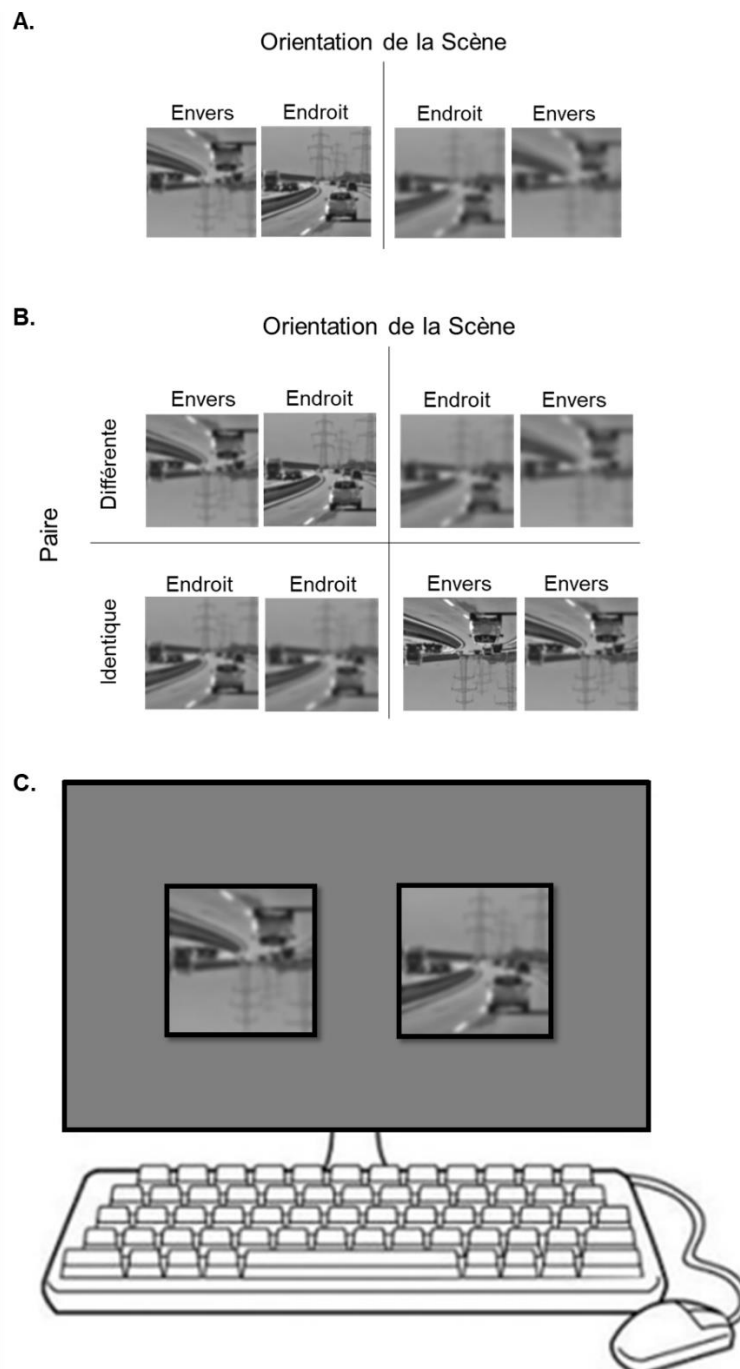
que trois niveaux de flou afin de nous assurer que, contrairement aux Expériences 1 et 2, tous les stimuli étaient vus à tous les niveaux de flou tout en gardant l'expérience relativement courte. Pour les essais test, les deux scènes avaient toujours une orientation différente (une scène prédictible à l'endroit et une scène non prédictible à l'envers). La scène à l'endroit était affichée à droite de l'écran et la scène à l'envers était affichée à gauche dans la moitié des essais test et l'inverse était fait dans l'autre moitié. Les participants ont reçu pour instruction de trouver laquelle des deux scènes était la plus floue ou la plus nette. Ils ont été informés que la différence de flou entre les scènes pouvait être subtile. Il est toutefois possible que les participants aient un doute sur l'existence d'une différence de flou entre les deux scènes et qu'ils soient alors enclins à sélectionner systématiquement la scène la plus familière, c'est-à-dire la scène à l'endroit, peu importe la tâche demandée (Han & VanRullen, 2016). Afin de vérifier la présence d'un tel biais de réponse, la moitié des participants ont reçu pour instruction d'indiquer la scène qu'ils jugeaient la plus floue, tandis que l'autre moitié devait signaler la scène qu'ils jugeaient la plus nette. Ils avaient également la possibilité de répondre qu'ils ne percevaient aucune différence de flou entre les deux scènes. Les essais test étaient les essais d'intérêt nous permettant de tester notre hypothèse selon laquelle, à un niveau de flou objectivement égal, les participants percevaient les scènes à l'endroit comme plus nettes que les scènes à l'envers. Nous avons également inclus des essais catch (Figure 2.10.B) pour nous assurer que les participants effectuaient la tâche correctement lorsque le flou des deux scènes était objectivement différent et pour éviter qu'ils ne se rendent compte que les deux scènes avaient souvent le même niveau de flou. Dans ces essais catch, les deux scènes avaient des niveaux de flou différents. L'une des deux scènes avait un niveau de flou de 5, 15 ou 25. L'autre avait un niveau de flou de 3 à 7 niveaux de flou au-dessus ou en-dessous. La scène la plus floue se trouvait à droite pour la moitié des essais et à gauche pour l'autre moitié. En outre, pour les essais catch, les deux scènes avaient soit une orientation différente (i.e., une scène prédictible à l'endroit et une scène non prédictible à l'envers), soit une orientation identique (i.e., deux scènes à l'endroit ou deux scènes à l'envers). Les essais catch avec une orientation identique avaient pour objectif de dissuader les participants d'adopter une stratégie consistant à sélectionner systématiquement la scène à l'endroit ou à l'envers. Étant donné que l'Expérience 2 a révélé que les participants attribuaient un niveau de flou plus élevé pour les scènes à l'endroit (jusqu'à 1.73 niveaux de flou en moyenne) que pour les scènes à l'envers pour percevoir les deux scènes comme étant aussi floues l'une que l'autre, la différence minimale de niveau de flou entre les deux scènes dans les essais catch a été fixée à 3 afin de garantir que les deux scènes étaient bien

Chapitre 2 – Contribution empirique

subjectivement perçues comme ayant un niveau de flou différent. Les participants devaient donner leur réponse (Figure 2.10.C) en appuyant sur la flèche du haut du clavier d'ordinateur s'ils ne percevaient pas de différence entre les deux scènes, sur la flèche de gauche s'ils jugeaient la scène de gauche la plus floue (ou la plus nette) et sur la flèche de droite s'ils jugeaient la scène de droite la plus floue (ou la plus nette).

Figure 2.10

Illustration des Conditions Expérimentales et du Déroulement de l'Expérience 3.



Chapitre 2 – Contribution empirique

Note. A. Exemple de stimuli présentés dans les essais test. Les deux scènes d'une paire avaient toujours une orientation différente (une scène à l'endroit et une scène à l'envers), mais le même niveau de flou. B. Exemple de stimuli présentés dans des essais catch. Les deux scènes d'une paire avaient soit la même orientation, soit une orientation différente et elles avaient des niveaux de flou différents. C. Illustration d'un essai. Deux scènes floues à l'endroit ou à l'envers apparaissaient simultanément avec des niveaux de flou identiques (essais test) ou différents (essais catch). Les participants devaient indiquer laquelle des deux scènes était la plus floue ou la plus nette, selon leur groupe expérimental.

Une session expérimentale comprenait 120 essais test (20 paires de scènes \times 2 combinaisons d'orientation [scène de gauche à l'endroit - scène de droite à l'envers vs. scène de gauche à l'envers - scène de droite à l'endroit] \times 3 niveaux de flou [5 vs. 15 vs. 25]) et 80 essais catch (20 paires de scènes \times 2 types de paires [les deux scènes ayant des orientations identiques vs. différentes] \times 2 orientations [scène à l'endroit vs. scène à l'envers]). Cela a donné un total de 200 essais apparaissant dans un ordre aléatoire avec une pause tous les 50 essais. Pour chaque essai, nous avons enregistré si les participants percevaient une différence de flou entre les scènes et si oui, quelle scène était choisie comme étant la plus floue ou la plus nette selon le groupe expérimental. Nous avons également enregistré les durées de fixation oculaire sur chaque scène. L'expérience durait environ 25 minutes, en fonction de la vitesse à laquelle les participants effectuaient leur jugement (durée moyenne \pm *SD* d'un essai : 3124 ± 2924 ms).

1.3.2. Analyse des données

1.3.2.1. Données comportementales

Les données ont été analysées à l'aide des mêmes logiciels et package que les Expériences 1 et 2. Concernant l'analyse des essais test (i.e., ceux pour lesquels les deux scènes avaient le même niveau de flou), nous avons utilisé deux modèles de régression logistique à effets mixtes. La première régression logistique visait à évaluer si les participants percevaient ou non une différence de flou entre les deux images (i.e., la variable dépendante était différence perçue vs. pas de différence perçue). La deuxième régression logistique nous a permis d'évaluer si les participants jugeaient plus souvent la scène prédictible ou non prédictible comme étant la plus floue dans les essais test (i.e., la variable dépendante était scène prédictible jugée comme étant la plus floue vs. scène non prédictible jugée comme étant la plus floue). Pour construire

Chapitre 2 – Contribution empirique

cette seconde variable dépendante, nous avons converti les réponses des participants ayant pour tâche d'indiquer quelle scène était la plus nette en réponse indiquant quelle scène était la plus floue afin d'uniformiser les réponses de tous les participants quel que soit leur groupe expérimental. Pour les deux régressions logistiques, l'intercept et le niveau de flou des scènes ont été définis comme effets fixes. Les intercepts pour les sujets et les stimuli, ainsi que les pentes aléatoires par sujet et par stimulus pour l'effet du niveau de flou ont été spécifiés comme effets aléatoires. Sur la base des résultats de l'Expérience 2, nous nous attendions à ce qu'à un niveau de flou objectivement égal, les scènes à l'envers soient explicitement perçues comme plus nettes par rapport aux scènes à l'envers et donc que la proportion de scènes à l'envers jugées comme les plus floues soit supérieure à la proportion de scènes à l'envers jugées comme les plus nettes. Pour vérifier cette hypothèse, nous avons testé dans les essais test (1) si la proportion de différence de flou perçue était significativement supérieure au hasard et (2) si la proportion de scènes à l'envers jugées comme étant les plus floues était significativement supérieure au hasard. Comme dans l'Expérience 2, nous avons également vérifié si cet effet était modulé par la fiabilité du signal (i.e., le niveau de flou) des scènes dans les essais test. Si l'influence des prédictions sur la netteté perçue augmente lorsque la fiabilité du signal diminue, nous nous attendions à ce que la tendance à sélectionner la scène à l'envers comme la plus floue augmente avec le niveau de flou des paires de scènes (i.e., interaction entre l'effet de la prédictibilité et le niveau de flou). Pour tester cette hypothèse, nous avons calculé l'effet du niveau de flou sur la proportion de scènes à l'envers jugées comme les plus floues (la complémentaire de cette proportion étant donc la proportion de scènes à l'envers jugées comme les plus nettes) et nous avons également effectué des comparaisons par paire entre chaque niveau de flou.

Concernant l'analyse des essais catch (i.e., lorsque les deux scènes présentaient des niveaux de flou objectivement différents), nous avons également mis en place un modèle de régression logistique à effets mixtes avec, comme variable dépendante, le taux de bonnes réponses pour détecter des différences de flou. Le type de paire (i.e., les deux scènes ayant ou non la même orientation) a été introduit dans le modèle comme effet fixe. Les intercepts pour les sujets et les stimuli, ainsi que les pentes aléatoires par sujet et par stimulus pour l'effet du type de paire ont été spécifiés comme effets aléatoires. Sur la base de ce modèle, nous avons calculé si le taux de bonnes réponses dans chaque condition différait significativement du hasard afin de nous assurer que les participants percevaient la différence de niveau de flou entre les deux images et effectuaient la tâche correctement.

Chapitre 2 – Contribution empirique

Pour toutes les analyses, la significativité des effets a été testée avec un test de Wald. Le seuil de significativité a été fixé à .025 pour les effets principaux du niveau de flou sur les essais test après correction de Bonferroni pour deux régressions logistiques non orthogonales planifiées ($.05/2 = .025$), à .017 pour les trois comparaisons par paires effectuées a posteriori sur les essais test après correction de Bonferroni pour trois tests non orthogonaux ($.05/3 = .017$) et à .017 pour les analyses des essais catch après correction de Bonferroni pour trois tests non orthogonaux sur les taux de bonnes réponses.

1.3.2.2. Données des mouvements oculaires

Concernant les analyses des mouvements oculaires, nous avons défini pour chaque essai deux zones d'intérêt correspondant à l'emplacement des deux scènes sur l'écran. Nous avons ensuite extrait la durée totale de fixation dans chacune de ces zones d'intérêt pour chaque essai. Nous avons effectué des analyses des mouvements oculaires sur les essais test uniquement, afin de comparer les durées de fixation en tant qu'indice d'attention visuelle entre les scènes à l'endroit et à l'envers lorsque les deux scènes avaient un niveau de flou objectivement égal et étaient orientées différemment. Nous voulions ainsi vérifier si l'augmentation de la netteté perçue des scènes à l'endroit par rapport aux scènes à l'envers dans les essais test pouvait s'expliquer par une augmentation de l'attention visuelle sur les scènes à l'endroit par rapport aux scènes à l'envers (Yeshurun & Carrasco, 1998). Nous avons effectué une analyse linéaire à effets mixtes de l'orientation de la scène (à l'endroit ou à l'envers) avec la durée totale de fixation de la scène comme variable dépendante. Les termes d'intercept et d'orientation de la scène ont été introduits dans le modèle comme effets fixes. Les intercepts pour les sujets et les stimuli, ainsi que les pentes aléatoires par sujet et par stimulus pour l'effet de l'orientation de la scène ont été spécifiés comme effets aléatoires. Nous avons calculé l'effet principal de l'orientation de la scène (à l'endroit ou à l'envers) sur la durée de fixation afin d'évaluer les différences d'attention visuelle entre les scènes à l'endroit et à l'envers.

Les modèles mixtes ont été construits en utilisant la même méthode que pour les Expériences 1 et 2 (i.e., méthode de Bates, Kliegl, et al., 2015 pour construire des modèles parcimonieux).

1.3.3. Résultats

1.3.3.1. Résultats comportementaux

Les proportions moyennes de réponses pour chaque niveau de flou des essais test sont présentées dans la Figure 2.11.A. Le taux de bonnes réponses moyen pour chaque condition expérimentale des essais catch est présenté dans la Figure 2.11.B. Huit participants ont été exclus des analyses en raison d'un taux de bonnes réponses égal ou inférieur au hasard dans les essais catch, ce qui indique qu'ils ne se sont pas impliqués dans la tâche (en effet, ces participants ont systématiquement indiqué « aucune différence de flou »), ce qui donne un échantillon final de 27 participants.

Premièrement, les analyses effectuées sur les essais catch ont révélé que lorsque les deux scènes apparaissaient avec un niveau de flou objectivement différent, le taux de bonnes réponses pour indiquer laquelle des scènes était la plus floue était élevé pour chaque condition (type de paire différente : $M = 0.87$, $SE = 0.07$; type de paire identique : $M = 0.89$, $SE = 0.06$) et significativement supérieur au hasard (type de paire différent : $\beta = 2.37$, $z = 7.92$, $p < .001$; type de paire identique : $\beta = 2.58$, $z = 8.57$, $p < .001$). Bien que la différence minimale entre les niveaux de flou des deux scènes ait été fixée à 3 pour que les deux scènes soient bien perçues comme ayant un niveau de flou différent, nous nous sommes demandées si le taux de bonnes réponses diminuait lorsque la scène à l'endroit était objectivement plus floue que la scène à l'envers par rapport à l'inverse (lorsque la scène à l'envers était objectivement plus floue que la scène à l'endroit). Selon nos hypothèses théoriques, la scène à l'endroit devrait être perçue comme plus nette que la scène à l'envers. Cependant, lorsque la scène à l'endroit était objectivement plus floue que la scène à l'envers, l'augmentation de la netteté perçue de la scène à l'endroit a pu compenser la différence objective de niveaux de flou entre les deux scènes, ce qui aurait donc réduit la différence perçue. À cette fin, nous avons effectué une régression logistique à effets mixtes a posteriori sur les essais catch où les deux scènes étaient orientées différemment. L'intercept et l'orientation de la scène objectivement la plus floue (i.e., à l'endroit vs. à l'envers) ont été définis comme effets fixes. Les intercepts et les pentes aléatoires pour l'effet de l'orientation de la scène la plus floue ont été spécifiés comme effets aléatoires pour les sujets et les stimuli. Les résultats ont montré que le taux de bonnes réponses était significativement plus faible lorsque la scène à l'endroit était plus floue que la scène à l'envers ($M = 0.80$, $SE = 0.08$) que lorsque la scène à l'envers était plus floue que la scène à l'endroit

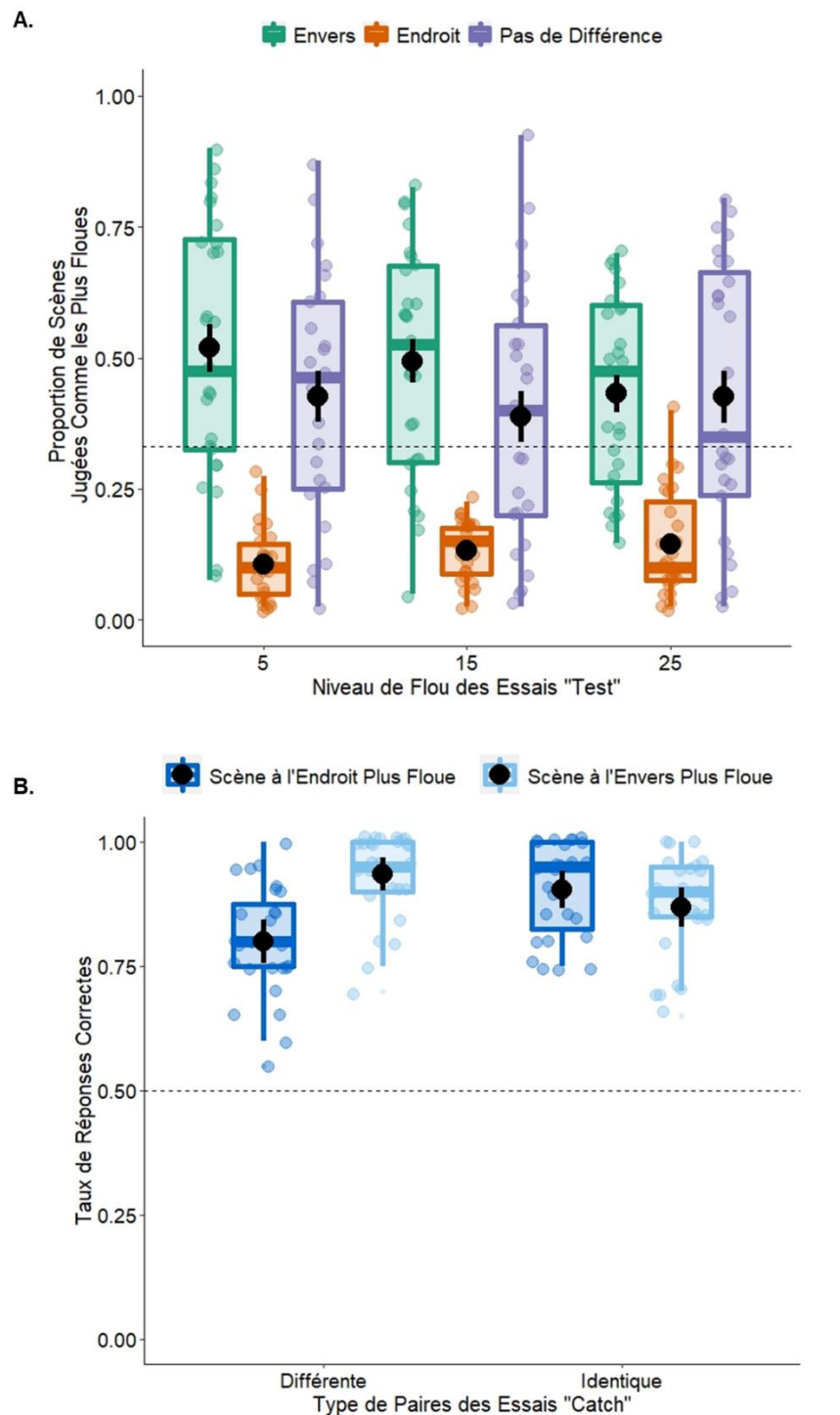
Chapitre 2 – Contribution empirique

($M = 0.94$, $SE = 0.05$), $\beta = 2.28$, $z = 5.45$, $p < .001$, ce qui plaide en faveur de nos hypothèses. Les participants auraient tendance à percevoir comme plus nette une scène à l'endroit, et dans le cas où son niveau de flou de départ serait plus élevé que celui de la scène à l'envers, elle serait alors perçue comme aussi nette que la scène à l'envers, conduisant à des réponses erronées.

En ce qui concerne les essais test dans lesquels une scène à l'envers était toujours comparée à une scène à l'endroit, les résultats ont montré que les participants ont perçu une différence de flou ($M = 0.61$, $SE = 0.09$) significativement plus fréquemment qu'une absence de différence de flou ($M = 0.39$, $SE = 0.09$), bien que les deux scènes aient le même niveau de flou objectif, $\beta = 0.68$, $z = 2.60$, $p = .009$, suggérant une influence de la prédictibilité des scènes sur la perception du flou. De plus, lorsqu'ils percevaient une différence de flou, les participants choisissaient la scène à l'envers comme étant la plus floue ($M = 0.79$, $SE = 0.08$) plus fréquemment que la scène à l'endroit ($M = 0.21$, $SE = 0.08$), $\beta = 1.42$, $z = 11.19$, $p < .001$. Ainsi, les participants ont explicitement jugé les scènes à l'endroit comme plus nettes que les scènes à l'envers à niveau de flou objectivement égal. Cependant, contrairement à ce qui était attendu, la proportion de choix de scènes à l'envers comme étant les plus floues a diminué à mesure que le niveau de flou des paires de scènes augmentait (pour un flou faible de 5 : $M = 0.62$, $SE = 0.09$; pour un flou moyen de 15 : $M = 0.63$, $SE = 0.09$; pour un flou élevé de 25 : $M = 0.57$, $SE = 0.10$), $\beta = -0.02$, $z = -2.32$, $p = .020$. Après correction de Bonferroni, les comparaisons par paire n'ont pas permis de mettre en évidence une différence significative sur la proportion de choix de scènes à l'envers comme étant les plus floues entre les différents niveaux de flou, ni d'un niveau de flou de 5 à 15, $\beta = 0.29$, $z = 1.75$, $p = .081$, ni d'un niveau de flou de 15 à 25, $\beta = 0.17$, $z = 1.13$, $p = .257$, ni d'un niveau de flou de 5 à 25, $\beta = 0.47$, $z = 2.32$, $p = .020$.

Figure 2.11

Proportion Moyenne de Scènes Jugées Comme Etant les Plus Floues Selon le Niveau de Flou Pour les Essais Test et Taux de Réponses Correctes Moyen Selon le Type de Paire et l'Orientation de la Scène Objectivement la Plus Floue Pour les Essais Catch.



Note. A. Proportion moyenne de scènes jugées comme les plus floues en fonction du niveau de flou pour les essais tests (i.e., les essais dans lesquels les deux scènes avaient le même niveau de flou objectif).

Chapitre 2 – Contribution empirique

Les deux scènes des essais test avaient toujours une orientation différente. Les points noirs et les barres d'erreur indiquent respectivement la moyenne et l'erreur standard de tous les participants. Les points de couleur correspondent aux observations individuelles. La ligne en pointillés représente le hasard. B. Taux de réponses correctes moyens pour les essais catch (i.e., les essais dans lesquels les deux scènes avaient un niveau de flou objectivement différent) en fonction de l'orientation de la scène objectivement la plus floue et du type de paire. Les deux scènes des essais catch avaient soit une orientation différente, soit une orientation identique. Les points noirs et les barres d'erreur indiquent respectivement la moyenne et l'erreur standard de tous les participants. Les points de couleur correspondent aux observations individuelles. La ligne en pointillés représente le hasard.

1.3.3.2. Résultats des mouvements oculaires

Les analyses effectuées sur les essais test ont révélé que les participants fixaient les scènes à l'envers ($M = 1849$ ms, $SE = 323$ ms) significativement plus longtemps que les scènes à l'endroit ($M = 1693$ ms, $SE = 286$ ms), $t(6433.00) = 5.32$, $p < .001$, $d_z = 1.024$, suggérant que plus d'attention visuelle était allouée aux scènes à l'envers qu'aux scènes à l'endroit. Cela suggère donc que la perception des scènes prédictibles comme plus nettes que les scènes non prédictibles ne résulte pas d'une attention visuelle plus importante portée aux scènes prédictibles.

1.3.4. Discussion de l'Expérience 3

Cette troisième expérience avait pour objectif de confirmer les résultats de l'augmentation de la netteté perçue des stimuli prédictibles par rapport aux stimuli non prédictibles, évalués implicitement dans les Expériences 1 et 2, en utilisant cette fois une tâche explicite de jugement de la netteté. À cette fin, les participants devaient décider laquelle des deux scènes présentées (à l'endroit ou à l'envers) était la plus floue, alors que les deux scènes étaient présentées dans la plupart des essais à un niveau de flou objectivement égal. Tout d'abord, les analyses ont montré que la plupart des participants se sont engagés dans la tâche car ils ont correctement perçu (plus de 85% du temps) quelle scène était la plus floue lorsque les deux scènes avaient un niveau de flou objectivement différent (analyse des essais catch). De façon critique, l'analyse des essais tests (lorsque les deux scènes avaient un niveau de flou objectivement égal) a mis en évidence (1) que les participants percevaient subjectivement une

Chapitre 2 – Contribution empirique

différence de flou et (2) qu'ils jugeaient les scènes à l'endroit comme plus nettes que les scènes à l'envers. Ces résultats viennent donc étayer les conclusions des Expériences 1 et 2 en montrant que les scènes prédictibles apparaissent explicitement plus nettes que les scènes non prédictibles, conformément aux hypothèses dérivées du mécanisme de sharpening.

De plus, l'analyse des essais catch apporte également des arguments en faveur du mécanisme de sharpening. En effet, lorsque les deux scènes avaient un niveau de flou objectivement différent, les performances pour déterminer quelle scène était la plus floue étaient plus faibles lorsque la scène à l'endroit était plus floue que la scène à l'envers que lorsque la scène à l'envers était plus floue que la scène à l'endroit. Ce résultat suggère que les participants étaient plus réticents à juger la scène prédictible comme étant la plus floue par rapport à la scène non prédictible, alors que c'était objectivement le cas. Cela pourrait s'expliquer par le mécanisme de sharpening dont bénéficie la scène prédictible par rapport à la scène non prédictible. En effet, lorsque la scène prédictible est objectivement la plus floue, elle serait cependant subjectivement perçue comme plus nette qu'elle ne l'est, conduisant à réduire la différence de flou perçue entre les deux scènes et donc à rendre la tâche plus difficile. Au contraire, lorsque la scène prédictible est objectivement plus nette que la scène non prédictible, le mécanisme de sharpening conduirait à ce qu'elle soit perçue comme encore plus nette, conduisant à accroître la différence de flou perçue entre les deux scènes et donc à rendre la tâche plus aisée.

Il a été montré que l'attention peut impacter la perception (Rungratsameetaweemana & Serences, 2019; Summerfield & de Lange, 2014) en améliorant le contraste perçu (Carrasco et al., 2004) et la résolution spatiale (Yeshurun & Carrasco, 1998) des stimuli bénéficiant de plus d'attention. Par conséquent, l'augmentation de la netteté perçue des stimuli prédictibles dans nos expériences pourrait s'expliquer par le fait que plus d'attention a été portée aux stimuli prédictibles qu'aux stimuli non prédictibles. Cependant, cette explication semble peu probable pour plusieurs raisons. Premièrement, la tâche utilisée dans cette expérience (ainsi que celle utilisée dans les Expériences 1 et 2) nécessitait de prêter attention à l'image de gauche comme à l'image de droite qui étaient prédictibles ou non prédictibles dans un nombre égal d'essais. Par conséquent, si une plus grande attention était, par exemple, accordée à l'image droite (ou au stimulus sample dans les Expériences 1 et 2) par rapport à l'image de gauche (ou au stimulus target dans les Expériences 1 et 2), les effets de l'attention sur la perception du flou des images auraient dû être similaires pour les stimuli prédictibles et non prédictibles, ce qui n'était pas le cas. De plus, plusieurs études ont montré que lorsqu'il s'agit de la perception d'objets et de

Chapitre 2 – Contribution empirique

scènes, l'attention a tendance à être captée par ce qui est nouveau ou surprenant (Johnston et al., 1990; Park et al., 2010). Dans le contexte de notre étude, une plus grande attention aurait donc dû être allouée aux stimuli non prédictibles - qui peuvent être considérés comme nouveaux puisqu'ils ne sont pas liés à une expérience antérieure - qu'aux stimuli prédictibles qui étaient plus familiers. Ceci a été confirmé par les analyses oculométriques de la durée de fixation des scènes de l'Expérience 3, qui ont montré que les stimuli non prédictibles étaient fixés plus longtemps que les stimuli prédictibles. Par conséquent, une plus grande attention aux stimuli non prédictibles ne peut pas expliquer l'augmentation de la netteté perçue des stimuli prédictibles observée dans nos études.

Enfin, contrairement à ce qui était attendu sur la base des Expériences 1 et 2, les participants n'ont pas nécessairement jugé la scène à l'envers comme étant la plus floue plus fréquemment à mesure que le niveau de flou des paires de scènes augmentait. Nous avons même observé la tendance inverse, à savoir que les participants ont jugé la scène à l'envers comme étant la plus floue moins souvent lorsque le niveau de flou des paires de scènes augmentait - et donc lorsque la fiabilité du signal diminuait. Une explication possible de ces résultats est que lorsque les scènes sont trop floues, les prédictions seraient trop diverses et donc trop peu informatives pour affiner davantage la représentation du stimulus. Il est également possible que lorsque le niveau de flou augmente dans la tâche, les scènes soient en fait trop floues pour être reliées à des connaissances préalables, même dans la condition où les scènes sont à l'endroit. Les scènes à l'endroit deviendraient donc moins prédictibles, réduisant ainsi l'effet de la prédictibilité sur la netteté perçue. Ces analyses en fonction du niveau de flou des stimuli ont été réalisées dans la suite des analyses post-hoc effectuées dans les Expériences 1 et 2 impliquant une tâche d'ajustement perceptif. L'Expérience 3 a alors été construite afin de tester explicitement cette hypothèse secondaire. Cependant, nous utilisons ici une tâche de comparaison de flou entre les deux images où nous mesurons la proportion de sélection de scènes à l'endroit/à l'envers comme étant les plus floues. Cette mesure pourrait ne pas être une mesure adéquate pour évaluer comment l'effet des prédictions sur la netteté perçue des stimuli est influencé par leur fiabilité. En effet, si cette mesure reflète la tendance à percevoir une scène comme plus nette/floue que l'autre, elle ne permet pas de quantifier cet effet (i.e., l'importance de la différence de netteté perçue entre les deux scènes). Par exemple, si l'effet de la prédictibilité des stimuli sur leur netteté perçue est plus important lorsque les stimuli ont une faible fiabilité, mais que les stimuli prédictibles sont systématiquement perçus comme plus nets que les stimuli non prédictibles pour toutes les conditions de fiabilité (pour tous les niveaux de

flou), la proportion de sélection des stimuli non prédictibles comme étant les plus flous risque d'être similaire dans toutes les conditions de fiabilité.

En conclusion, l'ensemble des trois premières expériences de ce travail de thèse fournit des arguments en faveur d'une influence des connaissances et des prédictions qui en découlent sur la perception subjective de la netteté de scènes et d'objets et d'un rehaussement perceptif (i.e., perception plus nette) des stimuli prédictibles (i.e., conformes à nos connaissances) par rapport aux stimuli non prédictibles. Ainsi, les connaissances que nous avons sur notre environnement visuel nous aident non seulement à reconnaître plus facilement les stimuli, mais nous permettent également d'améliorer leur perception. Par ailleurs, nos résultats suggèrent également que les effets perceptifs des prédictions seraient modulés par la qualité du signal visuel. Cependant, ces expériences n'étant au final pas adaptées pour tester spécifiquement cette hypothèse, nous avons mis en place une quatrième expérience pour la tester.

2. Facteurs influençant l'effet des prédictions sur la perception subjective

2.2. Expérience 4 – Etude de l'effet des prédictions en fonction de la fiabilité du signal

L'objectif de cette quatrième expérience était de préciser dans quelle mesure l'effet des prédictions sur la netteté perçue varie en fonction de la fiabilité du signal visuel et donc, dans le contexte de ce travail de thèse, en fonction du flou du signal. En effet, les théories du codage prédictif proposent que le poids relatif des signaux de prédiction et d'erreur de prédiction varierait en fonction de la fiabilité du signal visuel. Les signaux de prédiction auraient plus de poids lorsque le stimulus est bruité, tandis que les signaux d'erreur de prédiction auraient plus de poids lorsque le stimulus est fiable. Ainsi, l'influence des prédictions sur la perception visuelle augmenterait à mesure que la fiabilité du signal diminue, c'est-à-dire à mesure que le signal devient ambigu ou bruité (Feldman & Friston, 2010; Friston, 2005). Dans la même optique, l'opposing process theory (Press et al., 2020) propose également qu'un rehaussement perceptif de stimuli prédictibles (par rapport à des stimuli non prédictibles) aurait lieu seulement dans des conditions de signal visuel bruité. Dans le contexte des précédentes expériences (Expériences 1 et 2) qui n'avaient pas pour objectif de tester cette hypothèse, nous avons observé cet effet a posteriori pour l'Expérience 1 sur les objets en contexte, mais il n'était pas significatif pour l'Expérience 2 sur les scènes entières. Concernant l'Expérience 3 dans laquelle

Chapitre 2 – Contribution empirique

les participants devaient indiquer quelle scène était perçue comme la plus floue ou la plus nette, nous avons testé cette hypothèse qui ne s'est pas vérifiée. Cependant, la mesure que nous avons utilisée dans l'Expérience 3, c'est-à-dire la proportion de sélection de scènes à l'endroit/à l'envers comme étant les plus floues, pourrait ne pas être la mesure la plus adéquate pour quantifier comment l'effet des prédictions sur la netteté perçue des stimuli varie avec la fiabilité du signal.

Pour répondre à cette question, nous avons repris le paradigme d'ajustement perceptif utilisé dans l'Expérience 2 sur les scènes à l'endroit et à l'envers en ne conservant que la condition où les scènes en sample et en target étaient différentes. Nous avons donc présenté soit une scène sample à l'endroit dont les participants devaient ajuster le niveau de flou pour qu'il corresponde à celui de la scène target à l'envers, soit une scène sample à l'envers dont les participants devaient ajuster le niveau de flou pour qu'il corresponde à celui de la scène target à l'endroit. Afin de mieux contrôler le niveau de flou des scènes target et de présenter chaque target à tous les niveaux de flou (ce que nous n'avions pu faire dans les Expériences 1 et 2 afin de limiter la durée des expériences tout en évitant les effets d'apprentissage), chaque scène target pouvait être présentée à quatre niveaux de flou (8, 13, 18 et 23) allant d'une scène assez nette à une scène très floue. Le niveau de flou initial de la scène sample restait identique entre les essais, à un niveau de flou moyen. Comme pour les Expériences 1 et 2, nous nous attendions à ce que le niveau de flou attribué à la scène sample pour correspondre à celui de la scène target soit plus élevé pour les sample à l'endroit qu'à l'envers. Par ailleurs, si l'effet des prédictions sur la netteté perçue augmente à mesure que la fiabilité du signal diminue, nous nous attendions à ce que la différence entre le niveau de flou attribué à une scène sample à l'endroit et à l'envers augmente avec le niveau de flou de la scène target.

2.2.1. Méthode

2.2.1.1. Participants et Design

Trente-six étudiants en licence de psychologie de l'Université Grenoble Alpes (31 femmes, $M_{\text{age}} = 20.64$, $SD_{\text{age}} = 2.89$) ayant une vision normale ou corrigée à la normale ont participé à l'expérience. La taille de l'échantillon a été fixée sur la base d'une taille d'effet estimée de $d_z = 0.93$ à partir de l'effet de la prédictibilité pour les paires target-sample différentes observé dans l'Expérience 2 pour obtenir une puissance (i.e., correspondant à $1 - \beta$, β étant le

Chapitre 2 – Contribution empirique

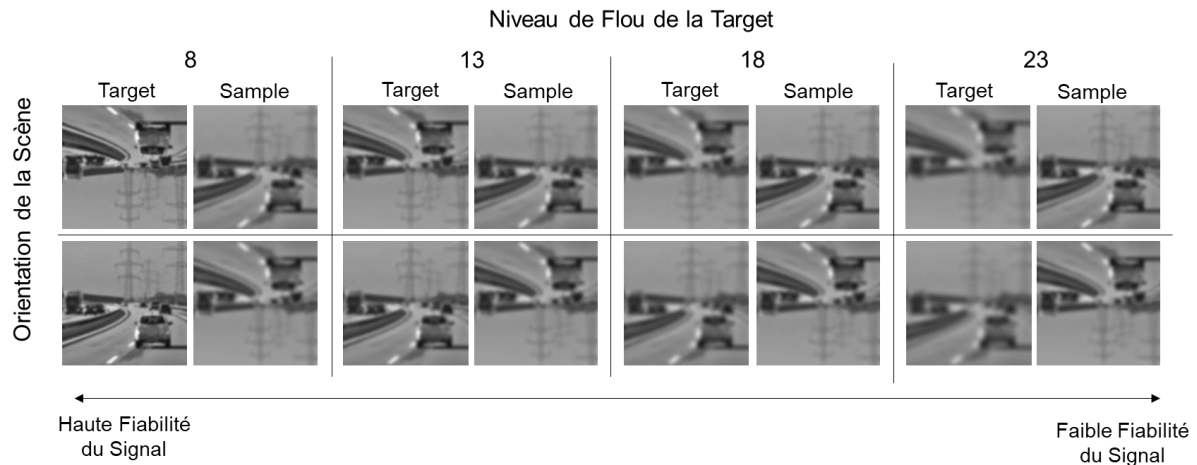
risque d'erreur de type 2) de 0.8 avec un seuil alpha de .017 après correction de Bonferroni pour les trois analyses non orthogonales prévues.

2.2.1.2. Stimuli

Les stimuli étaient exactement les mêmes scènes floues à l'endroit et à l'envers que celles utilisées dans l'Expérience 2. Comme pour l'Expérience 2, le niveau de flou initial de la scène sample a été fixé à 16 au début de chaque essai, mais chaque scène en target apparaissait cette fois-ci uniquement avec un niveau de flou de 8, 13, 18 ou 23, allant d'une scène relativement nette (8) à une scène très floue (23, Figure 2.12). Contrairement à l'Expérience 2, nous avons manipulé le niveau de flou de la scène target et l'identité du stimulus en target de façon indépendante (i.e., chaque stimulus target était présenté à chacun des quatre niveaux de flou) afin de tester que l'effet de la prédictibilité des scènes augmente réellement avec le niveau de flou de la scène en target.

Figure 2.12

Exemple d'une Scène Orientée à l'Endroit et à l'Envers à Différents Niveaux de Flou.



Note. Exemple des différentes conditions de l'expérience. La scène en target pouvait être orientée à l'endroit/envers avec une scène en sample orientée à l'envers/endroit. Le niveau de flou de la scène en target variait entre quatre niveaux (8, 13, 18 et 23) allant d'une scène assez nette ayant une haute fiabilité à une scène très floue ayant une faible fiabilité.

2.2.1.3. Procédure

Les paramètres expérimentaux (i.e., les stimuli affichés à l'aide du logiciel E-Prime Go 1.0 sur l'ordinateur personnel du participant) et la procédure étaient identiques à ceux de l'Expérience 2. Cependant, contrairement à l'Expérience 2, nous n'avons conservé que les types de paires target-sample ayant une orientation différente afin de limiter la durée de l'expérience et la scène en target apparaissait à un niveau de flou choisi aléatoirement parmi les quatre niveaux de flou possibles (8, 13, 18 ou 23), qui restait constant tout au long de l'essai.

Une session expérimentale contenait 160 essais apparaissant dans un ordre aléatoire (20 scènes \times 2 orientations [à l'endroit vs. à l'envers] \times 4 niveaux de flou) avec une pause aux premier et deuxième tiers de l'expérience. L'expérience durait environ 30-40 minutes, en fonction de la vitesse à laquelle chaque participant effectuait son jugement (durée moyenne \pm SD d'un essai : 1798 \pm 1982 ms). Pour chaque essai, nous avons enregistré le niveau de flou attribué à la scène sample pour correspondre à celui de la scène target (Matched Blur Level, MBL).

2.2.2. Analyse des données

Les données ont été analysées à l'aide du logiciel R (R Core Team, 2019) et du package *lme4* (Bates, Mächler, et al., 2015) pour effectuer une analyse linéaire à effets mixtes de l'orientation de la scène en sample (à l'endroit vs. à l'envers) et du niveau de flou de la scène en target (8, 13, 18 et 23) dans un design intra-sujet $2 \times$ continu¹, avec le MBL comme variable dépendante. Le niveau de flou de la scène target, l'orientation de la scène sample et leur terme d'interaction ont été introduits dans le modèle en tant qu'effets fixes. Les intercepts pour les sujets et les stimuli, ainsi que les pentes aléatoires par sujet et par stimulus pour l'effet du niveau de flou de la scène target, de l'orientation de la scène sample et de leur interaction ont été spécifiés comme effets aléatoires.

Similairement aux Expériences 1 et 2, nous avons testé si le MBL était plus élevé pour faire correspondre une scène sample à l'endroit à une scène target à l'envers que pour faire correspondre une scène sample à l'envers à une scène target à l'endroit. En outre, nous avons testé si cet effet de la prédictibilité était modulé par la fiabilité du signal (i.e., le niveau de flou)

¹ Le niveau de flou des scènes target a été traité comme une variable continue plutôt que catégorielle afin d'optimiser la puissance statistique des analyses.

Chapitre 2 – Contribution empirique

des scènes en target. Si l'influence des prédictions sur la netteté perçue augmente lorsque la fiabilité du signal diminue, nous nous attendions à ce que la différence de MBL entre les scènes sample à l'endroit et à l'envers augmente avec le niveau de flou de la scène target. Pour vérifier cette hypothèse, nous avons calculé l'interaction entre l'orientation de la scène en sample et le niveau de flou de la scène en target et nous avons également effectué des comparaisons par paires entre chaque niveau de flou.

Comme dans l'Expérience 2, nous avons utilisé la méthode proposée par Bates, Kliegl, et al. (2015) pour construire des modèles mixtes parcimonieux limitant les problèmes de convergence. L'inspection visuelle des graphiques de résidus n'a pas révélé de déviations évidentes de l'homoscédasticité ou de la normalité. Les p-valeurs ont été obtenues par approximation de Satterthwaite avec le package *lmerTest* (Kuznetsova et al., 2017). Le seuil de significativité a été fixé à .017 après correction de Bonferroni pour 3 tests non orthogonaux. Les tailles d'effet ont été estimées à l'aide du d_z de Cohen (Lakens, 2013).

2.2.3. Résultats

Les MBL moyens pour chaque condition expérimentale sont présentés dans la Figure 2.13. Les résultats ont révélé un effet principal de l'orientation de la scène en sample, $t(25.02) = 7.86$, $p < .001$, $d_z = 1.310$. Les participants attribuaient un MBL plus élevé à la scène en sample lorsqu'elle était à l'endroit ($M = 15.71$, $SE = 0.96$) qu'à l'envers ($M = 14.71$, $SE = 0.92$). De plus, l'interaction entre l'orientation de la scène en sample et le niveau de flou de la scène en target était significative, $t(18.77) = 3.15$, $p = .005$, $d_z = 0.525$. Conformément à nos hypothèses, la différence de MBL entre les conditions de sample à l'endroit et à l'envers augmentait à mesure que le niveau de flou de la scène en target augmentait (T-envers/S-endroit niveau de flou de la scène target de 8 : $M = 8.14$, $SE = 0.20$; T-endroit/S-envers niveau de flou de la scène target de 8 : $M = 7.66$, $SE = 0.23$; T-envers/S-endroit niveau de flou de la scène target de 13 : $M = 13.76$, $SE = 0.33$; T-endroit/S-envers niveau de flou de la scène target de 13 : $M = 12.64$, $SE = 0.33$; T-envers/S-endroit niveau de flou de la scène target de 18 : $M = 18.34$, $SE = 0.34$; T-endroit/S-envers niveau de flou de la scène target de 18 : $M = 17.16$, $SE = 0.35$; T-envers/S-endroit niveau de flou de la scène target de 23 : $M = 22.62$, $SE = 0.46$; T-endroit/S-envers niveau de flou de la scène target de 23 : $M = 21.39$, $SE = 0.45$). Plus précisément, la différence de MBL entre les conditions de scènes sample à l'endroit et à l'envers augmentait lorsque la scène en target passait d'un niveau de flou de 8 à 13, $t(2749) = 5.51$, $p < .001$, $d_z =$

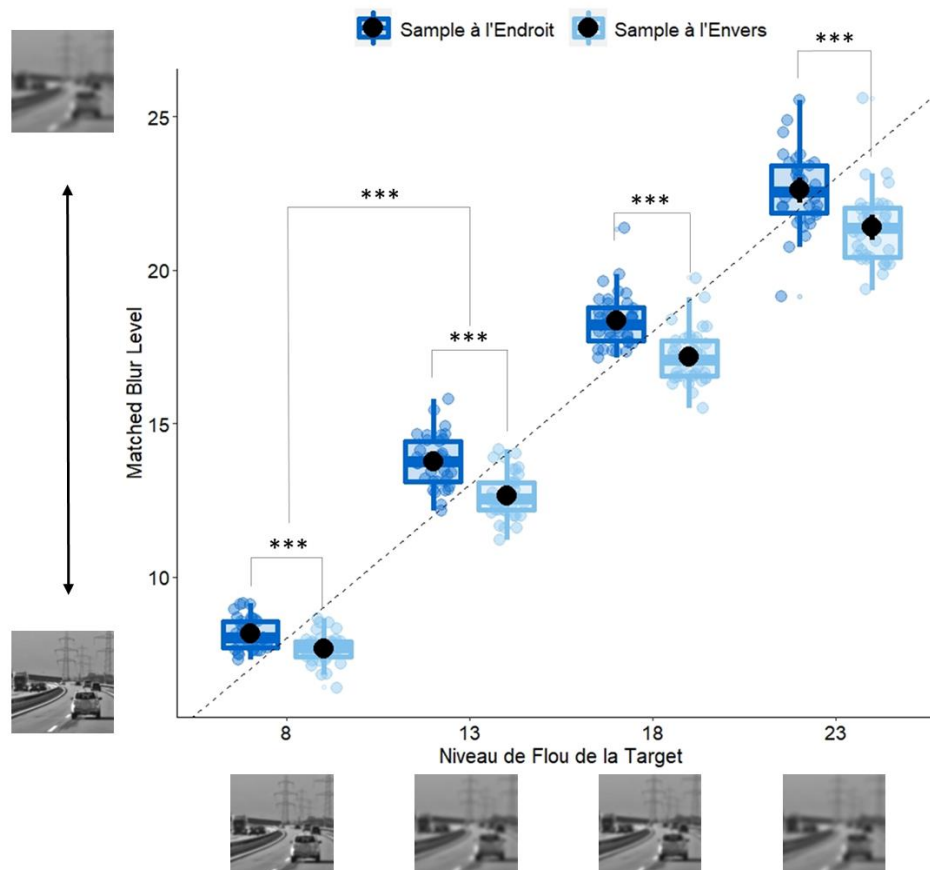
Chapitre 2 – Contribution empirique

0.918, mais pas lorsque la scène en target passait d'un niveau de flou de 13 à 18, $t(2749) = 0.47$, $p = .640$, $d_z = 0.078$, ni de 18 à 23, $t(2768) = 0.29$, $p = .772$, $d_z = 0.048$.

Nous avons également testé a posteriori si l'effet de la prédictibilité (i.e., l'effet de l'orientation de la scène en sample) était significatif pour chacun des quatre niveaux de flou. Les résultats ont montré que le MBL était significativement plus élevé pour les scènes sample à l'endroit que pour les scènes sample à l'envers, pour tous les niveaux de flou de la scène target (niveau de flou de 8, $t(5613) = 7.60$, $p < .001$, $d_z = 1.267$, niveau de flou de 13, $t(5613) = 15.65$, $p < .001$, $d_z = 2.608$, niveau de flou de 18, $t(5648) = 19.55$, $p < .001$, $d_z = 3.258$, niveau de flou de 23, $t(5648) = 15.44$, $p < .001$, $d_z = 2.573$).

Figure 2.13

Matched Blur Level Moyen pour Chaque Condition de l'Expérience 4.



Note. A. Niveau moyen de flou attribué à la scène sample pour correspondre au flou de la scène target (MBL) pour les scènes en sample à l'endroit et à l'envers en fonction du niveau de flou de la scène en target. Un niveau de flou élevé correspond à une scène très floue, tandis qu'un niveau de flou faible

Chapitre 2 – Contribution empirique

correspond à une scène relativement nette. La ligne en pointillés représente l'ajustement parfait de la scène en sample pour chaque niveau de flou de la scène en target. Les points noirs et les barres d'erreur indiquent respectivement la moyenne et l'erreur standard de tous les participants. Les points en couleur correspondent aux observations individuelles.

*** $p < .001$.

2.2.4. Discussion de l'Expérience 4

L'expérience 4 avait pour objectif de tester plus rigoureusement si l'effet des prédictions sur la netteté perçue varie en fonction de la fiabilité du signal visuel. Nous avons repris le paradigme d'appariement perceptif utilisé dans l'Expérience 2 sur les scènes à l'endroit et à l'envers en ne conservant que la condition où les scènes en sample et en target étaient différentes. Afin de contrôler le niveau de flou des stimuli et de présenter chaque scène à tous les niveaux de flou, nous avons manipulé seulement quatre niveaux de flou de la target (8, 13, 18 et 23) allant d'une scène assez nette à une scène très floue.

Nous avons répliqué les résultats de l'Expérience 2, à savoir que lorsque les deux scènes avaient une orientation différente, les participants attribuaient un niveau de flou plus élevé aux scènes sample à l'endroit qu'à l'envers pour ajuster le niveau de flou de la scène sample à celui de la scène target. Ce résultat renforce les résultats des expériences précédentes et soutient le fait que les scènes à l'endroit, et donc prédictibles, seraient perçues comme plus nettes que les scènes à l'envers, moins prédictibles. De plus, nos résultats ont confirmé les analyses post-hoc de l'Expérience 1 en montrant que l'effet des prédictions sur la netteté perçue augmentait à mesure que la fiabilité du signal diminuait (i.e., lorsque le niveau de flou de la scène target augmentait). L'influence des prédictions serait ainsi plus importante lorsque le signal est ambigu que lorsqu'il est fiable comme l'ont suggéré les versions les plus récentes des modèles du codage prédictif (Feldman & Friston, 2010). Cependant, contrairement à ce qui était attendu, l'effet des prédictions sur la netteté perçue n'augmentait pas linéairement avec le niveau de flou de la scène target. En effet, la différence entre le niveau de flou attribué à une scène sample à l'endroit et à l'envers augmentait lorsque la fiabilité de la scène target passait d'un niveau élevé (i.e., niveau de flou de 8) à un niveau moyen (i.e., niveau de flou de 13), puis restait constant jusqu'à ce que la scène target soit très floue et ait donc un niveau de fiabilité très faible (i.e., niveau de flou de 18 et 23). Ces résultats soutiennent donc une modulation de l'effet des prédictions sur la netteté perçue en fonction du degré de fiabilité du signal, mais au-delà d'un

Chapitre 2 – Contribution empirique

certain seuil de fiabilité, l'influence des prédictions sur la perception cesserait d'augmenter. Une explication possible de ces résultats est que lorsque les scènes sont trop floues (i.e., signal trop peu fiable), les prédictions seraient trop diverses et donc trop grossières (e.g., prédictions sur la catégorie de la scène plutôt que sur les détails des objets qui la composent) pour permettre d'affiner davantage la représentation du stimulus. En résumé, les résultats des Expériences 1 à 4 mettent en évidence un rehaussement perceptif des stimuli prédictibles, par rapport au stimuli non prédictibles et cet effet est d'autant plus fort que la fiabilité du signal visuel diminue (i.e., que les stimuli deviennent flous). Il faut toutefois noter que bien que les résultats de toutes ces expériences soutiennent l'idée d'un rehaussement perceptif des caractéristiques attendues des stimuli visuels, en accord avec le mécanisme de sharpening, ils ne remettent pas nécessairement en cause l'existence du mécanisme de dampening (de Lange et al., 2018; Kumar et al., 2017; Press et al., 2020; Richter et al., 2018) qui suppose le rehaussement perceptif des caractéristiques qui ne correspondent pas aux prédictions. En effet, nos manipulations expérimentales concernent jusqu'ici la présence ou l'absence de contenu informatif permettant d'accéder à des connaissances antérieures et donc d'élaborer des prédictions (i.e., présence/absence d'un contexte dans l'Expérience 1 ; présence/absence d'une expérience antérieure avec l'orientation des scènes dans les Expériences 2 à 4). Afin de distinguer les mécanismes de sharpening et de dampening, il est nécessaire d'inclure une condition dans laquelle les stimuli peuvent être inattendus/surprenants et donc contraires aux prédictions (e.g., une voiture dans un contexte intact de cuisine), ce que nous n'avons pas fait jusqu'ici. Dans les expériences suivantes (Expériences 5 et 6) nous avons inclus cette condition. Plus précisément, nous avons manipulé ce que nous appelons la validité des prédictions : nous avons utilisé des images de scènes contenant des objets congruents avec le contexte de la scène (stimuli conduisant à des prédictions valides sur l'objet, comme dans nos précédentes expériences), mais aussi des objets incongruents avec le contexte de la scène (stimuli conduisant à des prédictions non valides sur l'objet). Dans l'Expérience 5, nous avons étudié l'effet de la validité des prédictions en tenant compte de la fiabilité du signal visuel (plus précisément dans le cas de notre expérience, par le niveau de flou des stimuli). Dans l'Expérience 6, nous avons étudié comment l'effet de la validité était influencé par des contraintes de temps de traitement de l'information visuelle. Ces deux expériences visaient plus spécifiquement à tester les hypothèses formulées par l'opposing process theory (Press et al., 2020), jamais testées systématiquement dans le contexte d'une même étude.

Chapitre 2 – Contribution empirique

2.3. Expérience 5 – Etude de l'effet des prédictions en fonction de la validité des prédictions et de la fiabilité du signal

N.B. Dans cette partie expérimentale, nous reportons les résultats d'une expérience (Expérience 5) qui a été publiée dans la revue *Vision Research* (Rossel et al., 2023; Annexe D).

Objectif des Expériences 5 et 6

L'objectif de ces deux expériences était de tester explicitement les hypothèses issues de l'*opposing process theory*, qui propose que les mécanismes de *sharpening* et de *dampening* interviennent alternativement en fonction du décours temporel du traitement visuel et de la fiabilité du signal visuel (Press et al., 2020). Selon cette théorie, la perception serait initialement biaisée en faveur de ce qui est attendu (*sharpening*), permettant de confirmer les prédictions. Le mécanisme de *dampening* prendrait ensuite le relais pour réhausser le traitement des caractéristiques inattendues du signal (e.g., Yon & Press, 2017). Cette séquence serait en outre contrainte par la qualité (ou la fiabilité) des entrées sensorielles. Le mécanisme de *sharpening* prédominerait lorsque le signal est bruité mais que nous avons de fortes prédictions à son sujet (prédictions valides), tandis que le mécanisme de *dampening* ne prendrait le dessus que lorsqu'un signal fiable diverge fortement des prédictions (prédictions non valides). Ainsi, nous avons mené les Expériences 5 et 6 de façon complémentaire, afin de tester pour chacune une hypothèse spécifique de l'*opposing process theory* : l'effet des prédictions sur la perception subjective en fonction de leur validité et en tenant compte de la fiabilité du signal visuel dans l'Expérience 5 et en tenant compte des contraintes de temps de traitement de l'information visuelle dans l'Expérience 6.

Dans l'Expérience 5, nous avons donc étudié comment la perception subjective d'objets varie en fonction de leur caractère attendu ou inattendu (i.e., la validité des prédictions) ainsi que de leur niveau de flou (i.e., leur fiabilité, un stimulus flou étant peu fiable). Nous avons repris le paradigme d'ajustement perceptif du niveau de flou utilisé lors de l'Expérience 1 sur des objets en contexte. Dans l'Expérience 5, chaque objet apparaissait soit dans le contexte d'une scène sémantiquement congruente (e.g., un bateau dans un port), condition expérimentale qui devrait conduire à des prédictions correctes/valides sur l'objet, soit dans le contexte d'une scène sémantiquement incongruente (e.g., un bateau dans une chambre), condition expérimentale qui devait conduire à des prédictions incorrectes/non valides sur l'objet. À

Chapitre 2 – Contribution empirique

chaque essai, deux images (une image target et une image sample) contenant exactement le même objet étaient affichées à gauche (target) et à droite (sample) de l'écran sur un fond gris. L'objet en target apparaissait dans la première moitié des essais à un niveau de flou élevé de 18 (faible fiabilité du signal) et dans la seconde moitié des essais à un faible niveau de flou de 5 (haute fiabilité du signal). Enfin, pour chaque essai, les objets en target et en sample avaient soit un contexte différent (l'un congruent et l'autre incongruent avec l'objet, i.e., paires target-sample différentes), soit le même contexte (les deux congruents ou incongruents, paires target-sample identiques).

Les paires target-sample différentes ont été utilisées pour adresser spécifiquement notre problématique de recherche : Comment l'effet des prédictions sur la netteté perçue des objets varie en fonction de la validité de ces prédictions et de la fiabilité du signal visuel. D'après l'opposing process theory (Press et al., 2020), le mécanisme de sharpening (i.e., le rehaussement perceptif des caractéristiques prédites du stimulus) devrait s'appliquer lorsque les stimuli sont ambigus et donc que le signal visuel est peu fiable tandis que le mécanisme de dampening (i.e., le rehaussement perceptif des caractéristiques inattendues du stimulus) devrait s'appliquer lorsque le signal est fiable. Nous nous attendions donc à un effet d'interaction entre la congruence objet-contexte (i.e., validité des prédictions) et la fiabilité des objets. Plus spécifiquement, nous nous attendions à ce que dans la condition de faible fiabilité de l'objet (objet flou), les objets congruents avec leur contexte soient subjectivement perçus comme plus nets que ces mêmes objets dans un contexte incongruent, à un niveau de flou objectivement égal entre la target et le sample. Par conséquent, pour les paires target-sample différentes en condition de faible fiabilité, le niveau de flou attribué aux objets sample dans la condition T-incongruent/S-congruent devrait être plus élevé que le celui attribué aux objets sample dans la condition T-congruent/S-incongruent. A l'inverse, nous nous attendions à ce que dans la condition de haute fiabilité de l'objet (objet net), les objets incongruents avec leur contexte soient subjectivement perçus comme plus nets que ces mêmes objets congruents, à un niveau de flou objectivement égal entre la target et le sample. Par conséquent, pour les paires target-sample différentes en condition de haute fiabilité, le niveau de flou attribué aux objets sample dans la condition T-congruent/S-incongruent devrait être plus élevé que celui attribué aux objets sample dans la condition T-incongruent/S-congruent.

Comme dans les Expériences 1 et 2, nous avons également utilisé des paires target-sample identiques comme condition contrôle afin d'évaluer toute tendance générale à surestimer ou sous-estimer le flou de l'objet en sample par rapport à l'objet en target,

Chapitre 2 – Contribution empirique

indépendamment de son contexte (et donc indépendamment de la validité des prédictions), ou pour détecter un biais plus systématique de sous-estimation du flou des objets dans une condition, indépendamment de l'objet présenté en target. Pour rappel, si des facteurs de haut niveau (tels que les prédictions fondées sur des informations contextuelles) influencent la perception subjective des objets, ces effets perceptifs devraient affecter de la même manière tous les objets présentés simultanément. Ainsi, deux objets entourés du même contexte en target et en sample devraient bénéficier des mêmes effets perceptifs ce qui devrait résulter en l'absence de différence de flou perçue entre les deux (cf. Raisonnement du El Greco fallacy, Firestone & Scholl, 2014). Nous nous attendons donc à observer une interaction double entre le type de paire target-sample, la congruence objet-contexte et la fiabilité des objets, et plus spécifiquement à ce que l'interaction entre la congruence objet-contexte et la fiabilité des objets soit réduite pour les paires target-sample identiques par rapport aux paires target-sample différentes.

2.3.1. Méthode

2.3.1.1. Participants et Design

Cinquante-trois étudiants en licence de psychologie de l'Université Grenoble Alpes (50 femmes, $M_{\text{age}} = 21.43$, $SD_{\text{age}} = 5.54$) ayant une vision normale ou corrigée à la normale ont participé à l'expérience. La taille de l'échantillon a été fixée sur la base d'une taille d'effet estimée de $d_z = 0.45$ à partir de l'effet simple de la prédictibilité des objets sur la perception du flou pour les paires target-sample ayant des informations contextuelles différentes (i.e., comparaison entre (1) le niveau de flou attribué à un objet sample avec un contexte intact apparié à un objet target avec un contexte brouillé et (2) le niveau de flou attribué à un objet sample avec un contexte brouillé apparié à un objet target avec un contexte intact) dans l'Expérience 1 pour atteindre une puissance (i.e., correspondant à $1 - \beta$, β étant le risque d'erreur de type 2) de 0.8 avec un seuil alpha de .025 après correction de Bonferroni pour les deux analyses non orthogonales prévues ($.05/2 = .025$).

2.3.1.2. Stimuli

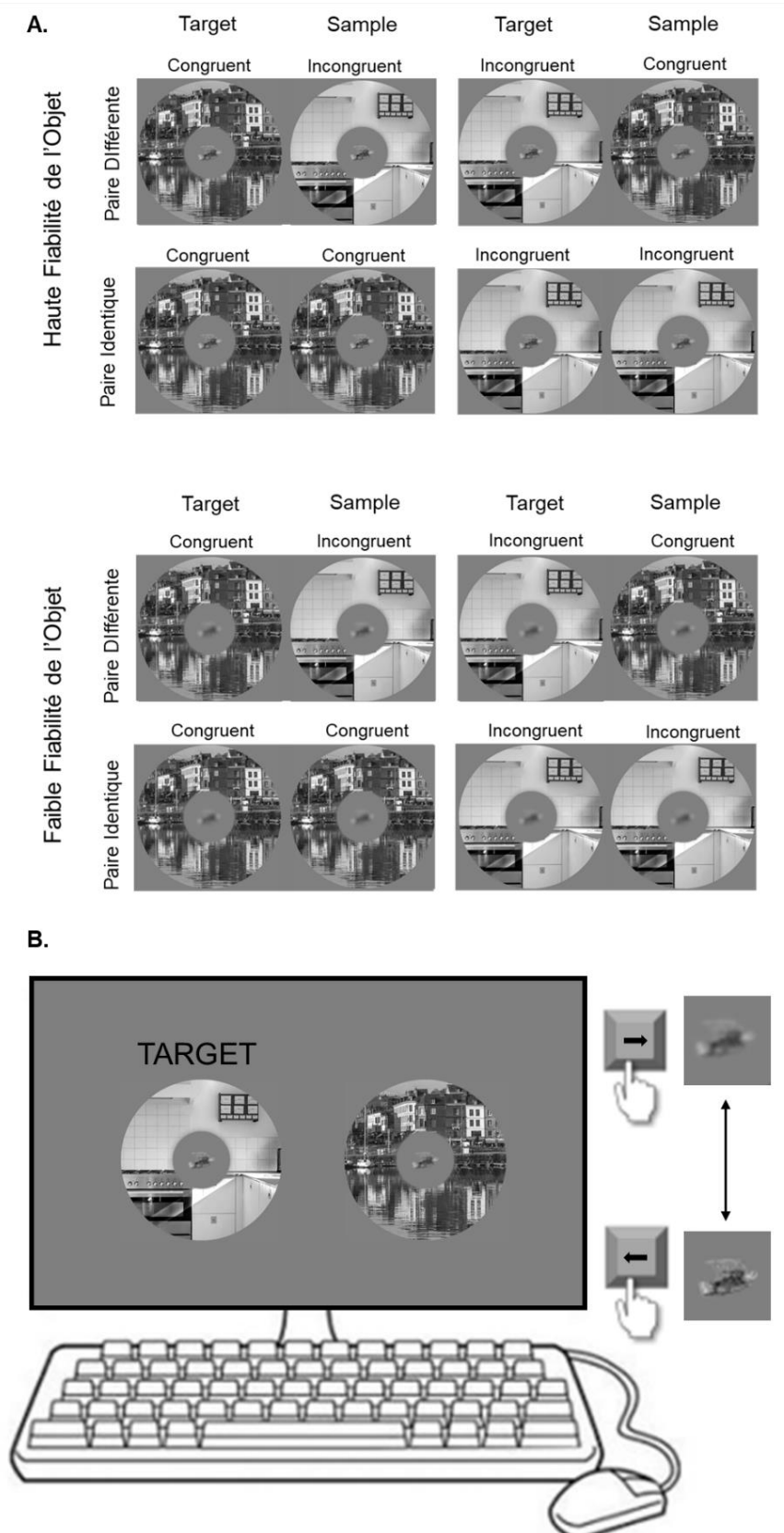
Pour cette étude nous avons construits de nouveaux stimuli. Ceux-ci étaient constitués de 20 photographies en couleur de scènes (10 scènes d'intérieur et 10 scènes d'extérieur) et de

Chapitre 2 – Contribution empirique

20 photographies en couleur d'objets (10 objets animés et 10 objets non animés). Les photographies ont été téléchargées sur le site Pixabay (<https://pixabay.com/fr>) et sur la base de données Google Images en filtrant pour la licence CC0. Les stimuli ont été créés à l'aide de MATLAB R2019b (MathWorks, Natick, MA, USA).

Figure 2.14

Illustration des Conditions Expérimentales et du Déroulement de l'Expérience 5.



Chapitre 2 – Contribution empirique

Note. A. Exemple de stimuli présentés dans chaque condition expérimentale. Pour chaque essai, les objets en sample et en target étaient identiques mais apparaissaient soit à un niveau de flou faible où l'objet était assez net (condition de haute fiabilité), soit à un niveau de flou élevé rendant l'identité de l'objet ambiguë (condition de faible fiabilité). Le contexte de scène pouvait être soit congruent avec l'objet (prédictions correctes/valides), soit incongruent avec l'objet (prédictions incorrectes/non valides). De plus, les contextes de scènes en target et en sample pouvaient être identiques ou différents. B. Illustration d'un essai avec une target incongruente et un sample congruent (type de paire target-sample différent). Deux objets flous apparaissaient simultanément, l'objet target à gauche et l'objet sample à droite. Les participants devaient ajuster le niveau de flou de l'objet en sample jusqu'à ce qu'ils jugent les deux objets comme aussi flous l'un que l'autre.

Les photographies originales des scènes ont d'abord été recadrées au format carré et redimensionnées à 700×700 pixels. Elles ont été converties en images en niveaux de gris (256 niveaux de gris) en faisant la moyenne des valeurs des trois canaux de couleur pour chaque pixel. La luminance moyenne et le contraste RMS de toutes les images ont été égalisés (pour les valeurs d'intensité des pixels entre 0 et 1) à 0.5 et la valeur moyenne du contraste RMS de l'ensemble des images (0.1947) a été appliquée à chaque image. Pour chaque image de scène, nous avons ensuite construit un disque de 700×700 pixels contenant la scène².

Les photographies originales des objets ont également été recadrées au format carré et redimensionnées à 120×120 pixels, ce qui correspondait à la taille maximale des objets (soit en largeur, soit en hauteur). Les objets ont été isolés de leur arrière-plan en délimitant leur contour sur le logiciel GIMP (The GIMP Development Team, 2019) et en appliquant des pixels transparents à leur arrière-plan. Ils ont ensuite été convertis en images en niveaux de gris. La luminance moyenne et le contraste RMS de toutes les images ont également été égalisés (pour les valeurs d'intensité des pixels entre 0 et 1) à 0.5 et la valeur moyenne du contraste RMS de l'ensemble des images (0.2117) a été appliquée à chaque image.

Nous avons ensuite flouté l'image de l'objet à l'aide de filtres gaussiens 2D (fonction *imgaussfilt* dans Matlab) de taille croissante (i.e., avec un écart-type croissant - SD) pour obtenir 22 versions de l'objet allant d'un objet net (proche de l'état intact ; niveau de flou 1, SD = 1) à un objet très flou et difficilement reconnaissable (niveau de flou 22, SD = 10). Nous avons

² Ce format a été choisi car nous avons au départ pour projet d'étudier également l'influence des informations de vision périphérique sur la perception des informations de vision centrale avec ces stimuli. Par conséquent, nous nous sommes basées sur le format des stimuli de Trouilloud et al. (2020) prévus pour étudier spécifiquement le rôle des informations de vision périphérique.

Chapitre 2 – Contribution empirique

ensuite assemblé les images des objets et des scènes pour construire les stimuli finaux contenant un objet central entouré d'un contexte de scène. Pour cela, nous avons créé un disque central gris de 250×250 pixels de luminance 0.5 qui a été filtré spatialement par une gaussienne avec un écart-type de 2 afin de lisser le bord du disque. Nous avons ensuite placé ce disque central sur l'image de la scène et nous avons ajouté l'objet flou au centre du disque gris. Ce disque gris nous a permis d'isoler l'objet de l'arrière-plan et d'éviter les différences de contraste locales entre l'objet et la scène dans les différentes conditions.

Les images de l'objet et de la scène ont été appariées afin que chaque objet apparaisse soit entouré d'une scène sémantiquement congruente (e.g., un bateau dans un port), condition expérimentale qui devrait conduire à des prédictions correctes/valides sur l'objet, soit entouré d'une scène sémantiquement incongruente (e.g., un bateau dans une cuisine), condition expérimentale qui devait conduire à des prédictions incorrectes/non valides sur l'objet (Figure 2.14).

2.3.1.3. Procédure

L'expérience a été construite à l'aide de la Psychtoolbox (Brainard, 1997; Pelli, 1997) de MATLAB R2019b (MathWorks, Natick, MA, USA). Les stimuli apparaissaient sur un écran de 30 pouces (DELL ULTRASHARP) avec une résolution de 2560×1600 pixels. Les participants ont placé leur tête sur une mentonnière à 72 cm de l'écran de sorte que le disque gris contenant l'objet corresponde à une taille angulaire de 5° d'angle visuel, que l'objet corresponde à une taille angulaire maximale de 2.4° d'angle visuel et que la scène circulaire corresponde à une taille angulaire de 14° d'angle visuel.

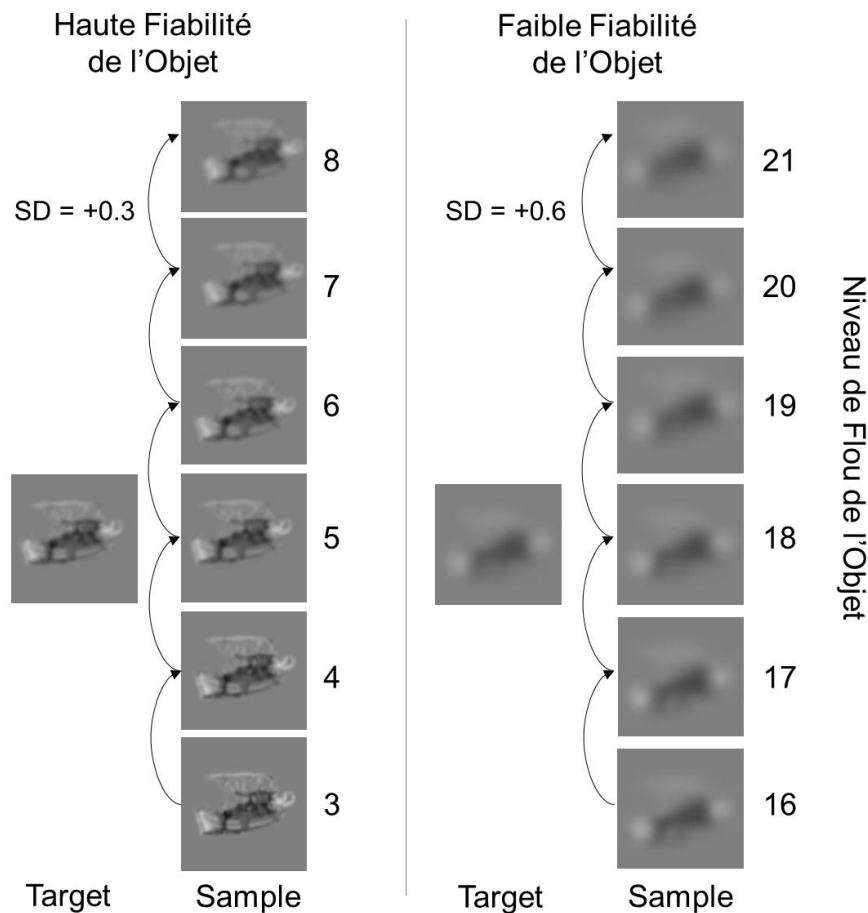
À chaque essai, deux images contenant le même objet central étaient affichées à gauche (target) et à droite (sample) de l'écran sur un fond gris (luminance moyenne de 0.5). Pour étudier l'influence de la fiabilité du signal visuel sur la netteté perçue des objets, l'objet en target apparaissait dans la première moitié des essais à un niveau de flou de 18 (SD des filtres gaussiens = 7.6, faible fiabilité du signal) et dans la seconde moitié des essais à un niveau de flou de 5 (SD des filtres gaussiens = 2.2, haute fiabilité du signal). Dans une étude pilote qui avait pour objectif de déterminer si le seuil de discrimination du flou reste constant quel que soit le niveau de flou ou s'il varie à mesure que le niveau de flou des stimuli augmente (voir Annexe B), nous avons constaté que la différence minimale de flou nécessaire pour percevoir une différence de flou entre deux objets était de 0.3 SD lorsque les objets avaient un niveau de

Chapitre 2 – Contribution empirique

flou inférieur à 4.6 SD (comme dans la condition de haute fiabilité du signal) et de 0.6 SD lorsque les objets avaient un niveau de flou supérieur ou égal à 4.6 SD (comme dans la condition de faible fiabilité du signal). La taille des filtres gaussiens utilisés pour obtenir les niveaux de flou des objets en target et en sample variait donc par incréments de 0.3 SD dans la condition de haute fiabilité du signal et de 0.6 SD dans la condition de faible fiabilité du signal. Nous avons ainsi obtenu un total de 22 versions floues de chaque objet. Nous n'avons pas contrebalancé l'ordre des conditions de fiabilité afin d'éviter que la perception préalable d'objets nets dans la condition de haute fiabilité puisse être utilisée ultérieurement pour désambiguïser le signal d'objets flous dans la condition de faible fiabilité, si elle était effectuée ensuite. Les essais de la condition de faible fiabilité étaient donc présentés en premier et suivis des essais de la condition de haute fiabilité. En ce qui concerne le niveau de flou initial de l'objet en sample, il variait de façon aléatoire entre plus ou moins trois niveaux de flou autour du niveau de flou de l'objet en target. Cependant, les niveaux variaient en fonction de la condition de fiabilité du signal de la target (faible ou haute). En effet, il est possible que dans la condition de faible fiabilité de l'objet en target (i.e., lorsque l'objet est très flou et donc ambigu), le fait d'avoir le même objet en sample à un niveau de flou moyen et donc mieux identifiable, biaiserait le traitement de l'objet en target qui deviendrait alors moins ambigu, même si très flou. Pour éviter cela, le niveau de flou initial de l'objet en sample restait proche de celui de l'objet en target dans la condition de faible fiabilité. Par conséquent, l'objet en sample pouvait avoir un niveau de flou compris entre 16 et 21 dans la condition de faible fiabilité du signal et un niveau de flou compris entre 3 et 8 dans la condition de haute fiabilité du signal (Figure 2.15). Les participants devaient ajuster le niveau de flou de l'objet en sample pour qu'il corresponde au niveau de flou de l'objet en target en augmentant ou en diminuant le flou de l'objet en sample, en utilisant respectivement les flèches gauche et droite du clavier (Figure 2.14). Dans la condition de faible fiabilité, l'objet en sample pouvait être ajusté à un flou maximal correspondant au niveau 22 et à un flou minimal correspondant au niveau 15. Dans la condition de haute fiabilité, l'objet en sample pouvait être ajusté à un flou maximal correspondant au niveau 14 et un flou minimal correspondant au niveau 1. Les participants devaient appuyer sur la barre espace pour valider leur réponse lorsqu'ils jugeaient que les objets en sample et en target étaient aussi flous l'un que l'autre. Les participants étaient encouragés à se fier à leur jugement subjectif pour faire correspondre le niveau de flou des objets et n'étaient pas limités en temps à chaque essai.

Figure 2.15

Illustration du Niveau de Flou des Stimuli dans les Différentes Conditions de Fiabilité.



Note. En condition de haute fiabilité, les stimuli en target apparaissaient à un niveau de flou de 5 et les stimuli en sample apparaissaient à un niveau de flou initial compris entre 3 et 8 inclus. En condition de faible fiabilité, les stimuli en target apparaissaient à un niveau de flou de 18 et les stimuli en sample apparaissaient à un niveau de flou initial compris entre 16 et 21 inclus. Une étude pilote a mis en évidence que la différence minimale de flou nécessaire pour percevoir une différence de flou entre deux objets était de 0.3 SD lorsque les objets avaient un niveau de flou inférieur à 4.6 SD (comme dans la condition de haute fiabilité du signal) et de 0.6 SD lorsque les objets avaient un niveau de flou supérieur ou égal à 4.6 SD (comme dans la condition de faible fiabilité du signal). La taille des filtres gaussiens utilisés pour obtenir les niveaux de flou des objets en target et en sample variait donc par incréments de 0.3 SD dans la condition de haute fiabilité du signal et de 0.6 SD dans la condition de faible fiabilité du signal.

Chapitre 2 – Contribution empirique

Pour chaque essai, les objets en target et en sample avaient soit un contexte de scène différent (l'un congruent et l'autre incongruent avec l'objet, i.e., des types de paires target-sample différents), soit le même contexte de scène (congruent ou incongruent avec l'objet, i.e., des types de paires target-sample identiques). Les types de paires target-sample différents nous ont permis de tester notre hypothèse principale, à savoir que les objets congruents avec leur contexte (prédictions valides) devraient être perçus comme plus nets que les mêmes objets dans un contexte incongruent (prédictions non valides) lorsque le signal visuel est peu fiable tandis que l'inverse devrait être observé lorsque le signal visuel est fiable. Similairement aux Expériences 1 et 2, des paires dans lesquelles les stimuli en target et en sample étaient identiques ont également été incluses pour évaluer si les effets observés dans la condition de paires target-sample différentes étaient bien de nature perceptive et non la conséquence d'un biais de réponse (Figure 2.14).

Une session expérimentale contenait 160 essais divisés en deux blocs de 80 essais : les 80 essais où l'objet était peu fiable apparaissaient toujours en premier et étaient suivis des 80 essais où l'objet était très fiable. Dans chaque bloc, nous avons 20 objets \times 2 conditions de congruence du contexte sémantique pour l'objet en sample (contexte congruent vs. incongruent) \times 2 types de paires target-sample (types de paires target-sample identiques vs. types de paires target-sample différentes). L'expérience durait environ 30-40 minutes, selon la vitesse à laquelle les participants ajustaient le flou de l'objet en sample, et comportait une pause tous les 40 essais. Pour chaque essai, nous avons enregistré le niveau de flou attribué à l'objet sample pour correspondre à celui de l'objet target (Matched Blur Level, MBL).

2.3.2. Analyse des données

Les données ont été analysées à l'aide du logiciel R (R Core Team, 2019) et du package *lme4* (Bates, Mächler, et al., 2015) pour effectuer une analyse linéaire à effets mixtes de l'effet de la congruence entre objet-contexte en sample (congruents vs. incongruents), de la fiabilité du signal de l'objet (haute vs. faible) et du type de paire target-sample (identique vs. différente) dans un design intra-sujet $2 \times 2 \times 2$. La congruence objet-contexte en sample, la fiabilité du signal de l'objet, le type de paire target-sample, et leurs termes d'interaction ont été introduits dans le modèle comme effets fixes. Les intercepts pour les sujets, ainsi que les pentes aléatoires par sujet pour l'effet de la congruence objet-contexte en sample, de la fiabilité du

Chapitre 2 – Contribution empirique

signal de l'objet, du type de paire target-sample et de leurs interactions ont été spécifiés comme effets aléatoires.

Nous avons effectué le nombre minimum d'analyses nécessaires pour tester nos hypothèses afin de ne pas augmenter le risque d'erreur de type 1. D'après l'opposing process theory (Press et al., 2020), le mécanisme de sharpening (perception plus nette des objets attendus par rapport aux objets inattendus) devrait s'appliquer lorsque les stimuli sont ambigus et que le signal visuel n'est pas fiable, tandis que le mécanisme de dampening (perception moins nette des objets attendus par rapport aux objets inattendus) devrait s'appliquer lorsque le signal est plus fiable. Nous nous attendions à ce que dans la condition de faible fiabilité de l'objet, les objets congruents soient subjectivement perçus comme plus nets que les objets incongruents à un niveau de flou objectivement égal entre les objets target et sample. Par conséquent, pour les paires target-sample (T/S) de contextes différents en condition de faible fiabilité, le MBL des objets sample dans la condition T-incongruent/S-congruent devrait être plus élevé que le MBL des objets en sample dans la condition T-congruent/S-incongruent. D'autre part, nous nous attendions à ce que dans la condition de haute fiabilité de l'objet, les objets incongruents soient subjectivement perçus comme plus nets que les objets congruents à un niveau de flou objectivement égal. Par conséquent, pour les paires target-sample de contextes différents en condition de haute fiabilité, le MBL des objets en sample dans la condition T-congruent/S-incongruent devrait être plus élevé que le MBL des objets en sample dans la condition T-incongruent/S-congruent. Pour répondre à cette hypothèse principale, nous avons donc testé l'interaction entre la congruence objet-contexte en sample et la fiabilité du signal pour les types de paires target-sample différents, ainsi que la différence entre la MBL des objets sample congruents et incongruents dans chaque condition de fiabilité du signal (i.e., deux tests a priori non orthogonaux).

De plus, si les prédictions basées sur des informations contextuelles congruentes ou incongruentes influencent la netteté perçue des objets (soit par le mécanisme de sharpening ou de dampening), de tels effets perceptifs devraient également se produire lors de l'appariement de paires target-sample identiques (i.e., deux stimuli congruents ou incongruents). Cependant, ils devraient affecter de la même manière les objets en target et en sample (Firestone & Scholl, 2014). Par conséquent, la différence de flou perçue entre les objets en target et en sample dans les types de paires identiques (et donc la différence de MBL entre les conditions T-congruent/S-congruent et T-incongruent/S-incongruent) dans les deux conditions de fiabilité des objets devrait être réduite (voire absente) par rapport aux essais avec des types de paires différents.

Chapitre 2 – Contribution empirique

Pour répondre à cette hypothèse, nous avons donc également testé l'interaction double entre la congruence objet-contexte du stimulus sample, la fiabilité de l'objet et le type de paire target-sample.

Nous avons utilisé la méthode proposée par Bates, Kliegl, et al. (2015) pour construire des modèles mixtes parcimonieux limitant les problèmes de convergence. L'inspection visuelle des graphiques de résidus n'a pas révélé de déviations évidentes de l'homoscédasticité ou de la normalité. Les p-valeurs ont été obtenues par approximation de Satterthwaite avec le package *lmerTest* (Kuznetsova et al., 2017). Le seuil de significativité a été fixé à .025 ($.05/2 = .025$) après correction de Bonferroni pour les deux tests non orthogonaux réalisés. Les tailles d'effet ont été estimées à l'aide du d_z de Cohen (Lakens, 2013).

2.3.3. Résultats

Les MBL moyens pour chaque condition expérimentale sont présentés dans la Figure 2.16. Étant donné que le niveau de flou de l'objet en target peut être de 5 dans la condition de haute fiabilité ou de 18 dans la condition de faible fiabilité, un MBL supérieur (inférieur) à ces valeurs dans chaque condition de fiabilité du signal indiquerait une surestimation (sous-estimation) du flou de l'objet en sample par rapport à celui de l'objet en target.

Les résultats ont d'abord révélé un effet principal de la fiabilité du signal, $t(51.99) = 375.66$, $p < .001$, $d_z = 51.601$, indiquant que le MBL était plus élevé dans la condition de faible fiabilité ($M = 17.93$, $SE = 0.13$) que dans la condition de haute fiabilité ($M = 5.10$, $SE = 0.10$). Cet effet pouvait être attendu étant donné que le niveau de flou de l'objet en target était plus élevé dans la condition de faible fiabilité que dans la condition de haute fiabilité. Il n'y avait pas d'effet principal significatif de la congruence du sample, $t(8368) = 1.26$, $p = .208$, $d_z = 0.173$, ni d'effet principal significatif du type de paire target-sample, $t(8368) = 1.13$, $p = .259$, $d_z = 0.155$.

En ce qui concerne notre hypothèse principale, les résultats ont révélé une interaction significative entre la congruence objet-contexte du stimulus sample et la fiabilité du signal pour les types de paires target-sample différents, $t(4132) = 3.27$, $p = .001$, $d_z = 0.449$. Plus précisément, dans la condition de faible fiabilité pour les types de paires target-sample différents, nous avons observé un effet significatif de la congruence objet-contexte du stimulus sample, $t(2066) = 2.54$, $p = .011$, $d_z = 0.349$, pour lequel les participants ont attribué un niveau de flou plus élevé aux objets sample avec un contexte congruent qu'incongruent (sample

Chapitre 2 – Contribution empirique

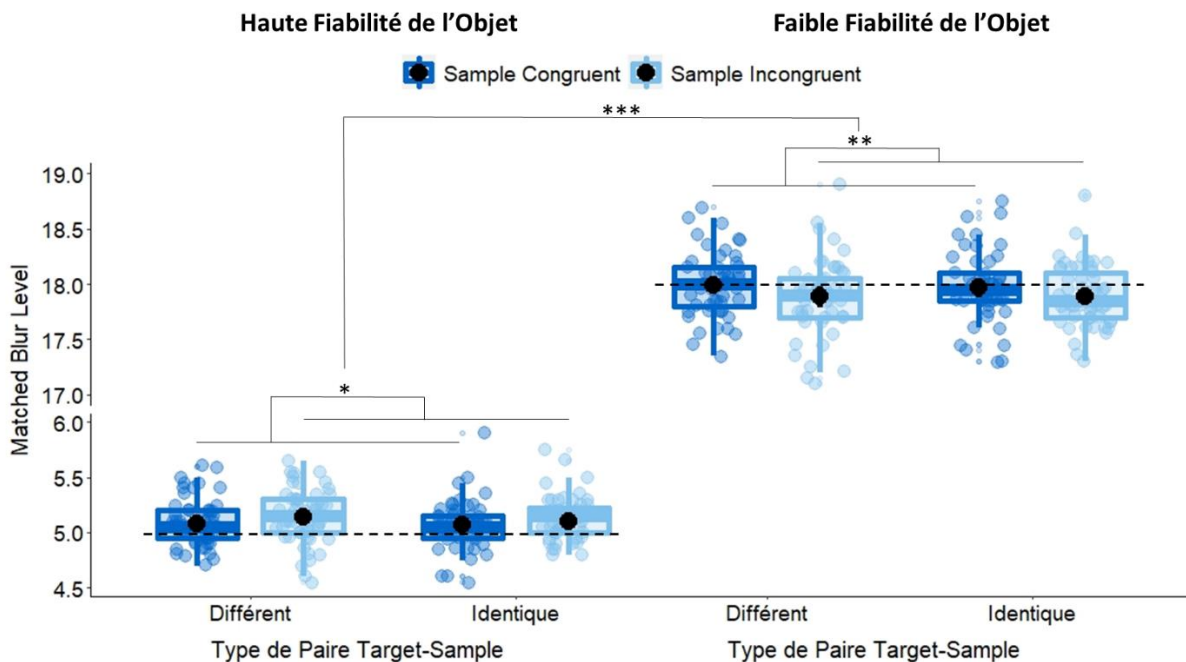
congruent : $M = 17.99$, $SE = 0.14$; sample incongruent : $M = 17.89$, $SE = 0.14$). Au contraire, dans la condition de haute fiabilité pour les types de paires target-sample différentes, les participants ont eu tendance à attribuer un niveau de flou plus élevé aux objets sample en contexte incongruent que congruents, mais cette différence n'était pas significative après correction de Bonferroni, $t(2066) = -2.07$, $p = .039$, $d_z = 0.284$ (sample congruent : $M = 5.08$, $SE = 0.10$; sample incongruent : $M = 5.14$, $SE = 0.10$).

En ce qui concerne notre deuxième hypothèse, l'interaction double entre la congruence objet-contexte du stimulus sample, le type de paire target-sample et la fiabilité du signal n'était pas significative, $t(8368) = 0.74$, $p = .462$, $d_z = 0.102$, indiquant que l'interaction entre la congruence objet-contexte du stimulus sample et la fiabilité du signal ne différait pas significativement dans les paires identiques par rapport aux paires target-sample différentes. Ce pattern de résultats suggère donc plutôt une interaction entre la congruence objet-contexte du stimulus sample et la fiabilité du signal, indépendamment du type de paire target-sample, qui était effectivement significative, $t(8368) = 4.02$, $p < .001$, $d_z = 0.552$. Nous avons testé l'effet de la congruence objet-contexte du stimulus sample a posteriori pour chacune des conditions de fiabilité du signal (indépendamment du type de paire target-sample). Dans la condition de faible fiabilité, nous avons trouvé un effet significatif de la congruence objet-contexte du stimulus sample, $t(4186) = 3.26$, $p = .001$, $d_z = 0.448$, où les participants ont attribué un niveau de flou plus élevé aux objets sample en contexte congruent qu'incongruents (sample congruent : $M = 17.98$, $SE = 0.13$; sample incongruent : $M = 17.89$, $SE = 0.13$), ce qui suggère une perception plus nette des objets congruents qu'incongruents, indépendamment du type de paire target-sample. Au contraire, dans la condition de haute fiabilité, les participants ont attribué un niveau de flou plus élevé aux objets sample en contexte incongruent que congruent et cette fois la différence était significative, $t(4186) = -2.34$, $p = .019$, $d_z = 0.321$ (sample congruent : $M = 5.07$, $SE = 0.10$; sample incongruent : $M = 5.12$, $SE = 0.10$), ce qui suggère une perception plus nette des objets incongruents que congruents, indépendamment du type de paire target-sample).

Il n'y avait aucune autre interaction significative (interaction entre le type de paire target-sample et la fiabilité du signal : $t(8368) = -0.47$, $p = .637$, $d_z = 0.065$, interaction entre le type de paire target-sample et la congruence objet-contexte du stimulus sample : $t(8368) = -0.08$, $p = .937$, $d_z = 0.036$).

Figure 2.16

Matched Blur Level Moyen pour Chaque Condition de l'Expérience 5.



Note. Niveau de flou moyen attribué aux objets sample (MBL) congruents et incongruents pour les paires target-sample différentes et identiques en condition de haute (gauche) et de faible (droite) fiabilité. Un niveau de flou élevé correspond à un objet très flou, tandis qu'un niveau de flou faible correspond à un objet relativement net. Les points noirs et les barres d'erreur indiquent respectivement la moyenne et l'erreur standard de tous les participants. Les points en couleur correspondent aux observations individuelles. Les lignes en pointillés correspondent à l'ajustement parfait pour chaque condition.

* $p < .05$. ** $p < .01$. *** $p < .001$

Par souci d'exhaustivité, nous avons répété les mêmes analyses statistiques, en ajoutant cette fois les effets aléatoires des stimuli. Ainsi, les intercepts des stimuli, ainsi que les pentes aléatoires par stimuli pour l'effet de la congruence objet-contexte du stimulus sample, de la fiabilité du signal, du type de paire target-sample et de leurs interactions ont été spécifiés comme effets aléatoires dans notre modèle à effets mixtes. Une fois encore, l'inspection visuelle des graphiques de résidus n'a pas révélé de déviations évidentes de l'homoscédasticité ou de la normalité.

Après inclusion des effets aléatoires des stimuli, les résultats n'ont révélé aucune interaction significative entre la congruence objet-contexte du stimulus sample et la fiabilité du

Chapitre 2 – Contribution empirique

signal pour les types de paires target-sample différents, $t(18.99) = 1.70$, $p = .106$, $d_z = 0.234$. De plus, l'interaction double entre la fiabilité du signal, le type de paire target-sample et la congruence objet-contexte du stimulus sample n'était pas significative, $t(19.02) = 0.34$, $p = .738$, $d_z = 0.047$ et il n'y avait pas non plus d'interaction entre la congruence objet-contexte du sample et la fiabilité du signal, $t(19.26) = 1.74$, $p = .098$, $d_z = 0.239$.

2.3.4. Discussion de l'Expérience 5

L'objectif de cette cinquième expérience était d'étudier comment la validité des prédictions, basées sur des informations contextuelles, influence la perception subjective des objets et ce, en tenant compte de la fiabilité du signal visuel. À cette fin, les participants ont effectué une tâche d'ajustement perceptif similaire à celle des expériences précédentes consistant à ajuster le niveau de flou d'un objet (sample) intégré dans un contexte de scène congruent ou incongruent (conduisant à des prédictions valides ou invalides sur l'objet, respectivement) pour faire correspondre le niveau de flou du même objet (target) intégré dans un contexte congruent ou incongruent. Pour tenir compte de l'effet de la fiabilité du signal, les objets pouvaient apparaître à un niveau de flou faible (i.e., haute fiabilité du signal) ou élevé (i.e., faible fiabilité du signal). D'après l'opposing process theory (Press et al., 2020) qui postule que les stimuli attendus sont perceptivement réhaussés par rapport aux stimuli inattendus (mécanisme de sharpening) lorsque le signal visuel est bruité, tandis qu'ils sont perceptivement atténués (mécanisme de dampening) lorsque le signal visuel est plus fiable, nous nous attendions à ce qu'à un niveau de flou objectivement égal, les objets congruents soient subjectivement perçus comme plus nets que les objets incongruents lorsque le niveau de flou est élevé et que les objets incongruents soient perçus comme plus nets que les objets congruents lorsque le niveau de flou est faible. Cependant, nous nous attendions à ce que la différence de flou perçue entre les objets en target et en sample dans les deux conditions de fiabilité du signal soit réduite lorsque les paires d'objets avaient un contexte identique plutôt que différent (Firestone & Scholl, 2014).

Contrairement à cette hypothèse, les résultats n'ont révélé aucune différence significative entre les effets observés dans les conditions de paires différentes et identiques. Cela suggère que même lorsqu'ils étaient identiques, les objets en target et en sample étaient cependant perçus comme différents. L'interaction significative entre la congruence objet-contexte du stimulus sample et la fiabilité du signal, ainsi que les tests post-hoc, suggèrent

Chapitre 2 – Contribution empirique

toutefois que l'effet de la congruence du sample s'inverse en fonction de la fiabilité du signal (i.e., les objets en sample congruents sont perçus comme plus nets que les objets incongruents lorsque la fiabilité du signal est faible, tandis que le contraire est observé lorsque la fiabilité du signal est élevée) et ce, indépendamment du type de paire target-sample.

Suivant le raisonnement de Firestone et Scholl (2014), si des effets perceptifs peuvent être trouvés lorsque les stimuli sont identiques, ces effets doivent plutôt refléter l'influence de jugements non perceptifs ou de biais de réponse. Dans le contexte de notre expérience, les objets qui ont plus de sens dans des contextes congruents peuvent par exemple être plus facilement jugés comme étant nets ou clairs que les objets dans des contextes incongruents sans que cela soit associé à des effets perceptifs. De tels jugements non perceptifs pourraient donc biaiser l'ajustement perceptif du flou des stimuli. Cependant, si nos effets étaient dus à un biais de réponse systématique, consistant par exemple à juger les objets congruents plus nets que les objets incongruents (et donc à attribuer systématiquement plus de flou aux objets congruents qu'aux objets incongruents en sample), un tel biais devrait être présent dans les deux conditions de fiabilité. Or, nous avons observé que l'effet de la congruence objet-contexte était inversé en fonction de la fiabilité du signal, et ce indépendamment du type de paire target-sample (i.e., interaction croisée entre la congruence du sample et la fiabilité du signal), ce qui rend peu probable l'explication de nos résultats par un biais de réponse.

Si les résultats dans les types de paires target-sample identiques ne peuvent pas être attribués à un biais de réponse, comment expliquer que le pattern de résultats dans cette condition soit similaire à celui de la condition des types de paires différents ? Les résultats d'une étude récente de Dubova et Goldstone (2022) peuvent fournir une piste intéressante pour répondre à cette question. Ces auteurs ont remis en question l'hypothèse selon laquelle les effets perceptifs induits par les connaissances préalables devraient affecter de la même manière tous les éléments du champ visuel en faisant valoir que lors de tâches d'ajustement perceptif comme les nôtres, les stimuli en target et en sample (même s'ils sont identiques) peuvent difficilement être perçus de manière similaire au même moment. Dans leur étude, ces auteurs ont enregistré les mouvements oculaires des participants au cours d'une tâche d'ajustement perceptif (de couleurs) comme la nôtre et ont observé que les participants fixaient plus souvent le stimulus sample en cours d'ajustement que le stimulus target, qui apparaissait donc le plus souvent de manière périphérique tout au long de la tâche. Dans le contexte théorique de la présente étude, un tel comportement pourrait avoir un impact critique sur la performance d'ajustement et pourrait expliquer les différences observées même dans les essais où les paires target-sample

Chapitre 2 – Contribution empirique

étaient identiques. En effet, dans cette Expérience 5, nos objets en target et en sample avaient une taille relativement petite (2.4° d'angle visuel) et ils étaient par contre chacun entourés d'un large contexte (contrairement aux Expériences 1 à 4 où les objets étaient plus gros et les scènes plus petites, donc vus à une moindre excentricité). Lorsque les participants fixaient l'un des objets, l'autre apparaissait avec une excentricité de 16.3° et le bord intérieur de son contexte était à 14.3° . Si les participants ont adopté une stratégie similaire à celle de Dubova et Goldstone consistant à fixer l'objet en sample la plupart du temps, cela implique que l'objet en target a été perçu avec une résolution spatiale plus faible inhérente à la vision périphérique (Wertheim, 1894). Plus préjudiciable encore, le contexte de l'objet en target peut également être apparu flou et donc moins informatif ou plus ambigu en ce qui concerne sa congruence avec l'objet en target. Les effets perceptifs induits par les prédictions contextuelles pourraient donc avoir affecté l'objet en target moins fortement que celui en sample. Notre pattern de résultats suggère en effet que la performance des participants (i.e., le MBL) varie en fonction de la congruence de l'objet en sample mais pas de la congruence de l'objet en target (i.e., du type de paire target-sample). Une autre explication possible de la similitude des résultats obtenus pour les paires target-sample identiques et différentes peut provenir de la conception de notre expérience, dans laquelle le niveau de flou de l'objet en target est resté constant tout au long des blocs de haute fiabilité et de faible fiabilité (niveaux de flou de 5 et 18, respectivement), tandis que le flou de l'objet en sample variait d'un essai à l'autre. Cela a pu conduire les participants à créer une estimation précise du niveau de flou de l'objet en target dans chaque condition, qui est devenue moins sujette aux effets induits par l'information contextuelle par rapport à celle de l'objet en sample, éliminant ainsi l'effet de congruence de l'objet target.

Par ailleurs, l'effet de la congruence objet-contexte du stimulus sample qui s'est inversé en fonction de la fiabilité du signal (indépendamment du type de paire target-sample), est conforme à nos hypothèses. Dans la condition de faible fiabilité, les participants ont attribué un niveau de flou plus élevé aux objets en sample congruents qu'aux objets en sample incongruents. Cela indique qu'à un niveau de flou physiquement égal entre l'objet en sample et en target, les objets attendus apparaissent subjectivement plus nets que les objets inattendus. Ce pattern de résultats était inversé dans la condition de haute fiabilité. Les participants ont attribué un niveau de flou plus élevé aux objets en sample dans un contexte de scène incongruent qu'aux objets en sample dans un contexte de scène congruent, ce qui indique qu'à un niveau de flou physiquement égal entre l'objet en sample et en target, les objets inattendus sont subjectivement perçus comme plus nets que les objets attendus. Ces résultats soutiennent donc l'idée d'effets

Chapitre 2 – Contribution empirique

perceptifs opposés des prédictions en fonction de la fiabilité du signal, conformément à l’opposing process theory (Press et al., 2020). Il est toutefois intéressant de noter que ces effets perceptifs semblent surtout affecter les stimuli inattendus plutôt que les stimuli attendus. En effet, comme on peut le voir sur la Figure 2.16, le MBL pour les objets en sample congruents était très proche d’une correspondance parfaite, ce qui suggère une estimation précise de leur flou, indépendamment de la fiabilité du signal. Au contraire, le flou des objets en sample incongruents tend à être sous-estimé lorsque la fiabilité du signal est élevée - ce qui suggère que l’objet est perçu comme plus net qu’il ne l’est en réalité -, tandis qu’il est surestimé lorsque la fiabilité du signal est faible - ce qui suggère que l’objet est perçu comme plus flou qu’il ne l’est en réalité. La perception relativement plus nette (plus floue) des stimuli attendus que des stimuli inattendus lorsque la fiabilité du signal est faible (élevée), mise en avant par l’opposing process theory, pourrait donc en fait seulement résulter d’un effet des prédictions sur les stimuli inattendus.

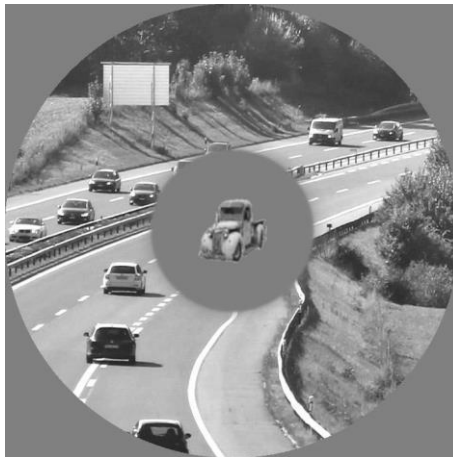
Enfin, il faut noter que les analyses supplémentaires incluant les effets aléatoires des stimuli dans nos modèles n’ont montré aucune interaction significative entre la congruence du sample et la fiabilité du signal. Cela indique que ces résultats, qui étaient significatifs dans les analyses principales n’incluant pas la variabilité des stimuli, ne peuvent pas être généralisés à tous les stimuli visuels mais restent plutôt limités aux stimuli utilisés pour cette expérience. En effet, la variabilité du stimulus incluse dans les analyses supplémentaires masque les effets fixes globaux testés car ces effets fixes n’étaient pas assez importants par rapport à la variabilité des stimuli pour atteindre le seuil de significativité. Nous avons donc mené des analyses complémentaires au niveau des stimuli afin de déterminer si l’absence de résultats significatifs s’explique effectivement par leur variabilité. En effet, il est possible que bien qu’elles soient congruentes sémantiquement, certaines paires objet-contexte congruents ne soient pas plausibles physiquement. Par exemple, dans la Figure 2.17 ci-dessous, on peut voir l’objet dépanneuse qui est placé sur une autoroute pour la condition de contexte congruent, mais on peut remarquer que la dépanneuse n’est pas dans le bon sens de circulation et que sa taille n’est pas physiquement plausible dans le contexte de cette scène particulière. Il est donc possible que les participants aient considéré cette paire dépanneuse-autoroute comme un stimulus incongruent au lieu de congruent. Nous avons donc mis en place une expérience contrôle (voir Annexe C) où nous avons demandé à des participants n’ayant jamais vu les stimuli de juger leur congruence sémantique (« À quel point la présence de cet objet paraît-elle plausible à vos yeux dans ce contexte ? »), mais également physique (« À quel point la taille et la position de

Chapitre 2 – Contribution empirique

l'objet sont-elles plausibles dans ce contexte »). Après avoir reclassé l'ensemble de nos stimuli comme congruents et incongruents selon les réponses des participants, nous avons renouvelé les analyses statistiques en modèles mixtes incluant la variabilité des stimuli. Malgré cela, les résultats n'ont pas changé car aucun effet n'a atteint le seuil de significativité (voir Annexe C). Toutefois, le fait que nous ne puissions pas généraliser nos résultats à tous les stimuli n'est pas nécessairement problématique puisque nos stimuli n'étaient pas très écologiques et ils étaient spécifiquement construits pour répondre à nos hypothèses. Des pistes pour pallier ce problème seront développées dans la Discussion Générale.

Figure 2.17

Exemple de Stimulus de l'Expérience 5.



Note. L'objet dépanneuse placé dans un contexte d'autoroute pour la condition de contexte congruent n'est pas dans le bon sens de circulation et sa taille n'est pas physiquement plausible dans le contexte de cette scène particulière.

2.4. Expérience 6 – Etude de l'effet des prédictions en fonction de la validité des prédictions et de contraintes temporelles sur le traitement du signal visuel

L'objectif de l'Expérience 6 était de tester le second postulat de l'opposing process theory (Press et al., 2020), à savoir que le mécanisme de sharpening interviendrait en premier, permettant de confirmer les prédictions, puis le mécanisme de dampening prendrait ensuite le relais afin de privilégier le traitement des autres caractéristiques inattendues. Ainsi, l'influence de la validité des prédictions sur la perception subjective des objets devrait s'inverser avec le

Chapitre 2 – Contribution empirique

décours temporel du traitement des stimuli visuels, avec une préférence temporelle du sharpening sur le dampening. Un moyen de tester cette hypothèse est de manipuler le temps de présentation des stimuli visuels. Le paradigme d'ajustement perceptif utilisé jusque-là n'étant pas propice à la manipulation du temps de présentation des images, nous avons donc mis en place un nouveau paradigme inspiré des travaux de Spaak et al. (2022). Dans ce paradigme, nous avons utilisé le même type de stimuli que dans l'Expérience 5, c'est à dire des scènes avec un objet congruent ou incongruent en leur centre, mais en les présentant individuellement. Afin d'étudier spécifiquement l'influence des contraintes temporelles, nous avons manipulé le temps de présentation des stimuli. La scène était suivie de la présentation de deux objets côte à côte, l'un correspondant à l'objet présent dans la scène (objet cible), l'autre correspondant à un autre exemplaire du même objet (objet distracteur). Les participants avaient pour tâche de discriminer parmi les deux objets celui qui avait été vu dans la scène. Cette tâche de discrimination parmi deux exemplaires du même objet a été privilégiée à une tâche plus classique d'identification simple ou de catégorisation de l'objet présent dans la scène. En effet, si nous demandions d'identifier/de catégoriser directement l'objet présent dans la scène, la tâche pourrait être réalisée sur la seule base du contexte, plutôt que sur la base de l'objet. En utilisant une tâche de discrimination entre deux objets à la suite de la présentation de la scène, notre objectif était d'évaluer plus spécifiquement la qualité de l'encodage perceptif des objets en limitant le recours à des stratégies autres que perceptives (e.g., sémantiques basées sur le contexte) pour résoudre la tâche. Enfin, les travaux antérieurs (Press et al., 2010; Yon & Press, 2017) mettent en évidence un mécanisme de sharpening entre 50 à 110 ms après la stimulation visuelle et un mécanisme de dampening entre 200 et 290 ms après la stimulation. Nous avons donc présenté les objets, intégrés dans un contexte de scène congruent ou incongruent, soit pendant une durée très brève de 50 ms, condition expérimentale qui devrait impliquer le mécanisme de sharpening, soit pendant une durée plus longue de 200 ms, condition expérimentale qui devrait impliquer le mécanisme de dampening.

En se basant sur les hypothèses formulées par l'opposing process theory (Press et al., 2020), nous nous attendions donc à une interaction entre le temps de présentation des stimuli et la congruence objet-contexte. Pour un temps de présentation de très bref de 50 ms, les objets congruents avec le contexte devraient être plus rapidement et mieux identifiés que les objets incongruents avec le contexte. En revanche, pour un temps de présentation plus long de 200 ms, les objets incongruents avec le contexte devraient être plus rapidement et mieux identifiés que les objets congruents avec le contexte.

Chapitre 2 – Contribution empirique

2.4.1. Méthode

2.4.1.1. Participants et Design

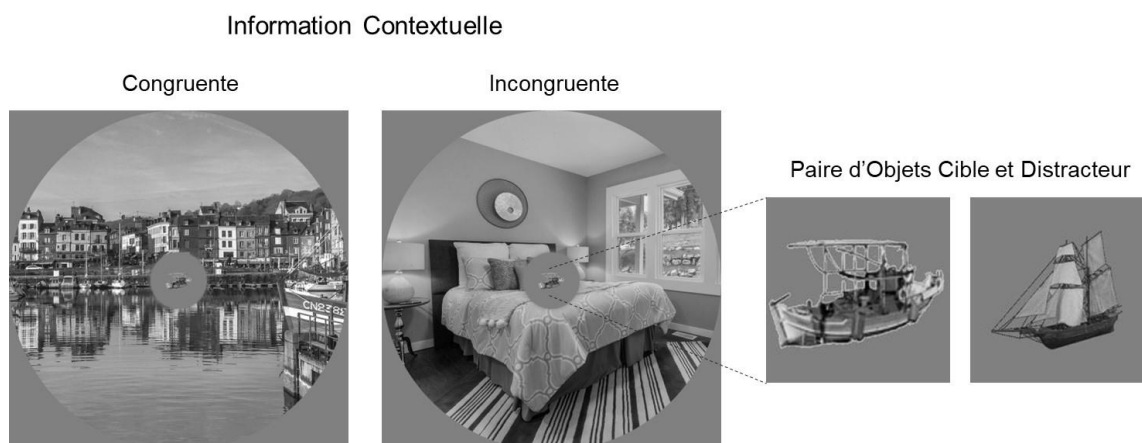
Quatre-vingt-treize étudiants en licence de psychologie de l'Université Grenoble Alpes (87 femmes, $M_{age} = 21.06$, $SD_{age} = 3.94$) ayant une vision normale ou corrigée à la normale ont participé à l'expérience. La taille de l'échantillon a été fixée sur la base d'une taille d'effet estimée de $d_z = 0.30$ à partir de l'effet de la congruence sur les temps de réponse en condition *probe-key* dans l'Expérience 2 de Spaak et al. (2022) pour obtenir une puissance (i.e., correspondant à $1 - \beta$, β étant le risque d'erreur de type 2) de 0.8 avec un seuil alpha de .05.

2.4.1.2. Stimuli

Les stimuli étaient constitués de 20 photographies en couleurs de scènes (10 scènes d'intérieur et 10 scènes d'extérieur) et de 40 photographies en couleurs d'objets (20 objets animés et 20 objets non animés). Les photographies ont été téléchargées sur le site Pixabay (<https://pixabay.com/fr/>) et sur la base de données Google Images en filtrant pour la licence CC0. Les stimuli ont été créés à l'aide de MATLAB R2019b (MathWorks, Natick, MA, USA).

Figure 2.18

Exemple de Paires d'Objets Présentées en Conditions Congruente et Incongruente dans l'Expérience 6.



Note. Chaque objet cible apparaissait soit entouré d'un contexte congruent sémantiquement, soit entouré d'un contexte incongruent sémantiquement. Chaque objet cible était apparié à un objet distracteur sémantiquement et physiquement similaire afin que les participants identifient quel objet de la paire leur avait été présenté dans la scène.

Chapitre 2 – Contribution empirique

Les photographies originales des scènes ont d'abord été recadrées au format carré et redimensionnées à 880×880 pixels. Elles ont été converties en images en niveaux de gris (256 niveaux de gris) en faisant la moyenne des valeurs des trois canaux de couleur pour chaque pixel. La luminance moyenne et le contraste RMS de toutes les images ont été égalisés (pour les valeurs d'intensité des pixels entre 0 et 1) à 0.5 et la valeur moyenne du contraste RMS de l'ensemble des images (0.1947) a été appliquée à chaque image. Pour chaque image de scène, nous avons ensuite construit un disque de 880×880 pixels contenant la scène. Les photographies originales des objets ont également été recadrées au format carré et redimensionnées dans deux formats : un à 70×70 pixels et l'autre à 200×200 pixels. Les objets ont été isolés de leur arrière-plan en délimitant leur contour sur le logiciel GIMP (The GIMP Development Team, 2019) et en appliquant des pixels transparents à leur arrière-plan. Ils ont ensuite été convertis en images en niveaux de gris en faisant la moyenne des valeurs des trois canaux de couleur pour chaque pixel. La luminance moyenne et le contraste RMS de toutes les images ont également été égalisés (pour les valeurs d'intensité des pixels entre 0 et 1) à 0.5 et la valeur moyenne du contraste RMS de l'ensemble des images (0.2070) a été appliquée à chaque image. Nous avons ensuite assemblé les images des objets et des scènes pour construire les stimuli finaux contenant un objet central entouré d'un contexte de scène. Pour cela, nous avons créé un disque central gris de 140×140 pixels de luminance 0.5 qui a été filtré spatialement par une gaussienne avec un écart-type de 2 afin de lisser le bord du disque. Nous avons ensuite placé ce disque central sur l'image de la scène et nous avons ajouté l'objet au format 70×70 pixels au centre du disque gris. Ce disque gris nous a permis d'isoler l'objet de l'arrière-plan et d'éviter les différences de contraste locales entre l'objet et la scène dans les différentes conditions.

Les 40 images d'objet ont été utilisées pour être superposées aux scènes en tant qu'objets cibles (Figure 2.18). Les images de l'objet et de la scène ont été appariées afin que chaque objet apparaisse soit entouré d'une scène sémantiquement congruente (e.g., un bateau dans un port), condition expérimentale qui devrait conduire à des prédictions correctes/valides sur l'objet, soit entouré d'une scène sémantiquement incongruente (e.g., un bateau dans une chambre), condition expérimentale qui devait conduire à des prédictions incorrectes/non valides sur l'objet (Figure 2.18). Les 40 images d'objets ont également été appariées deux à deux entre elles afin de créer des paires d'objets sémantiquement et physiquement similaires (e.g., deux bateaux ; Figure 2.18) dans lesquelles chaque objet de la paire était tour à tour soit objet cible (superposé à une scène), soit objet distracteur.

2.4.1.3. Procédure

L'expérience a été construite à l'aide de la Psychtoolbox (Brainard, 1997; Pelli, 1997) de MATLAB R2019b (MathWorks, Natick, MA, USA) et a été conduite à l'aide du logiciel Octave 5.1.0.0 (Eaton et al., 2019). Les stimuli apparaissaient sur un écran de 23.8 pouces (DELL P2419H) avec une résolution de 1920×1080 pixels. Les participants posaient leur tête sur une mentonnière placée à 45 cm de l'écran afin que le disque gris contenant l'objet corresponde à une taille angulaire de 5° d'angle visuel. Cela nous a permis de contrôler que l'objet était entièrement perçu en vision centrale avec la meilleure acuité visuelle possible.

Au cours de cette expérience, nous avons manipulé le temps de présentation des objets en contexte. Les travaux antérieurs (Press et al., 2010; Yon & Press, 2017) mettent en évidence un mécanisme de sharpening entre 50 à 110 ms après la stimulation visuelle et un mécanisme de dampening entre 200 et 290 ms après la stimulation. Nous avons donc présenté les objets, intégrés dans un contexte de scène congruent ou incongruent, soit pendant une durée très brève de 50 ms, condition expérimentale qui devrait impliquer le mécanisme de sharpening, soit pendant une durée plus longue de 200 ms, condition expérimentale qui devrait activer le mécanisme de dampening. Un essai se déroulait comme suit (Figure 2.19) : Un point de fixation apparaissait pendant 500 ms au centre de l'écran, suivi du contexte de scène sans l'objet en son centre pendant 150 ms. La présentation du contexte en amont de l'objet avait pour objectif de maximiser l'élaboration de prédictions (Roux-Sibilon et al., 2019). Cette scène restait affichée à l'écran pendant la présentation de l'objet. L'objet était ensuite présenté au centre de la scène pendant 50 ms ou 200 ms selon la condition expérimentale. L'objet et la scène étaient ensuite suivis d'un masque du même temps de présentation (50 ms ou 200 ms selon la condition expérimentale). Le masque était ensuite suivi d'un écran présentant deux objets à droite et à gauche de l'écran. L'un des objets était celui qui avait été présenté précédemment dans la scène (objet cible) et le second objet était un objet apparié sur la base de la similarité sémantique et physique (objet distracteur). Les participants devaient indiquer le plus rapidement et précisément possible lequel des deux objets avait été présenté auparavant dans le contexte de la scène. Les participants répondaient avec les flèches de gauche et de droite du clavier d'ordinateur pour choisir l'objet de gauche ou de droite respectivement. Les deux objets restaient affichés à l'écran jusqu'à la réponse du participant, puis ils étaient suivis d'un écran gris de 1000 ms. Nous avons enregistré les taux de bonnes réponses ainsi que les temps de réponse des participants.

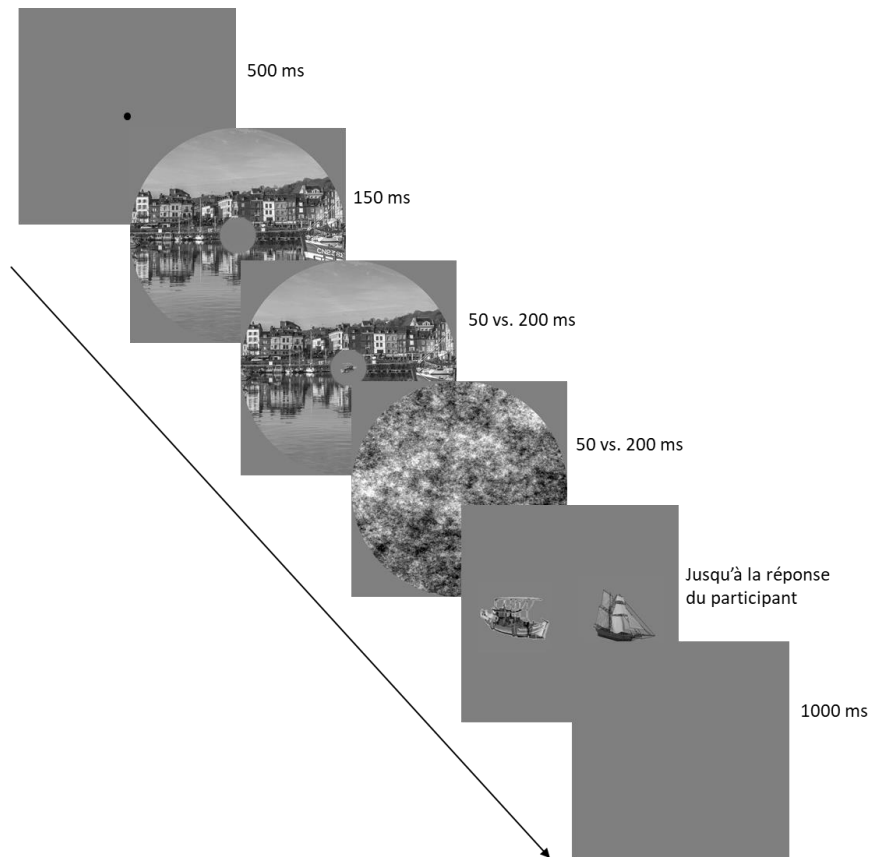
Chapitre 2 – Contribution empirique

Lors de la présentation des deux objets à gauche et à droite de l'écran, les deux objets avaient la même orientation mais l'objet qui avait été précédemment présenté dans le contexte de scène a subi une rotation de 180° sur le plan vertical afin de mieux contrôler que les participants identifiaient les objets sur la base de leur identité plutôt que sur la base d'informations seulement physiques (leur forme et leur orientation). Nous avons également contrebalancé le côté de présentation des objets cibles et distracteurs entre les essais. Un objet cible apparaissant autant de fois à droite qu'à gauche.

Une session expérimentale contenait 160 essais répartis en 16 blocs de 10 essais ayant le même temps de présentation (i.e., 10 essais à 50 ms ou 10 essais à 200 ms) présentés dans un ordre aléatoire. Au sein de chaque bloc, l'ordre des essais était également aléatoire. Nous avons donc 40 objets × 2 conditions de congruence objet/contexte (objet dans un contexte congruent vs. objet dans un contexte incongruent) × 2 temps de présentation des objets en contexte (50 ms vs. 200 ms). L'expérience durait environ 8 minutes et contenait une pause tous les 40 essais.

Figure 2.19

Procédure Expérimentale de l'Expérience 6.



Note. Chaque essai débutait par un point de fixation de 500 ms, suivi de la présentation du contexte de scène sans objet pendant 150 ms, de la présentation du contexte de scène avec un objet congruent ou incongruent au centre pendant 50 ms ou 200 ms, d'un masque pendant 50 ms ou 200 ms, de la présentation de deux objets (i.e., l'objet cible présenté précédemment en contexte et l'objet distracteur qui lui était apparié sur la base de la similarité sémantique et physique) et d'un écran gris de 1000 ms. La tâche des participants consistait à indiquer quel objet, parmi les deux proposés, avait été présenté dans le contexte de scène.

2.4.2. Analyse des données

Les données ont été analysées à l'aide du logiciel R (R Core Team, 2019) et du package *lme4* (Bates, Mächler, et al., 2015) pour effectuer deux analyses, une pour chaque variable dépendante, à savoir le taux de bonnes réponses et le temps de réponse correcte. Concernant tout d'abord l'analyse menée sur les taux de bonnes réponses, nous avons conduit une

Chapitre 2 – Contribution empirique

régression logistique à effets mixtes de l'effet du temps de présentation de l'objet (50 ms vs. 200 ms) et de la congruence sémantique entre l'objet et le contexte de scène (objet et contexte congruents vs. incongruents) dans un design intra-sujet 2×2 . Concernant l'analyse des temps de réponse correcte, nous avons mis en place une analyse linéaire à effets mixtes de l'effet du temps de présentation de l'objet (50 ms vs. 200 ms) et de la congruence sémantique entre l'objet et le contexte de scène (objet et contexte congruents vs. incongruents) dans un design intra-sujet 2×2 . Pour chacune des analyses, le temps de présentation de l'objet, la congruence entre l'objet et son contexte et leur terme d'interaction ont été introduits dans le modèle comme effets fixes. Les intercepts pour les sujets et les stimuli, ainsi que les pentes aléatoires par sujet et par stimulus pour l'effet du temps de présentation de l'objet, de la congruence entre l'objet et son contexte et de leur terme d'interaction ont été spécifiés comme effets aléatoires. Les essais dans lesquels les temps de réponse étaient supérieurs à 2 secondes ont été supprimés préalablement aux analyses statistiques (cela représentait 3.9% des essais).

D'après l'*opposing process theory* (Press et al., 2020), le mécanisme de sharpening interviendrait en premier, permettant de confirmer les prédictions, puis le mécanisme de dampening prendrait ensuite le relais afin de privilégier le traitement des autres caractéristiques inattendues. Nous nous attendions donc à ce que pour un temps de présentation de l'objet en contexte très bref de 50 ms, les objets congruents avec le contexte soient ensuite plus correctement et plus rapidement identifiés dans la paire d'objets que les objets incongruents avec le contexte. En revanche, pour un temps de présentation de l'objet en contexte plus long de 200 ms, nous nous attendions à ce que les objets incongruents avec le contexte soient ensuite plus correctement et plus rapidement identifiés dans la paire d'objets que les objets congruents avec le contexte. Nous avons donc testé, pour nos deux variables dépendantes, l'interaction entre le temps de présentation de l'objet et la congruence objet-contexte.

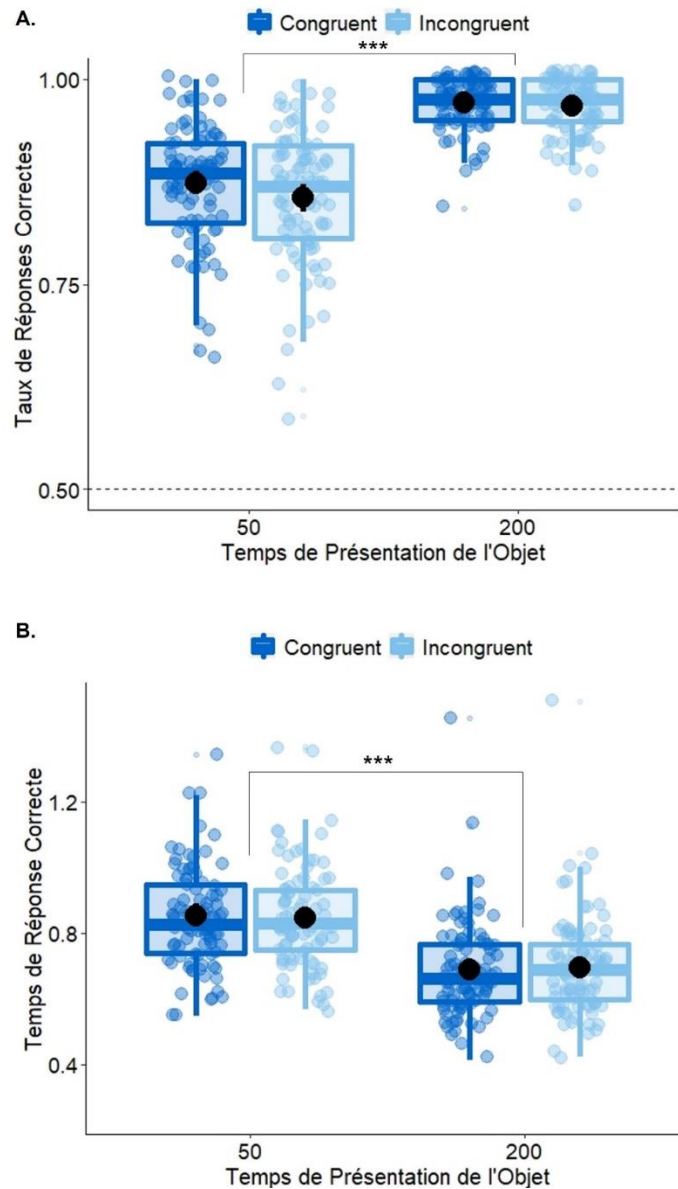
Concernant l'analyse des taux de réponses correctes, la significativité des effets a été testé grâce à un test de Wald. Concernant l'analyse des temps de réponse correcte, nous avons utilisé la méthode proposée par Bates, Kliegl, et al. (2015) pour construire des modèles mixtes parcimonieux limitant les problèmes de convergence. L'inspection visuelle des graphiques de résidus n'a pas révélé de déviations évidentes de l'homoscédasticité ou de la normalité. Les p-valeurs ont été obtenues par approximation de Satterthwaite avec le package *lmerTest* (Kuznetsova et al., 2017). Le seuil de significativité a été fixé à .05 pour chacune des deux analyses. Les tailles d'effet ont été estimées à l'aide du d_z de Cohen (Lakens, 2013).

2.4.3. Résultats

Les taux de réponses correctes moyens ainsi que les temps de réponse correcte sont présentés dans la Figure 2.20. Concernant l'analyse des taux de réponses correctes, les résultats ont révélé un effet principal significatif du temps de présentation de l'objet, $\beta = 2.03$, $z = 13.52$, $p < .001$. Les taux de réponses correctes étaient plus élevés lors de la présentation de l'objet durant 200 ms ($M = 0.97$, $SE = 0.02$) que durant 50 ms ($M = 0.86$, $SE = 0.04$). En revanche, ni l'effet principal de la congruence, $\beta = - 0.28$, $z = - 1.50$, $p = .133$, ni l'effet d'interaction entre le temps de présentation de l'objet et la congruence, $\beta = - 0.04$, $z = - 0.15$, $p = .881$, n'étaient significatifs. Concernant l'analyse des temps de réponse correcte, les résultats ont également révélé un effet principal significatif du temps de présentation de l'objet, $t(77.42) = -17.20$, $p < .001$, $d_z = 1.784$, indiquant que les temps de réponse correcte étaient plus courts lors de la présentation de l'objet durant 200 ms ($M = 688$ ms, $SE = 29$ ms) que durant 50 ms ($M = 840$ ms, $SE = 33$ ms). En revanche, ni l'effet principal de la congruence, $t(39.03) = 0.26$, $p = .797$, $d_z = 0.027$, ni l'effet d'interaction entre le temps de présentation de l'objet et la congruence, $t(12840) = 1.46$, $p = .146$, $d_z = 0.151$, n'étaient significatifs.

Figure 2.20

Taux de Réponses Correctes et Temps de Réponse Correcte Pour Chaque Condition de l'Expérience 6.



Note. A. Taux de réponses correctes moyens pour la discrimination de l'objet présenté en contexte congruent et incongruent durant 50 ms et 200 ms. B. Temps de réponses correctes moyens pour la discrimination de l'objet présenté en contexte congruent et incongruent durant 50 ms et 200 ms. Les points noirs et les barres d'erreur indiquent respectivement la moyenne et l'erreur standard de tous les participants. Les points en couleur correspondent aux observations individuelles. La ligne en pointillés correspond au hasard.

*** $p < .001$.

2.4.4. Discussion de l'Expérience 6

L'objectif de cette sixième expérience était de tester la seconde hypothèse de l'opposing process theory (Press et al., 2020). A savoir, le mécanisme de sharpening interviendrait en premier, permettant de confirmer les prédictions, puis le mécanisme de dampening prendrait ensuite le relais afin de privilégier le traitement des autres caractéristiques inattendues. Pour tester cette hypothèse nous avons donc mis en place un nouveau paradigme dans lequel les participants devaient discriminer des objets entourés d'une scène présentant un contexte sémantiquement congruent ou incongruent avec l'objet et nous avons manipulé leur temps de présentation. Les objets entourés d'un contexte étaient présentés soit pendant une durée très brève de 50 ms, condition expérimentale qui devrait impliquer le mécanisme de sharpening, soit pendant une durée plus longue de 200 ms, condition expérimentale qui devrait impliquer le mécanisme de dampening. Les participants avaient pour tâche de discriminer parmi deux objets physiquement et sémantiquement similaires celui qui avait été présenté en contexte précédemment. Nous nous attendions à ce que pour un temps de présentation de l'objet en contexte très bref de 50 ms, les objets congruents avec le contexte soient ensuite plus correctement et plus rapidement identifiés que les objets incongruents avec le contexte. En revanche, pour un temps de présentation de l'objet en contexte plus long de 200 ms, nous nous attendions à ce que les objets incongruents avec le contexte soient ensuite plus correctement et plus rapidement identifiés que les objets congruents avec le contexte.

Contrairement à nos hypothèses, les résultats n'ont pas mis en évidence de préférence du mécanisme de sharpening sur le mécanisme de dampening lors de la perception visuelle d'objets en contexte. En effet, l'effet d'interaction entre le temps de présentation et la congruence de l'objet n'était pas significatif ni sur les taux de bonnes réponses, ni sur les temps de réponse correcte. Seul l'effet principal du temps de présentation était significatif pour ces deux mesures, suggérant simplement de meilleures performances d'identification de l'objet cible lorsqu'il était présenté plus longtemps en contexte (200 ms). D'ailleurs, les taux de réponses correctes très élevés (97%) dans la condition 200 ms suggèrent que la tâche était peut-être trop facile, ce qui a pu masquer les effets attendus. Effectivement, contrairement à la tâche de Spaak et al. (2022) où les stimuli étaient des photographies de scènes (Figure 1.5) contenant plusieurs objets en plus des objets cibles qui pouvaient se trouver en divers endroits, nos stimuli correspondaient à une superposition artificielle d'un objet sur une scène, l'objet étant toujours bien dissocié du reste de la scène et toujours présentés au centre de l'écran afin de nous assurer

Chapitre 2 – Contribution empirique

qu'il avait bien été vu par les participants, même à un temps très bref de 50 ms de présentation. Les participants n'avaient donc qu'à fixer le centre de l'écran pour voir l'objet cible et la tâche était grandement facilitée par rapport à celle de Spaak et al. dans laquelle les objets cibles étaient moins saillants et les participants ne savaient pas où se trouvait l'objet cible dans la scène. Des propositions pour pallier ce biais méthodologique dans de futures études seront discutées dans la Discussion Générale.

Chapitre 3 – Discussion générale et perspectives

1. Apports des résultats aux questions théoriques et limites

Les modèles actuels dits « proactifs » de la perception visuelle supposent que le contenu de la perception visuelle n'est pas déterminé uniquement par les caractéristiques extrinsèques des entrées sensorielles, mais reflète plutôt leur combinaison avec les prédictions, ou attentes, qui découlent des connaissances et expériences antérieures de l'observateur avec son environnement visuel. Bien que de nombreuses études soutiennent l'idée que les prédictions influencent le traitement de l'information visuelle, dans quelle mesure et de quelle manière celles-ci contraignent voire modifient la perception et l'expérience subjective des stimuli visuels continue de faire l'objet de nombreux débats. Ce travail de thèse a eu pour objectif de caractériser comment ces prédictions basées sur nos connaissances et expériences antérieures avec notre environnement visuel influencent la perception visuelle subjective de stimuli visuels complexes tels que les scènes et les objets.

1.1. Dans quelle mesure ce que nous savons influence-t-il ce que nous voyons ?

1.1.1. Synthèse et apport des résultats

Si de nombreux travaux tendent à suggérer une influence des prédictions sur le traitement de l'information visuelle, dans quelle mesure ces prédictions influencent effectivement la perception ? Comme nous l'avons vu dans l'Introduction de cette thèse, la réponse à cette question se heurte à la définition même que l'on donne à la perception, ainsi qu'à la difficulté de la mesurer de manière fiable et isolée de tout autre processus qui serait considéré comme « non-perceptif ». En effet, les auteurs qui étudient l'influence des prédictions sur la perception à travers des mesures objectives basées sur les performances lors de tâches de reconnaissance ou de discrimination (Brandman & Peelen, 2017; Cheadle et al., 2015; Davenport, 2007; Davenport & Potter, 2004; Greene et al., 2015; Joubert et al., 2007, 2008; Kok, Jehee, et al., 2012; Roux-Sibilon et al., 2019; Teufel et al., 2018; Wyart et al., 2012) se voient généralement critiqués sur le fait que ces mesures ne reflèteraient pas uniquement la

Chapitre 3 – Discussion générale et perspectives

perception et seraient contaminées par des processus mnésiques, sémantiques ou décisionnels sur lesquels pourrait s'exercer l'effet des prédictions (Bang & Rahnev, 2017; Firestone & Scholl, 2015; Rungratsameetaweemana et al., 2018; Rungratsameetaweemana & Serences, 2019). Pour limiter ces biais potentiels, plusieurs auteurs se sont intéressés à la perception d'un point de vue phénoménologique en utilisant des tâches de reports subjectifs concernant l'apparence de stimuli visuels (Chalk et al., 2010; Denison et al., 2016; Han & VanRullen, 2016; Hansen et al., 2006; Lupyan, 2015, 2017; Perez et al., 2020). Si ces études suggèrent effectivement que les connaissances et les prédictions qui en découlent modifient l'apparence des stimuli visuels, elles ont généralement été menées sur des stimuli simples et peu écologiques qui ne reflètent pas la complexité de notre environnement visuel. **Le premier objectif de ce travail de thèse** était donc de préciser dans quelle mesure les connaissances que nous avons sur notre environnement et les prédictions qui en découlent sont susceptibles de modifier l'apparence subjective de stimuli visuels complexes tels que les objets et les scènes.

Les Expériences 1, 2 et 3 ont donc été menées dans le but de répondre à cette problématique au niveau de la perception d'objets pour l'Expérience 1 et de la perception de scènes entières pour les Expériences 2 et 3. Pour ce faire, nous avons utilisé dans les Expériences 1 et 2 une tâche d'ajustement perceptif du niveau de flou adaptée de Lupyan (2017) qui montrait qu'à niveau de flou objectivement égal, les stimuli conformes aux connaissances antérieures (des mots) étaient subjectivement perçus comme plus nets que les stimuli sans signification (pseudo-mots). L'utilisation de cette tâche visait à limiter au maximum l'influence de facteurs non perceptifs pouvant expliquer une différence d'ajustement entre les stimuli. Dans cette perspective, nous avons par exemple présenté des stimuli simultanément plutôt que successivement afin de limiter le recours aux facteurs mnésiques. Nous avons également utilisé des tâches d'ajustement plutôt que des tâches de discrimination qui impliquent dans une plus grande mesure des aspects décisionnels. De plus, les tâches d'ajustement portaient sur des paires de stimuli identiques où seule la caractéristique à ajuster (i.e., le niveau de flou) pouvait varier, de façon à limiter toute différence objective de signal autre que le flou entre les deux stimuli pouvant expliquer une différence de perception. Nous avons manipulé la disponibilité d'informations sémantiques pertinentes pour générer des prédictions sur le contenu de l'image (contexte de l'objet intact ou brouillé pour l'Expérience 1 ; scène à l'endroit ou à l'envers pour l'Expérience 2). Les résultats des deux expériences ont montré que pour percevoir les stimuli sample et target comme aussi flous l'un que l'autre, les participants attribuaient (1) un niveau de flou plus élevé à un stimulus sample prédictible qu'à un stimulus target non prédictible et

Chapitre 3 – Discussion générale et perspectives

(2) un niveau de flou plus faible à un stimulus sample non prédictible qu'à un stimulus target prédictible. Cela signifie qu'à un niveau de flou objectivement identique, les objets et les scènes prédictibles seraient subjectivement perçus comme plus nets que les objets et scènes non prédictibles. Ces résultats ont été répliqués dans les Expériences 4 et 5 et confirmés par l'Expérience 3 demandant cette fois aux participants une comparaison explicite sur la netteté des stimuli. Dans l'Expérience 3, lorsque les deux scènes avaient un niveau de flou physiquement égal, nous avons observé que les participants jugeaient explicitement les scènes à l'endroit comme plus nettes que les scènes à l'envers (ou les scènes à l'envers comme plus floues que les scènes à l'endroit).

Ces résultats sont en accord avec les résultats d'études précédentes suggérant que des silhouettes d'objets (Perez et al., 2020) ou des mots (Lupyan, 2017) floutés sont subjectivement perçus comme plus nets lorsqu'ils peuvent être associés à des connaissances antérieures. Nos résultats permettent donc d'étendre ces conclusions à la perception de photographies de scènes et d'objets en contexte, et ainsi de démontrer l'influence des connaissances et des prédictions qui en découlent sur l'apparence subjective de stimuli plus complexes et écologiques. En ce sens, nos travaux fournissent également des arguments à l'appui des modèles de la reconnaissance de scènes et d'objets tels que le modèle proactif de reconnaissance d'objets (Bar, 2003, 2007; Trapp & Bar, 2015) ou le modèle fréquentiel de la catégorisation de scènes (Kauffmann et al., 2014; Peyrin et al., 2010, 2021; Roux-sibilon, 2020) qui supposent un rôle crucial des prédictions générées sur la base du traitement rapide de l'information contextuelle des scènes pour guider leur perception. Nos résultats permettent donc de préciser ces modèles en suggérant que les prédictions générées sur la base de l'information contextuelle des scènes facilitent non seulement la reconnaissance des scènes et des objets, mais améliorent également leur perception.

En accord avec les nombreux travaux ayant étudié l'influence des prédictions sur le traitement de l'information visuelle (Brandman & Peelen, 2017; Cheadle et al., 2015; Davenport, 2007; Davenport & Potter, 2004; Greene et al., 2015; Han & VanRullen, 2016; Hansen et al., 2006; Joubert et al., 2008; Lai et al., 2020; Lupyan, 2017, 2015; Olkkonen et al., 2008; Palmer, 1975; Perez et al., 2020; Roux-Sibilon et al., 2019; Stein & Peelen, 2015; Teufel et al., 2018; Witzel et al., 2011; Wyart et al., 2012), nos résultats suggèrent plus largement que nos prédictions influencent non seulement « l'efficacité avec laquelle nous voyons » (i.e., la rapidité du traitement ou l'exactitude de la reconnaissance), mais aussi la phénoménologie de la perception : « comment nous voyons ». En ce sens, nos résultats fournissent des arguments

Chapitre 3 – Discussion générale et perspectives

supplémentaires en faveur de la pénétrabilité cognitive de la perception (Lupyan, 2015; O’Callaghan et al., 2017) selon laquelle la perception n'affecte pas seulement l'état cognitif ou les connaissances d'une personne : la perception est également contrainte par ceux-ci. Cognition et perception ne fonctionneraient donc pas de manière indépendante l'une de l'autre mais s'influenceraient mutuellement.

1.1.2. Limites et perspectives

Nous ne pouvons évidemment pas exclure la possibilité que certains de nos effets soient non perceptifs mais nous avons essayé de les limiter au maximum en intégrant une condition contrôle présentant deux stimuli parfaitement identiques en target et en sample dans nos Expériences 1, 2 et 5. Notre raisonnement était que si les connaissances et les prédictions qui en découlent influencent la perception subjective des objets et des scènes, ces effets perceptifs devraient affecter tous les objets et scènes présents à l'écran. Par exemple, si un objet attendu dans un contexte donné est perçu comme plus net qu'il ne l'est en réalité, les objets identiques affichés simultanément et partageant le même contexte devraient également bénéficier de cette amélioration perceptive, ce qui se traduirait par l'absence de différence de flou perçue entre eux (cf. raisonnement du El Greco fallacy, Firestone & Scholl, 2014). Selon cette logique, la différence de niveau de flou entre les stimuli sample prédictibles et non prédictibles observée dans les paires target-sample différentes devrait donc diminuer voire disparaître dans les paires target-sample identiques.

Si nous avons bien observé un tel résultat dans les Expériences 1 et 2, ce n'est pas le cas de l'Expérience 5 dans laquelle les patterns de résultats étaient similaires entre les conditions de paires target-sample identiques et différentes. Cependant, dans l'Expérience 5, nous avons utilisé des stimuli bien plus grands que dans les deux précédentes expériences et leur perception a pu être influencée par des différences d'acuité visuelle dans le champ visuel. Pour rappel, Dubova et Goldstone (2022) ont remis en question le fait que les effets des prédictions sur la perception subjective devraient affecter de la même manière tous les éléments du champ visuel en mettant en évidence que lors de tâches d'ajustement perceptif comme les nôtres, les stimuli en target et en sample (même s'ils sont identiques) peuvent difficilement être perçus de manière similaire au même moment. Par exemple, dans le cas de leur étude, les participants fixaient plus souvent le sample pendant la tâche que la target. Cette dernière apparaissait ainsi le plus souvent dans le champ visuel périphérique et donc avec une acuité visuelle moindre. Il est possible que

Chapitre 3 – Discussion générale et perspectives

ce biais perceptif (cette fois) soit présent dans notre Expérience 5 et que les stimuli objectivement identiques des paires target-sample identiques soient perçus différemment selon la position de la fixation oculaire. Pour que cette condition de paires target-sample identiques soit un meilleur contrôle, il semble donc important que les deux stimuli du sample et de la target soient situés dans le champ visuel central afin qu'ils soient perçus tout deux avec la meilleure acuité possible. Cette sensibilité à l'acuité visuelle est au final cohérente avec l'influence que peut avoir la fiabilité du signal sur l'effet des prédictions (ce que nous avons démontré au cours des Expériences 4 et 5). La fiabilité étant meilleure en vision centrale qu'en vision périphérique, les effets des prédictions pourraient donc différer entre ces deux zones du champ visuel et induire une perception subjective différente de la target et du sample même lorsqu'ils sont identiques. Afin de résoudre ce problème, il pourrait être envisageable de répliquer nos expériences en ajoutant une mesure des mouvements oculaires. Cela permettrait ainsi de vérifier si, comme Dubova et Goldstone, les participants ont plus tendance à fixer le sample que la target.

D'autres paradigmes expérimentaux pourraient également être envisagées pour s'assurer que les résultats ne découlent pas de biais de réponse ou d'effets non perceptifs. Par exemple, Williams et al. (2022) ont conduit une expérience dans laquelle ils utilisaient des *morphs* construits à partir d'une paire de silhouettes objet-animal. Plus précisément, ces stimuli morphs étaient construits en prenant une silhouette d'un objet (e.g., un avion), une silhouette d'un animal (e.g., un aigle) et en créant un continuum passant de l'image de l'objet à l'image de l'animal en déformant petit à petit le stimulus (Figure 3.1). Il y avait donc plusieurs images morphs pour chaque paire objet-animal allant de morphs ressemblant plus à l'objet à des morphs ressemblant plus à l'animal. Au cours de cette expérience, les auteurs faisaient d'abord entendre un son aux participants qui pouvait être soit relié à l'une des deux images de la paire objet-animal (e.g., un bruit d'avion ou un cri d'aigle) soit non relié à la paire (e.g., le bruit d'un train). Pendant la diffusion du son, un morph target d'une paire objet-animal apparaissait progressivement dans du bruit visuel. Les participants devaient cliquer sur un bouton lorsqu'ils estimaient avoir assez d'informations sur le morph target pour pouvoir le reproduire. Ils devaient ensuite effectuer une tâche de report visuel dans laquelle un morph sample de la même paire objet-animal que le morph target apparaissait et la tâche consistait à ajuster la forme de ce morph sample pour qu'il soit identique au morph target présenté précédemment (i.e., rajouter un peu plus d'animal et moins d'objet dans le morph sample ou un peu plus d'objet et un peu moins d'animal pour correspondre à ce qui avait été présenté en target). Les résultats ont montré

Chapitre 3 – Discussion générale et perspectives

que lorsque le son était relié à l'une des images du morph target, les participants estimaient pouvoir reproduire le morph avec moins d'information (i.e. lorsque le morph était mêlé à plus de bruit) que lorsque le son n'était pas relié aux images du morph. Cela suggère que les sons reliés facilitaient la perception des morphs dans le bruit. De plus, les sons reliés influençaient également l'ajustement des morphs sample lors de la phase de report dans le sens de morphs plus proches de l'animal ou de l'objet congruent avec le son entendu (e.g., les morphs sample étaient ajustés à une forme qui ressemblait plus à un aigle si le son entendu précédemment était le cri d'un aigle alors qu'ils ressemblaient plus à un avion si le bruit entendu était celui d'un avion).

Figure 3.1

Exemple des Stimuli de Williams et al. (2022).



Note. Les stimuli étaient des silhouettes sur un continuum allant d'une silhouette d'objet à une silhouette d'animal. Figure adaptée de Williams et al. (2022).

Pour s'assurer que ces résultats étaient bien la conséquence de mécanismes perceptifs et non de processus post-perceptifs comme la prise de décision, les auteurs ont mis en place une expérience contrôle. Pour cela, ils ont présenté les sons au moment où ils devraient avoir le plus d'impact sur les processus décisionnels, à savoir pendant la phase d'ajustement du morph sample plutôt que pendant la présentation du morph target. Avec ce changement du moment de diffusion du son, les auteurs ont observé une réduction significative des effets du son sur l'ajustement du morph sample. Cela suggère donc que l'influence du son sur l'ajustement des morphs ne découlait pas d'effets des prédictions sur la prise de décision mais plutôt sur la perception subjective des morphs target. Ce paradigme de report perceptif pourrait donc être intéressant à adapter dans le contexte de nos travaux puisqu'il nous permettrait de ne présenter qu'un seul stimulus à la fois (e.g., un objet dans un contexte intact ou brouillé suivi d'une phase de report du flou de l'objet isolé) et ainsi éviter les biais perceptifs potentiels liés aux fixations oculaires entre les stimuli target et sample et la baisse d'acuité visuelle en vision périphérique

Chapitre 3 – Discussion générale et perspectives

dans nos études. L'inclusion d'une condition contrôle similaire à celle de Williams et al. (2022) où le contexte serait cette fois-ci présenté pendant la phase de report permettrait en outre de vérifier si les effets observés peuvent être attribuables à une influence du contexte sur les processus décisionnels plutôt que perceptifs.

Un autre avantage de ne présenter qu'un seul stimulus à chaque essai est que cela permet d'éviter que, lors de la présentation simultanée de deux stimuli, la perception de l'un des deux (e.g., un objet dans un contexte intact) ne contamine directement la perception de l'autre stimulus (e.g., un objet dans un contexte brouillé). Effectivement, dans le cadre de nos expériences d'ajustement perceptif, la perception du contexte intact ou congruent peut permettre de désambiguïser l'objet qui ne bénéficie pas de ce contexte et qui a un contexte brouillé ou incongruent. Cela peut donc biaiser sa perception. L'utilisation d'un paradigme inspiré de celui de Williams et al. (2022) pourrait donc permettre de vérifier les résultats de l'Expérience 5 en présentant des objets congruents ou incongruents avec le reste de la scène, tout en évitant de présenter plusieurs stimuli simultanément pour réaliser la tâche.

En plus de montrer que nos connaissances sur l'environnement visuel influencent la perception d'objets, l'étude de Williams et al. (2022) suggère de façon intéressante que les prédictions influencent le traitement d'autres caractéristiques que le flou (comment on voit) et en particulier la perception de la forme (ce que l'on voit). Dans le cadre de cette thèse, nous avons étudié la perception par le biais du flou car il s'agissait d'une caractéristique sur laquelle les hypothèses dérivées du mécanisme de sharpening étaient faciles à vérifier. Mais il est effectivement probable que le traitement d'autres caractéristiques (notamment traitées par les aires visuelles de bas niveau) soit affecté, comme plusieurs études l'ont observé dans le contexte de la perception des couleurs (Hansen et al., 2006; Lupyan, 2015; Olkkonen et al., 2008; Witzel et al., 2011), du contraste (Han & VanRullen, 2016) ou de la forme (Williams et al., 2022). Dans cette optique, il pourrait être intéressant de répliquer l'expérience de Williams et al. avec des images morphs d'objet en contexte afin d'étudier l'influence de l'information contextuelle sur la perception de la forme des objets (Figure 3.2). Par exemple, un morph meuble-animal serait dans un premier temps présenté dans un contexte de scène congruent ou incongruent à un niveau de morphing ambigu (i.e., à un niveau où les participants voient autant le meuble que l'animal dans le morph). Dans un second temps, le morph présenté précédemment apparaîtrait à un niveau aléatoire de morphing sur le continuum meuble-animal. Les participants devraient alors ajuster le niveau de morphing du second morph pour qu'il corresponde au morph présenté en contexte précédemment. Dans ce cas, nous pourrions nous attendre à ce que les participants

Chapitre 3 – Discussion générale et perspectives

reproduisent le premier morph avec un biais d'ajustement dans le sens de la forme qui est cohérente avec le contexte dans lequel il était présenté. Par exemple, le morph tabouret-mangouste (Figure 3.2) serait ajusté dans le sens d'un morph ressemblant plus à un tabouret que le morph initial lorsqu'il a été présenté dans une scène d'intérieur tandis qu'il serait ajusté dans le sens d'un morph ressemblant plus à une mangouste que le morph initial lorsqu'il a été présenté dans une scène d'extérieur.

Figure 3.2

Exemple d'Image Morph Animal-Meuble en Contexte.



Note. En haut, exemples d'un morph tabouret-mangouste présenté dans le contexte d'une scène d'intérieur à gauche (congruent avec le tabouret et incongruent avec la mangouste) et dans le contexte d'une scène d'extérieur à droite (congruent avec la mangouste et incongruent avec le tabouret). En bas, exemples de morphs sur le continuum tabouret-mangouste.

Enfin, il est important de noter que les expériences de ce travail de thèse ont été majoritairement réalisées sur des participantes de sexe féminin et comptent peu de participants masculins. Les participants aux expériences étant des étudiants de psychologie de l'Université Grenoble Alpes, la proportion de femmes dans ce cursus universitaire est très largement supérieure à la proportion d'hommes, ce qui se retrouve dans nos échantillons. Il est donc possible que nos résultats se généralisent plus difficilement à une population masculine du fait

Chapitre 3 – Discussion générale et perspectives

de potentielles différences au niveau de la perception visuelle entre hommes et femmes (McGuinness, 1976; Sabatinelli et al., 2004). A notre connaissance, il n'existe pas d'étude ayant montré une influence du sexe sur les effets des prédictions sur la perception subjective. Il serait néanmoins important de répliquer ces études avec un échantillon plus représentatif de la population et comportant autant d'hommes que de femmes afin de s'assurer que nos résultats peuvent être généralisés aux participants de sexe masculin.

1.2. Modulation des effets des prédictions sur la perception subjective

1.2.1. Synthèse et apport des résultats

Comme nous l'avons vu dans le Chapitre 1, les études précédentes rapportent des résultats contradictoires concernant les mécanismes par lesquels les connaissances a priori et les prédictions qui en découlent modulent le traitement des entrées sensorielles. Alors que certains travaux suggèrent que celui-ci serait biaisé en faveur de ce qui est attendu et donc conforme aux connaissances a priori (mécanisme de sharpening), d'autres rapportent au contraire que le traitement des caractéristiques attendues des signaux sensoriels serait atténué, de façon à privilégier le traitement de ce qui n'est pas conforme aux connaissances et donc inattendu (mécanisme de dampening).

Le développement récent de l'opposing process theory, qui fait écho à certaines formalisations des modèles du codage prédictif (Feldman & Friston, 2010; Friston, 2005), fournit un cadre théorique permettant de réconcilier ces résultats a priori contradictoires. Selon ce modèle, les mécanismes de sharpening et de dampening coexistent mais interviennent alternativement selon le décours temporel du traitement visuel, ainsi que selon la fiabilité des signaux sensoriels (Press et al., 2020). Le mécanisme de sharpening interviendrait en premier pour traiter rapidement les caractéristiques attendues, permettant alors de confirmer les prédictions, puis le mécanisme de dampening prendrait ensuite le relais pour stimuler le traitement des caractéristiques inattendues. Cette séquence dépendrait néanmoins de la fiabilité des signaux sensoriels. Le mécanisme de sharpening prédominerait en situation de signal visuel bruité et peu fiable où le système visuel aurait tout intérêt à donner plus de poids aux caractéristiques prédites et à supprimer le traitement des caractéristiques non prédites qui pourraient être attribuables à du bruit. Lorsque le signal visuel est fiable et non ambigu, il serait par contre peu pertinent de supprimer le traitement de ce qui est inattendu et donc potentiellement informatif. Le mécanisme de dampening prendrait alors le dessus en atténuant

Chapitre 3 – Discussion générale et perspectives

le traitement des caractéristiques prédites de façon à privilégier le traitement des caractéristiques inattendues, permettant de mettre à jour les connaissances. Ce modèle est d'ailleurs cohérent avec les théories du codage prédictif qui proposent d'une part que la représentation initiale du stimulus dans les unités de prédictions correspondrait à ce qui est attendu (sharpening) avant que cette représentation ne soit mise à jour sur la base des signaux d'erreur de prédiction (dampening), et d'autre part que le poids relatif des signaux de prédiction et d'erreur de prédiction varierait en fonction de la fiabilité du signal visuel. Les signaux de prédiction auraient plus de poids lorsque le stimulus est bruité (ce qui correspondrait au mécanisme de sharpening), tandis que les signaux d'erreur de prédiction auraient plus de poids lorsque le stimulus est fiable (ce qui correspondrait au mécanisme de dampening). Ainsi, l'influence des prédictions sur la perception visuelle augmenterait à mesure que la fiabilité du signal diminue, c'est-à-dire à mesure que le signal devient ambigu ou bruité (Friston, 2005; Press et al., 2020; Rao & Ballard, 1999). Cette théorie semble compatible avec les résultats de la littérature. En particulier, les études ayant mis en évidence un mécanisme de dampening utilisaient généralement des stimuli non ambigus (e.g., images de scènes ou d'objets), tandis que les études en faveur d'un mécanisme de sharpening portaient sur la perception de stimuli bruités ou ambigus (e.g., flous). Cependant, à notre connaissance, les postulats de ce modèle théorique n'ont jamais été testés systématiquement dans une même étude. **Le second objectif de cette thèse** était donc de mieux comprendre les mécanismes par lesquels les prédictions influencent la perception subjective et de tester explicitement les hypothèses issues de l'opposing process theory en examinant de quelle manière la perception d'objets attendus et inattendus (et donc l'effet de la validité des prédictions) pouvait varier en fonction de la fiabilité du signal visuel et du décours temporel du traitement de l'information.

Bien que cela n'était pas l'objectif principal de ces expériences, l'influence de la fiabilité du signal a pu être examinée dans un premier temps dans le contexte des Expériences 1 et 2. Ces expériences ont montré que les scènes et objets prédictibles étaient perçus comme plus nets que les mêmes scènes et objets non prédictibles. De plus, nous avons observé que lorsque les stimuli en sample et en target étaient différents, la différence entre le niveau de flou attribué à un stimulus sample prédictible et non prédictible (i.e., l'effet des prédictions) augmentait à mesure que le niveau de flou du stimulus target augmentait. En revanche, pour l'Expérience 2 présentant des scènes entières à l'endroit ou à l'envers, l'effet des prédictions ne variait pas significativement selon le niveau de flou du stimulus target. Nous nous sommes cependant intéressées à ces effets a posteriori puisque nos expériences n'étaient pas construites dans le but

Chapitre 3 – Discussion générale et perspectives

de les tester. Concernant l'Expérience 3 dans laquelle les participants devaient indiquer quelle scène était perçue comme la plus floue ou la plus nette, nous avons également testé cette hypothèse. Si l'influence des prédictions sur la netteté perçue augmente lorsque la fiabilité du signal diminuait (i.e., lorsque le niveau de flou des scènes augmentait), la tendance à sélectionner la scène à l'envers comme la plus floue (ou la scène à l'endroit comme la plus nette) devrait augmenter avec le niveau de flou des paires de scènes. Contrairement à ce qui était attendu, les participants n'ont pas nécessairement jugé la scène à l'envers comme étant la plus floue plus fréquemment à mesure que les paires de scènes devenaient plus floues. Nous avons même observé la tendance inverse, à savoir que les participants jugeaient la scène à l'envers comme étant la plus floue moins fréquemment lorsque les deux scènes apparaissaient à un niveau de flou très élevé - et avaient donc une faible fiabilité - par rapport à celles qui apparaissaient à un niveau de flou faible ou moyen - et avaient donc une plus grande fiabilité. Cependant, comme détaillé dans la Discussion de l'Expérience 3, la mesure que nous avons utilisée, c'est-à-dire la proportion de sélection de scènes à l'endroit/à l'envers comme étant les plus floues, pourrait ne pas être la mesure la plus adéquate pour évaluer comment l'effet des prédictions sur la netteté perçue des stimuli varie avec la fiabilité du signal. Si cette mesure reflète la tendance à percevoir une scène comme plus nette/floue que l'autre, elle ne permet pas de quantifier l'amplitude de cet effet (i.e., l'importance de la différence de netteté perçue entre les deux scènes). Il était donc nécessaire de mener d'autres expériences pour répondre à cette question.

Dans cette optique, nous avons mis en place une quatrième expérience (Expérience 4) permettant de tester plus rigoureusement si l'effet des prédictions sur la netteté perçue variait en fonction de la fiabilité du signal visuel. Pour cela, nous avons repris le paradigme d'ajustement perceptif utilisé dans l'Expérience 2 sur les scènes à l'endroit et à l'envers en ne conservant que la condition où les stimuli en sample et en target étaient différents et chaque image target était présentée selon quatre niveaux de flou. Nos résultats ont confirmé les analyses post-hoc de l'Expérience 1 en montrant que l'effet de la prédictibilité des scènes sur leur netteté perçue augmentait à mesure que la fiabilité du signal diminuait (i.e., lorsque le niveau de flou du stimulus target augmentait). Cependant, l'effet de la prédictibilité sur la netteté perçue n'augmentait pas linéairement avec le niveau de flou du stimulus target. En effet, la différence entre le niveau de flou attribué à un stimulus sample à l'endroit et à l'envers augmentait lorsque la fiabilité du stimulus target passait d'un niveau élevé (i.e., niveau de flou de 8) à un niveau moyen (i.e., niveau de flou de 13), puis restait constant jusqu'à ce que le stimulus en target soit

Chapitre 3 – Discussion générale et perspectives

très flou et ait donc un niveau de fiabilité faible (i.e., niveau de flou de 18 et 23). Ces résultats indiquent donc que l'effet des prédictions sur la perception augmente avec la diminution de la fiabilité du signal visuel, comme l'ont suggéré les théories du codage prédictif, et soutiennent l'idée d'une prédominance du mécanisme de sharpening lorsque le signal est bruité (Feldman & Friston, 2010; Friston, 2005) et l'opposing process theory (Press et al., 2020). Toutefois, au-delà d'un certain seuil de fiabilité, l'influence des prédictions sur la perception subjective cesserait d'augmenter.

Les manipulations expérimentales des Expériences 1 à 4 concernaient la présence ou l'absence de contenu informatif ou de connaissances préalables permettant de générer des prédictions par le biais du brouillage des informations contextuelles ou de l'inversion de scènes. Cependant, elles ne nous ont pas permis d'évaluer l'effet de la validité des prédictions et donc de tester l'hypothèse d'une prédominance du mécanisme de dampening (atténuation perceptive des stimuli attendus au profit des stimuli inattendus) lorsque le signal est fiable, car nous n'avons pas inclus une condition dans laquelle les stimuli pouvaient être inattendus ou surprenants. L'Expérience 5 avait donc pour objectif de tester l'effet de la validité des prédictions selon la fiabilité du signal visuel. Pour cela, nous avons repris le paradigme d'ajustement perceptif utilisé précédemment que nous avons appliqué à des objets en contexte congruent (permettant la génération de prédictions valides sur l'objet) et incongruent (entraînant la génération de prédictions non valides sur l'objet). La fiabilité du signal a été manipulée en appliquant aux paires d'objets soit un niveau de flou élevé (faible fiabilité du signal), soit un niveau de flou faible (haute fiabilité du signal). Conformément à l'hypothèse de l'opposing process theory (Press et al., 2020), les résultats ont mis en évidence une inversion de l'effet de la validité des prédictions sur la netteté perçue selon la fiabilité du signal visuel. En effet, en condition de faible fiabilité, les participants ont attribué un niveau de flou plus élevé aux objets en sample congruents qu'incongruents. Cela indique qu'à un niveau de flou physiquement égal entre l'objet en sample et en target, les objets attendus apparaissent subjectivement plus nets que les objets inattendus. En condition de haute fiabilité, ce pattern de résultats était inversé. Les participants ont attribué un niveau de flou plus élevé aux objets en sample dans un contexte incongruent que congruent, ce qui indique qu'à un niveau de flou physiquement égal entre l'objet en sample et en target, les objets inattendus sont subjectivement perçus comme plus nets que les objets attendus.

Dans l'ensemble, les résultats de nos Expériences 1 à 5 apportent donc pour la première fois dans le contexte d'une même étude un soutien empirique en faveur des théories du codage

Chapitre 3 – Discussion générale et perspectives

prédictif (Feldman & Friston, 2010; Friston, 2005) et de l’opposing process theory (Press et al., 2020) en ce qui concerne l’effet des prédictions sur la perception en fonction de leur validité et de la fiabilité du signal visuel. En effet, nos résultats sont compatibles avec l’idée que la perception visuelle est dominée par ce que nous attendons lorsque l’environnement est incertain (Expériences 1 à 5), mais tend à être biaisée par ce qui nous surprend lorsque les entrées sensorielles deviennent plus fiables (Expérience 5). Comme suggéré par les auteurs de l’opposing process theory, ces observations pourraient aider à réconcilier les résultats a priori contradictoires de la littérature sur la façon dont les prédictions façonnent la perception. Ces conclusions soulignent également l’importance de considérer explicitement la fiabilité des stimuli utilisés dans les études qui visent à tester l’influence des prédictions sur la perception pour l’interprétation de leurs résultats. L’amélioration de la perception d’un stimulus (e.g., une voiture sur une autoroute) en fonction de la validité des prédictions et de la fiabilité du signal pourrait donc être la conséquence des mécanismes proposés au niveau neuronal par les théories du codage prédictif (Feldman & Friston, 2010; Friston, 2005) et de l’opposing process theory (Press et al., 2020). Bien que non pertinentes pour la tâche, les informations contextuelles des stimuli auraient néanmoins été utilisées pour former des prédictions concernant les entrées sensorielles. Ces prédictions seraient ensuite véhiculées par des connexions descendantes entre les zones adjacentes de la hiérarchie corticale. Ces signaux de prédiction (e.g., contexte d’une scène d’autoroute) pré-activeraient la représentation neuronale des caractéristiques attendues du stimulus de façon descendante aux différents niveaux de la hiérarchie (e.g., représentation d’une voiture dans les aires de la voie ventrale impliquées dans la reconnaissance d’objets qui entraînerait à son tour la pré-activation de la représentation des contours associés dans les aires de plus bas niveau) en renforçant la réponse des neurones dédiés à ces caractéristiques et en inhibant la réponse des neurones dédiés aux caractéristiques inattendues (mécanisme de sharpening). Lorsque les entrées sensorielles correspondent aux prédictions (e.g., présence d’une voiture), aucun autre processus ne serait nécessaire, ce qui accélère le traitement et la reconnaissance du stimulus. En revanche, lorsque les entrées sensorielles divergent fortement des prédictions (e.g., présence d’un bateau dans un contexte d’autoroute), les représentations neuronales dans les aires de bas niveau devraient être dominées par la réponse des neurones dédiés aux caractéristiques inattendues. La réponse de ces neurones augmenterait alors que celle des neurones codant les caractéristiques attendues diminuerait (mécanisme de dampening). A noter que selon l’opposing process theory, il est également possible que le rehaussement perceptif des caractéristiques inattendues résulte simplement d’un poids plus

important accordé au traitement des caractéristiques non-prédites bénéficiant d'une allocation attentionnelle plus importante – sans que cela ne soit associé à une diminution du poids accordé au traitement des caractéristiques prédites. Le poids plus important accordé aux caractéristiques non prédites favoriserait le traitement et la perception d'un contenu inattendu et donc potentiellement informatif utilisé pour mettre à jour les prédictions. Cependant, lorsque les entrées sensorielles sont bruitées, leur représentation neuronale devrait l'être aussi. La réponse des neurones codant les caractéristiques non prédites pourrait donc ne pas différer suffisamment de la réponse des neurones codant les caractéristiques prédites pour déclencher le mécanisme de dampening. Dans ce cas, la réponse accrue des neurones dédiés aux caractéristiques prédites serait toujours prédominante, conduisant à la perception plus nette des stimuli attendus relativement aux stimuli inattendus. Dans ce contexte, la fiabilité du signal serait donc estimée comme la variance de la distribution des réponses neuronales au stimulus (plus la variance est grande, moins le signal est fiable, Ernst & Banks, 2002). Cette estimation sensorielle de la fiabilité du signal pourrait également être combinée à des prédictions descendantes basées sur les connaissances de l'observateur (e.g., savoir que les entrées sensorielles sont susceptibles d'être (non) fiables ; Yon & Frith, 2021). Les résultats de nos études permettent donc de préciser ces mécanismes en mettant en évidence leurs conséquences sur la perception subjective.

1.2.2. Limites et perspectives

Nos résultats concernant le rehaussement perceptif des stimuli inattendus (mécanisme de dampening) restent toutefois à confirmer. En effet, une des limites inhérentes au paradigme utilisé dans la plupart de nos études concerne le fait qu'en manipulant la fiabilité du signal via le niveau de flou des stimuli (stimuli très flous vs. relativement nets), il devient impossible d'ajuster un stimulus qui est déjà net comme encore plus net. De plus, il est probable que lorsque les stimuli sont relativement nets et non ambigus, les effets des prédictions sur la netteté perçue soient plus subtils puisque leur rehaussement perceptif aurait un bénéfice moindre, relativement à des stimuli plus bruités et ambigus. La manipulation de la fiabilité des stimuli via leur niveau de flou dans nos études a donc pu limiter les effets du mécanisme de dampening sur la perception subjective lorsque nos stimuli se trouvaient en condition de haute fiabilité du signal (i.e., lorsque les stimuli étaient assez nets). Il serait donc intéressant de manipuler d'autres caractéristiques du signal (e.g., contraste, couleur), ce qui pourrait permettre de mieux mettre en évidence le rehaussement perceptif des stimuli inattendus. Par exemple, il est possible que lorsque des stimuli attendus et inattendus sont présentés avec un contraste ou une luminance

Chapitre 3 – Discussion générale et perspectives

moyenne permettant de bien les voir, les stimuli inattendus soient perçus comme plus contrastés ou de couleur plus vive que les stimuli attendus (Kok & de Lange, 2015).

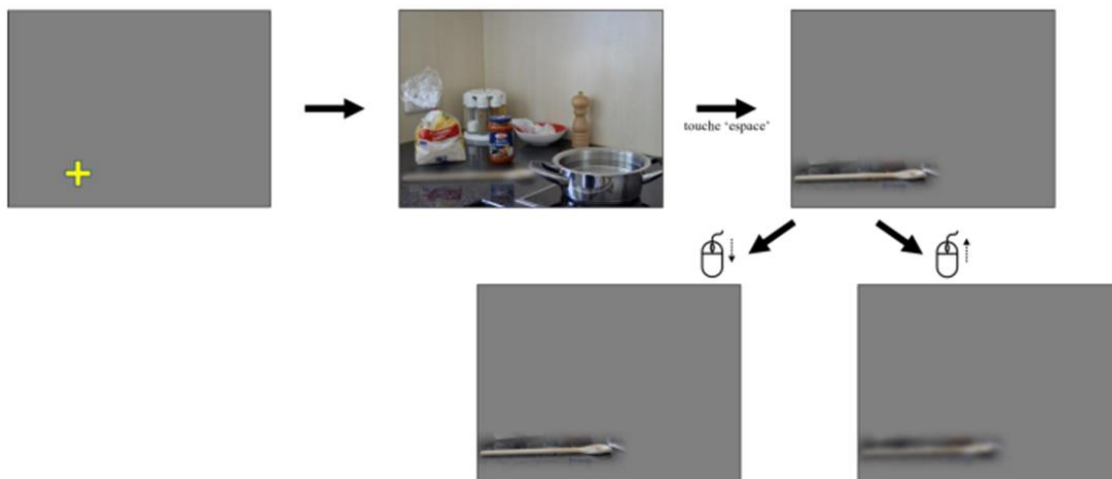
Notre paradigme d'ajustement perceptif présente également des limites en ce qui concerne la manipulation de la validité des prédictions et les stimuli utilisés dans l'Expérience 5. Effectivement, notre paradigme requiert que les objets à ajuster soient identiques dans les stimuli sample et target. Par conséquent, la manipulation de la validité des prédictions sur la base de l'information contextuelle ne pouvait se faire qu'en rajoutant artificiellement un contexte congruent ou incongruent autour de ces objets, ce qui donne des stimuli artificiels assez peu écologiques. Il pourrait donc être pertinent de conforter nos résultats par des études impliquant des stimuli plus naturels, proches de ce que nous pouvons rencontrer au quotidien comme par exemple les stimuli de la base d'images SCEGRAM (Figure 1.5 ; Öhlschläger & Vö, 2017), utilisés dans l'étude de Spaak et al. (2022). En revanche, avec un tel type d'image notre paradigme n'est plus adapté. En effet, comme il s'agit de photographies réelles d'objets congruents (e.g., une tasse dans un lave-vaisselle) ou incongruents avec leur contexte (e.g., une tasse dans un porte-papier toilette), un même objet n'est jamais vu selon les mêmes conditions dans chaque contexte (e.g., taille, angle de vue et position dans la scène), ce qui pourrait induire une différence de perception entre deux objets similaires en fonction de leur contexte en utilisant un paradigme d'ajustement perceptif tel que celui que nous avons utilisé. Dans ce cas, le paradigme de reproduction de percept de Williams et al. (2022) pourrait encore une fois être pertinent puisqu'il ne requiert la présentation que d'un seul stimulus à chaque essai, suivi de la reproduction du percept de ce même stimulus. Dans ce contexte, nous avons récemment débuté une expérience basée sur ce paradigme dans laquelle nous présentons aux participants une scène de la base d'image SCEGRAM contenant un objet congruent ou incongruent avec son contexte dont la position est indiquée par une croix de fixation préalable (Figure 3.3). L'objet de la scène était soit très flouté (faible fiabilité), soit peu flouté (haute fiabilité). Les participants avaient pour consigne de bien regarder l'objet, puis d'appuyer sur une touche lorsqu'ils pensaient être en mesure de reproduire le flou de l'objet. L'objet apparaissait ensuite isolé du contexte à un niveau de flou aléatoire. Les participants devaient ajuster le flou de l'objet de façon à reproduire le flou de l'objet précédemment perçu en contexte. Les résultats préliminaires (N = 10) soutiennent en partie les résultats obtenus avec notre paradigme d'ajustement perceptif. Lorsque l'objet en contexte était peu fiable, les participants avaient tendance à le reproduire comme plus net lorsqu'il était congruent qu'incongruent avec son contexte, en accord avec la prédominance du mécanisme de sharpening lorsque le signal est bruité. Cependant, lorsque l'objet en contexte

Chapitre 3 – Discussion générale et perspectives

était plus fiable, les participants avaient tendance à le reproduire plus fidèlement et de façon similaire qu'il soit congruent ou incongruent avec son contexte. Là encore, il est possible que l'absence d'effet dans cette dernière condition puisse s'expliquer par le fait que lorsque l'objet est net, les effets perceptifs des prédictions sont probablement plus subtils et difficiles à détecter. Cependant, la réplication de nos résultats dans la condition de haute fiabilité est encourageante quant à la pertinence de ce paradigme pour mesurer les effets des prédictions sur la perception de façon similaire à notre paradigme d'ajustement perceptif. Par conséquent, une future étudiante en doctorat (dans le contexte du projet EXPER, financé par l'ANR) continuera cette expérience pour analyser les résultats sur un plus grand nombre de participants. Pour pallier le problème de l'ajustement de la netteté, à savoir qu'un stimulus net ne peut pas être ajusté comme encore plus net, ce paradigme expérimental pourra être utilisé pour tester les effets des prédictions sur d'autres caractéristiques de bas niveau telles que le contraste ou encore la couleur.

Figure 3.3

Illustration du Déroulement de l'Expérience.



Note. Tâche d'ajustement du flou de l'objet. Chaque essai débutait par une croix de fixation indiquant l'emplacement de l'objet d'intérêt dans la scène, suivie de la présentation de la scène contenant l'objet. Les participants devaient bien regarder l'objet, puis appuyer sur la barre espace lorsqu'ils pensaient être en mesure de reproduire le flou de l'objet. L'objet apparaissait alors ensuite isolé du contexte à un niveau de flou aléatoire. Les participants devaient alors ajuster le flou de l'objet de façon à reproduire le flou de l'objet précédemment présenté en contexte.

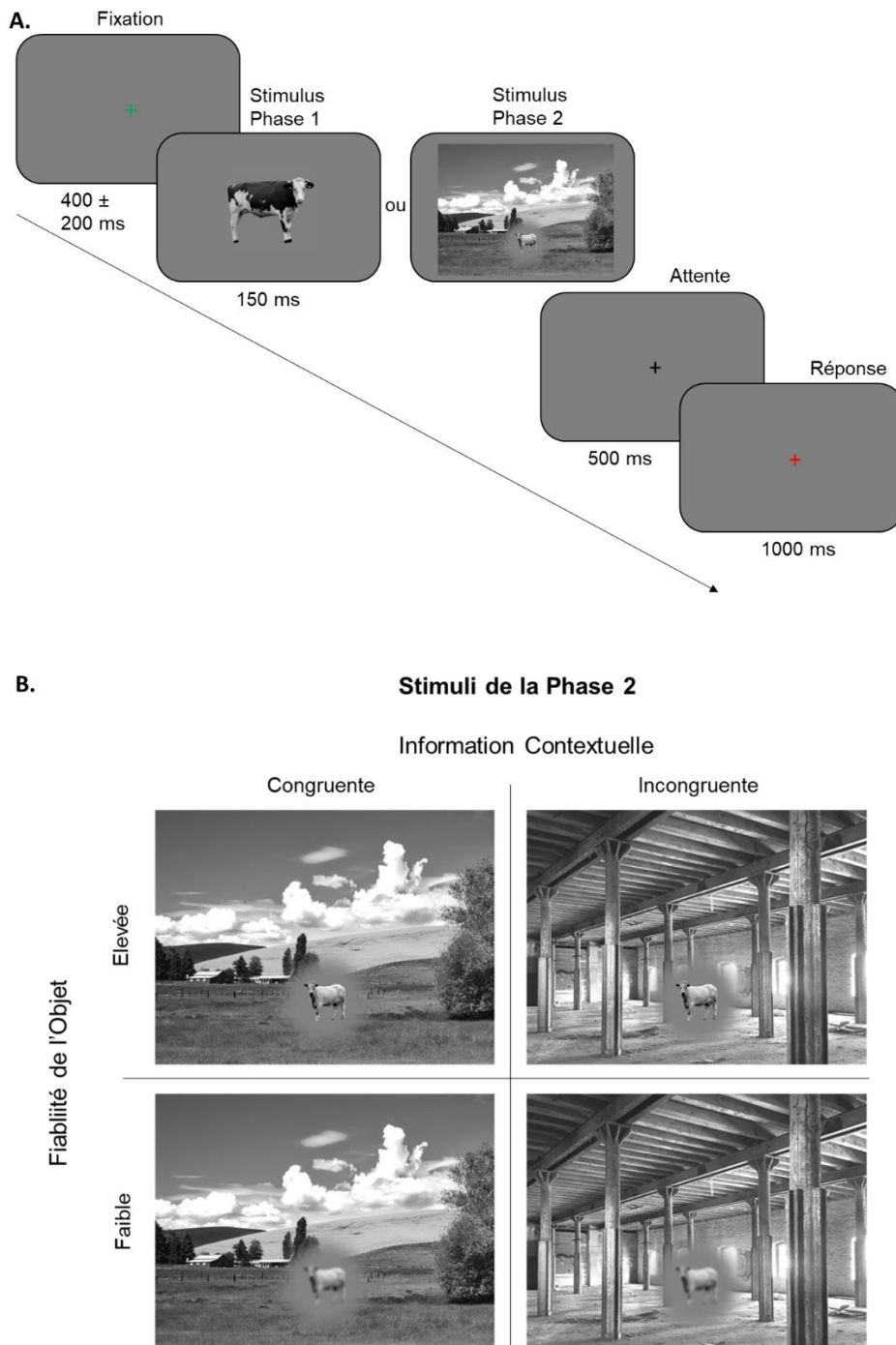
Si nos résultats soutiennent l'idée que l'effet de la validité des prédictions sur la perception visuelle subjective est modulé en fonction de la fiabilité du signal, dans l'Expérience 6, nous n'avons cependant pas réussi à valider la seconde hypothèse de l'opposing process theory (Press et al., 2020) selon laquelle l'effet de la validité des prédictions sur la perception serait également modulé selon le décours temporel du traitement de l'information visuelle. Dans cette dernière expérience, nous avons utilisé un nouveau paradigme inspiré de l'étude de Spaak et al. (2022) car le paradigme d'ajustement perceptif utilisé précédemment n'était pas adapté pour tester l'hypothèse des contraintes temporelles de l'opposing process theory. Les participants devaient donc discriminer parmi deux exemplaires (paradigme de choix forcé) d'un même objet (e.g., deux bateaux), celui qui avait été présenté précédemment dans un contexte de scène congruent ou incongruent. L'objet entouré de son contexte apparaissait soit pendant une durée très brève de 50 ms (condition expérimentale qui devrait favoriser le mécanisme de sharpening), soit pendant une durée plus longue de 200 ms (condition expérimentale qui devrait favoriser le mécanisme de dampening). Contrairement à nos hypothèses, les résultats n'ont pas mis en évidence de modulation de l'effet de la validité des prédictions selon le temps de présentation des stimuli, ce qui ne permet pas de conclure en faveur de la seconde hypothèse de l'opposing process theory sur les contraintes temporelles. Par ailleurs, le paradigme de choix forcé utilisé lors de l'Expérience 6 évaluait indirectement la qualité de l'encodage perceptif des objets en se basant sur le postulat que si les participants avaient des meilleures performances de discrimination des objets, cela indiquerait qu'ils les avaient perçus de meilleure qualité. Cependant notre paradigme ne permettait pas de mesurer directement la perception subjective (i.e., netteté perçue) que les participants avaient des objets. Encore une fois, le paradigme de reproduction de percept de Williams et al. (2022) pourrait se révéler intéressant pour préciser cette question et se rapprocher des mesures de notre paradigme d'ajustement perceptif. Dans ce contexte, nous pourrions par exemple présenter une image de scène contenant un objet congruent ou incongruent pendant un temps court ou long, puis présenter ensuite l'objet cible isolé de son contexte de scène et demander aux participants d'ajuster son niveau de flou pour qu'il corresponde au niveau de flou de l'objet cible présenté précédemment dans la scène. Dans le contexte de l'opposing process theory, on pourrait ainsi s'attendre à ce qu'un objet congruent soit reproduit comme plus net qu'un objet incongruent lorsque l'image cible est présentée brièvement, tandis que l'inverse devrait être observé lorsque l'image cible est présentée plus longtemps.

Chapitre 3 – Discussion générale et perspectives

Une autre manière d'étudier le traitement des stimuli attendus/inattendus au cours du temps serait de se rapprocher du signal cérébral. Dans ce contexte, nous avons débuté une expérience en EEG dans le but de préciser les mécanismes par lesquels les prédictions influencent le traitement au niveau cérébral de stimuli visuels attendus et inattendus au cours du temps, et en tenant compte de la fiabilité du signal visuel (Press et al., 2020). Pour cette étude, nous nous sommes inspirées de l'étude de Brandman et Peelen (2017) impliquant une méthode de décodage de patterns d'activité électrophysiologique. L'expérience comprenait deux phases (Figure 3.4). Durant la Phase 1, nous avons présenté aux participants des images d'objets animés (animaux) et inanimés (meubles) pendant que nous enregistrons leur activité cérébrale à l'aide d'un casque EEG. Ces mesures EEG ont été utilisées pour entraîner un algorithme à classifier l'activité cérébrale évoquée par les différentes catégories d'objets. Durant la Phase 2, nous avons également mesuré l'activité cérébrale des participants en EEG, mais nous avons cette fois-ci présenté des objets animés et inanimés (non utilisés dans la Phase 1) dans un contexte de scène (intérieur ou extérieur) congruent ou incongruent avec l'objet, conduisant à générer respectivement des prédictions valides ou non valides sur la catégorie de l'objet. Durant la Phase 2, nous avons également manipulé la fiabilité du signal visuel des objets en faisant varier leur niveau flou (stimuli nets vs. flous).

Figure 3.4

Illustration des Conditions Expérimentales et du Déroulement de l'Expérience EEG.



Note. A. Déroulement d'un essai pour les Phases 1 et 2. Chaque essai débutait par la présentation d'une croix de fixation verte dont la durée variait aléatoirement entre 200 ms et 600 ms, puis était suivi par la présentation d'un stimulus (objet ou meuble pour la Phase 1, objet ou meuble différents en contexte de scène pour la Phase 2) pendant 150 ms, d'une croix de fixation noire pendant 500 ms et enfin d'une croix de fixation rouge pendant 1 s. La tâche des participants consistait à appuyer, durant la présentation

Chapitre 3 – Discussion générale et perspectives

de la croix de fixation rouge, sur la barre d'espace si aucun objet n'était apparu à l'écran pendant l'essai. Cette tâche avait pour seul intérêt de s'assurer que les participants regardaient bien les stimuli présentés.

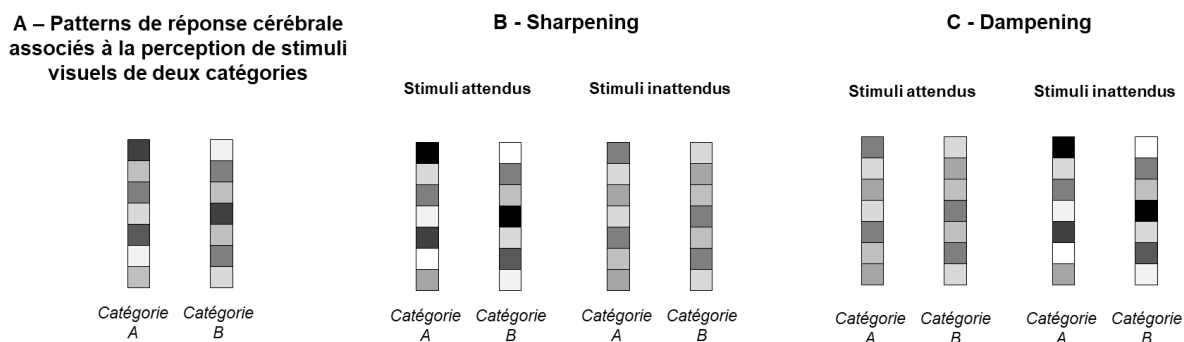
B. Exemple de stimuli utilisés dans la Phase 2. Les objets pouvaient apparaître à un niveau de fiabilité faible ou élevé et avec un contexte de scène congruent ou incongruent.

L'objectif était d'entraîner le classifieur sur les données de la Phase 1, puis de tester les performances du classifieur lors du décodage de la catégorie des objets sur la base des signaux EEG enregistrés lors de Phase 2 selon nos différentes conditions (objet attendu net ; objet attendu flou ; objet inattendu net ; objet inattendu flou) et selon le décours temporel du traitement de l'information visuelle. Le raisonnement était le suivant : si lors de la perception visuelle d'objets, l'activité cérébrale reflète essentiellement le mécanisme de sharpening (i.e., augmentation sélective de la sensibilité des neurones codant les caractéristiques prédites d'un stimulus et inhibition des neurones codant les caractéristiques non prédites de dampening), les patterns d'activité spatio-temporels devraient contenir des informations plus représentatives de la catégorie de l'objet lorsque celle-ci est conforme que non conforme aux prédictions. Le classifieur devrait donc décoder la catégorie d'objets attendus avec de meilleures performances que la catégorie des objets inattendus (Figure 3.5). Au contraire, si l'activité cérébrale reflète plutôt le mécanisme de dampening (i.e., inhibition des neurones codant les caractéristiques prédites), les patterns d'activité spatio-temporels devraient contenir des informations plus représentatives de la catégorie d'un objet lorsque celle-ci est non conforme que conforme aux prédictions. Le classifieur devrait donc décoder la catégorie d'objets inattendus avec de meilleures performances que la catégorie des objets attendus (Figure 3.5). De plus, si le mécanisme de sharpening précède le mécanisme de dampening, les performances de l'algorithme devraient être meilleures pour classifier dans la bonne catégorie l'activité cérébrale évoquée par les objets attendus qu'inattendus sur des fenêtres temporelles précoces des signaux EEG, tandis que les performances de l'algorithme devraient être meilleures pour classifier dans la bonne catégorie l'activité cérébrale évoquée par les objets inattendus qu'attendus sur des fenêtres temporelles plus tardives des signaux EEG. Enfin, si le mécanisme de sharpening prédomine lorsque les stimuli visuels sont peu fiables alors que le mécanisme de dampening prédomine lorsque les stimuli sont fiables et non ambigus, les performances de l'algorithme devraient être meilleures pour classifier dans la bonne catégorie l'activité cérébrale évoquée par les objets attendus qu'inattendus lorsque les stimuli sont flous, tandis que les performances de

l'algorithme devraient être meilleures pour classifier dans la bonne catégorie l'activité cérébrale évoquée par les objets inattendus qu'attendus lorsque les stimuli sont nets. Nous avons pour l'instant acquis les données de sept participants dans le cadre de cette étude. Les analyses préliminaires sur ces sept participants sont encore en cours et ne seront pas présentées dans ce manuscrit de thèse.

Figure 3.5

Illustration des Hypothèses Dérivées des Mécanismes de Sharpening et Dampening sur la Distinction de Deux Catégories d'Objets.



Note. A. Illustration des patterns d'activité cérébrale associés à la perception de stimuli visuels appartenant à deux catégories différentes (e.g., objets animés vs. inanimés). B. Illustration des hypothèses concernant les patterns d'activité cérébrale pour les stimuli attendus et inattendus sous l'hypothèse du mécanisme de sharpening. Les stimuli des catégories A et B devraient être différenciés avec plus de facilité lorsqu'ils sont attendus qu'inattendus. C. Illustration des hypothèses concernant les patterns d'activité cérébrale pour les stimuli attendus et inattendus sous l'hypothèse du mécanisme de dampening. Les stimuli des catégories A et B devraient être différenciés avec plus de facilité lorsqu'ils sont inattendus qu'attendus.

Enfin, un autre aspect qui n'a pas été testé dans le cadre de cette thèse est une autre hypothèse sous-jacente à l'opposing process theory (Press et al., 2020) qui est que l'influence des prédictions au niveau perceptif serait également modulée par la fiabilité de ces prédictions (le fait que les informations utilisées pour générer des prédictions soient de bonne qualité, comme un contexte de scène net, ou de mauvaise qualité, comme un contexte de scène brouillé ou flouté). Cet aspect a été évoqué pour discuter les résultats de l'Expérience 4. Pour rappel, les résultats de cette expérience indiquent que l'effet des prédictions sur la netteté perçue

augmente avec la diminution de la fiabilité du signal visuel mais au-delà d'un certain seuil de fiabilité (lorsque les scènes sont très floues), l'influence des prédictions sur la perception cesse d'augmenter. Nous avons proposé d'expliquer ces résultats par le fait que lorsque les scènes étaient trop floues (i.e., signal peu fiable), les prédictions seraient trop diverses et donc trop peu informatives pour permettre d'affiner davantage la représentation du stimulus. Pour tester cette hypothèse, il pourrait donc être intéressant de mettre en place une expérience d'ajustement perceptif manipulant plus explicitement la fiabilité des informations permettant de générer des prédictions. C'est justement ce qui a été fait par une étudiante de Master 2 que j'ai co-encadrée qui a répliqué l'Expérience 1 sur les objets en contexte mais en faisant varier la fiabilité du contexte de scène, par le biais d'un brouillage (i.e., scramble de phase) plus ou moins important du contexte entre les stimuli target et sample. Les résultats de cette expérience ont montré que l'effet des prédictions issues de l'information contextuelle sur la perception subjective de la netteté de l'objet était d'autant plus faible que le contexte était brouillé, mais ce uniquement lorsque les objets avaient un niveau de flou élevé et étaient donc plus difficilement identifiables et pas lorsqu'ils avaient un niveau de flou faible et étaient facilement reconnaissables. Comme proposé par l'opposing process theory, il semblerait donc que l'influence des prédictions diminue avec la diminution de leur fiabilité.

2. Perspectives théoriques et implications cliniques

2.1. Application des effets perceptifs des prédictions à la vision active et leurs conséquences pour la reconnaissance

La problématique de recherche de cette thèse impliquait l'utilisation de paradigmes et tâches très éloignés des tâches perceptives que nous réalisons au quotidien. En effet, la perception visuelle est un processus actif au cours duquel nous explorons sans cesse l'environnement via les mouvements oculaires. L'environnement est par ailleurs lui-même rarement statique puisque qu'il contient des objets en mouvement et se modifie constamment en fonction de nos propres déplacements. Dans ce contexte, comment les effets perceptifs liés aux connaissances et aux prédictions qui en découlent étudiés dans cette thèse pourraient participer à faciliter la perception visuelle en tenant compte de sa dimension active ?

Comme nous l'avons évoqué en Introduction, le recours aux connaissances a priori pourrait par exemple permettre d'améliorer la perception de stimuli rendus bruités ou ambigus du fait de différentes contraintes de l'environnement (e.g., faible luminosité, brouillard). On

Chapitre 3 – Discussion générale et perspectives

peut cependant rappeler qu'au-delà de contraintes extrinsèques liées à l'environnement, l'information visuelle peut également être bruitée ou ambiguë du fait de contraintes intrinsèques, liées aux propriétés du système visuel. Par exemple, l'inhomogénéité de la répartition des photorécepteurs et des cellules ganglionnaires sur la rétine implique une différence d'acuité visuelle entre la vision centrale et la vision périphérique (Wertheim, 1894). Alors que l'information en vision centrale apparaît nette et bien résolue, l'information en périphérie est moins précise et de faible résolution spatiale. Pourtant nous n'avons absolument pas conscience de ces différences et percevons tout notre champ visuel comme uniforme. Nous faisons l'hypothèse que les mécanismes prédictifs étudiés dans le cadre de cette thèse pourraient contribuer à expliquer ce phénomène. Sur la base des mécanismes proposés dans le cadre du modèle fréquentiel de catégorisation de scènes (Peyrin et al., 2021; Roux-sibilon, 2020), l'extraction rapide d'une information grossière issue de la vision périphérique pourrait non seulement permettre de générer des prédictions sur les objets et les détails présents en vision centrale, mais également dans tout le champ visuel. Ces prédictions pourraient alors permettre d'améliorer, grâce au mécanisme de sharpening, la perception de l'information conforme aux prédictions dans le champ visuel périphérique. Il est également possible que les informations de bonne résolution issues de la vision centrale soient aussi utilisées pour prédire l'information contextuelle en vision périphérique. Ainsi, nous pouvons proposer une extension de l'opposing process theory (Press et al., 2020) en ajoutant l'hypothèse d'un mécanisme de sharpening au niveau du champ visuel parafovéal et périphérique. En ce sens, des travaux récents menés par Brandman et Peelen (2019, 2023) utilisant des méthodes de classification des patterns d'activité cérébrale ont montré que le traitement des objets (qui se fait généralement en vision centrale) affine la représentation cérébrale du contexte des scènes (qui se fait généralement en vision périphérique) lorsque ce dernier est ambigu. Cependant, ces auteurs n'ont pas contrôlé si les objets et leurs contextes étaient vus en vision centrale et périphérique, respectivement. Il pourrait donc être intéressant de répliquer cette étude en contrôlant explicitement cela afin de déterminer si la perception d'objet en vision centrale affine la représentation cérébrale de l'information périphérique.

Le rehaussement perceptif des caractéristiques pouvant être prédites dans le champ visuel périphérique pourrait par ailleurs s'avérer utile lors de l'exploration visuelle. Par exemple, cela pourrait permettre de désambiguïser une partie de l'information périphérique et ainsi mieux identifier les points du champ visuel ne bénéficiant pas de cette amélioration et vers lesquels orienter le regard pour un traitement plus approfondi en vision centrale. En ce sens,

Chapitre 3 – Discussion générale et perspectives

une étude récente couplant des enregistrements EEG et oculométriques suggère que la détection d'objets incongruents avec le contexte d'une scène (qui ne bénéficieraient donc pas du mécanisme de sharpening) se produit dès la fixation précédente (i.e., lorsqu'ils apparaissent dans le champ visuel périphérique) et que ces objets sont priorisés et fixés plus longtemps lors de l'exploration relativement aux objets congruents (Coco et al., 2020). Par ailleurs, ce mécanisme de sharpening en vision périphérique pourrait également faciliter la perception post-saccadique de stimuli conformes aux prédictions, dont seules les caractéristiques non prédites seraient alors extraites une fois le stimulus amené en vision centrale (i.e., mécanisme de dampening). Cette hypothèse est supportée par une étude EEG récente (Buonocore et al., 2020) au cours de laquelle les participants devaient effectuer une saccade vers une cible périphérique (image de visage ou scramble d'une image de visage). Lorsque la saccade était initiée, soit le stimulus restait inchangé (une image de visage était vue en vision périphérique, puis en vision centrale après la saccade), soit il était remplacé par une autre image (un scramble de visage était vu en vision périphérique, puis une image de visage était présentée en vision centrale après la saccade). Ainsi, l'image apparaissant en vision centrale après la saccade pouvait être soit congruente, soit incongruente avec les prédictions élaborées sur la base de l'information périphérique pré-saccadique. Les résultats ont montré que l'amplitude de la composante N170 (connue pour être sélective à la perception de visages) post-saccadique était réduite lorsque les images pré- et post-saccadiques étaient congruentes (i.e., deux visages) qu'incongruentes (i.e., un scramble et un visage). Ces résultats sont donc compatibles avec l'idée que le traitement périphérique d'un stimulus favoriserait ensuite le mécanisme de dampening (i.e., atténuation du traitement des caractéristiques prédites - ici, le visage) une fois ce stimulus en vision centrale. Les résultats d'une autre étude utilisant un paradigme similaire (Edwards et al., 2018) suggèrent par ailleurs une facilitation du traitement post-saccadique de stimuli congruents avec l'information périphérique pré-saccadique, relativement aux stimuli incongruents. Il pourrait donc être intéressant d'adapter ce type de paradigme en utilisant une tâche perceptive à réaliser sur des objets en contexte de façon à vérifier comment le traitement de l'information pré-saccadique influence la perception subjective post-saccadique.

Les mécanismes prédictifs pourraient également contribuer à la perception stable et fluide de l'environnement visuel. En effet, en plus des prédictions générées sur la base d'informations extraites il y a longtemps (e.g., connaissances basées sur l'expérience antérieure) ou sur le moment (e.g., information contextuelle pendant la fixation d'un objet), les prédictions basées sur les informations extraites dans un passé récent (e.g., prédiction que

Chapitre 3 – Discussion générale et perspectives

l'environnement futur sera similaire voire identique à l'environnement présent) pourraient jouer un rôle important. Ainsi, la perception à un temps t serait déterminée par ce qui a été perçu à un temps $t-1$. Ce mécanisme a été appelé la dépendance sérielle et correspond à une modulation de la perception dans le sens des informations sensorielles récemment vues (Fischer & Whitney, 2014). Cela permettrait notamment de rendre la perception visuelle stable et cohérente malgré la variabilité du signal visuel. Par exemple, cela permettrait de continuer à percevoir les objets à la même position malgré toutes les saccades oculaires ou les mouvements de tête que nous effectuons et qui modifient donc la position des objets sur l'image rétinienne. Dans ce sens, une étude de Collins (2020) a mis en évidence que l'intégration des informations visuelles du présent et du passé très récent (i.e., la dépendance sérielle) impactait l'apparence des stimuli au niveau perceptif (voir également Collins, 2022). Dans cette étude, des patches de Gabor apparaissaient sur l'écran. Tous étaient des distracteurs présentés à des emplacements aléatoires sauf un gabor test qui restait au même emplacement à chaque essai. Dans certains essais (*oddball*), l'orientation du gabor test différait de celle des distracteurs. La tâche des participants consistait donc à indiquer s'ils avaient ou non perçus un *oddball* dans l'essai (i.e., si l'orientation du gabor test était différente ou non de celle des distracteurs). Les résultats ont mis en évidence que les performances de détection d'un *oddball* variaient selon l'orientation du gabor test dans l'essai précédent. Lorsque l'orientation du gabor test de l'essai précédent était plus proche de l'orientation des distracteurs dans l'essai actuel, les participants avaient tendance à moins détecter de *oddball*. Au contraire, lorsque l'orientation du gabor test de l'essai précédent était moins proche de l'orientation des distracteurs dans l'essai actuel, les participants avaient tendance à plus détecter de *oddball*. Cela indique donc que la perception de l'orientation du gabor test de l'essai actuel était biaisée dans le sens de l'orientation du gabor test de l'essai précédent. L'apparence subjective des stimuli (des gabors dans le cas de cette expérience) ne dépend donc pas seulement de leurs caractéristiques physiques présentes mais également de celles vues très récemment. Par ailleurs, la perception d'un stimulus est biaisée en faveur du stimulus vu précédemment seulement lorsqu'il est physiquement proche. Lorsqu'il est trop éloigné, sa perception n'est plus influencée par l'apparence du précédent stimulus (Collins, 2019). Ces résultats pourraient être interprétés dans le contexte de l'*opposing process theory* (Press et al., 2020). Lorsque la différence d'apparence entre les stimuli n'est pas fiable (i.e., lorsque l'apparence des stimuli est proche), la perception serait biaisée en faveur de ce qui est attendu sur la base de l'expérience précédente et donc, la perception du second stimulus serait biaisée dans le sens de l'apparence du premier stimulus. Au contraire, quand la différence entre

Chapitre 3 – Discussion générale et perspectives

les deux stimuli est fiable (i.e., lorsque l'apparence des stimuli est bien différente), la perception serait biaisée en faveur de ce qui est nouveau, c'est-à-dire en faveur des caractéristiques du second stimulus qui diffèrent de celles du premier stimulus.

Dans la continuité de ces travaux, il est également intéressant de se demander si les prédictions sont nécessaires à une bonne perception. Ainsi, il serait intéressant de déterminer si une perception plus nette des caractéristiques visuelles attendues explique les effets des prédictions dans la facilitation de la reconnaissance visuelle, par exemple, en termes de vitesse ou de précision de reconnaissance, comme le rapportent de nombreuses études comportementales (Brandman & Peelen, 2017; Davenport & Potter, 2004; Greene et al., 2015; Joubert et al., 2007, 2008; Palmer, 1975; Richter et al., 2018; Roux-Sibilon et al., 2019). Les stimuli attendus pourraient-ils être reconnus plus rapidement ou plus précisément parce qu'ils sont mieux perçus ? A notre connaissance, aucune étude n'a directement abordé cette question jusqu'à présent. D'autres expériences utilisant une tâche de reconnaissance explicite et des enregistrements de temps de réponse pourraient enrichir ce point. Dans le contexte de cette thèse, on pourrait, par exemple, émettre l'hypothèse qu'un stimulus prédictible flou serait reconnu plus rapidement qu'un stimulus non prédictible avec un niveau de flou similaire ou même inférieur. Les résultats d'études en IRMf soutiennent l'hypothèse selon laquelle une meilleure perception et de meilleures performances de reconnaissance sont corrélées. Par exemple, dans l'étude de Kok, Jehee, et al. (2012), des représentations plus nettes des stimuli dans le cortex visuel, basées sur les prédictions, ont été associées à des performances comportementales plus élevées (bien qu'aucun lien de causalité n'ait pu être établi sur la base d'analyses corrélationnelles). Une étude récente de Gandolfo et Downing (2019) s'est directement intéressée à la relation causale entre les prédictions et les performances de reconnaissance visuelle, ainsi qu'à l'activité corticale associée. Dans une première phase, les auteurs ont demandé aux participants de catégoriser des images de corps (forts vs. minces) et de scènes (à l'endroit vs. à l'envers). Ces stimuli étaient précédés d'un indice valide ou invalide qui était orthogonal à la tâche de catégorisation mais qui pouvait apporter des informations utiles pour faciliter la catégorisation. Par exemple, l'indice pour une image corporelle pouvait être la lettre « m » (pour masculin) ou la lettre « f » (pour féminin) implicitement associée à un corps fort ou mince, respectivement. Les résultats ont montré qu'un indice valide améliorait les performances de catégorisation des corps et des scènes. Dans une seconde phase, pendant la présentation de l'indice, l'activité cérébrale dans les aires corticales sélectives du corps ou de la scène a été perturbée par stimulation magnétique transcrânienne (TMS). L'effet de facilitation

par un indice valide a disparu lorsque la TMS a été appliquée pendant la présentation de l'indice sur les aires sélectives à la perception des corps (extrastriate body area) pour la tâche de catégorisation des corps ou sur les aires sélectives à la perception de scènes (occipital place area) pour la tâche de catégorisation de scènes. Ces résultats indiquent que l'activité liée aux prédictions dans ces aires est directement liée à la performance de catégorisation. Dans l'ensemble, ces études soutiennent l'idée que l'expression des prédictions dans les aires visuelles facilite la reconnaissance visuelle. D'autres études impliquant à la fois des tâches de reconnaissance et de reports subjectifs sur l'apparence de stimuli en utilisant les mêmes stimuli seraient nécessaires pour déterminer si des performances de reconnaissance plus élevées pour des stimuli visuels attendus sont également associées à une netteté perçue accrue.

2.2. Implications cliniques

En considérant les effets perceptifs des prédictions sous des conditions de signal visuel bruité, les travaux de cette thèse peuvent également fournir des indications utiles, par exemple pour mieux identifier dans quelle mesure les déficits sensoriels peuvent être compensés - ou au contraire amplifiés - par ces mécanismes prédictifs lors du traitement de stimuli attendus et inattendus. Ceci est particulièrement important dans le contexte de nos sociétés vieillissantes dans lesquelles les déficiences visuelles sont répandues, que ce soit en raison de maladies ophtalmologiques (e.g., le glaucome, la dégénérescence maculaire liée à l'âge, la cataracte) ou du vieillissement normal (Owsley, 2011). Par exemple, il a été suggéré que la perte du champ visuel périphérique chez les patients atteints de glaucome altère leur capacité à générer des prédictions concernant les objets situés au centre de leur champ visuel (Trouilloud et al., 2023). Ces patients peuvent donc ne pas bénéficier des effets perceptifs des prédictions. Dans le contexte de l'opposing process theory (Press et al., 2020), cela pourrait se traduire par des difficultés à identifier les stimuli ambigus, mais aussi à traiter les stimuli inattendus. Il est donc essentiel de caractériser la façon dont les prédictions influencent la perception visuelle chez les individus sains pour mieux comprendre comment la vie quotidienne de ces patients peut être affectée au-delà de leur perte de champ visuel. Montrer que les prédictions influencent non seulement la facilité avec laquelle nous comprenons notre environnement visuel, mais aussi la qualité de notre vision, peut également fournir des pistes intéressantes pour le développement de méthodes de rééducation basées sur l'utilisation accrue des prédictions afin d'aider la perception visuelle chez les patients atteints de déficiences visuelles (e.g., par un entraînement

à l'utilisation d'informations provenant d'autres modalités sensorielles ou par l'imagerie visuelle de stimuli attendus, Dijkstra et al., 2022; Williams et al., 2022).

Les effets perceptifs des prédictions étudiés dans ce projet de thèse peuvent également constituer une mesure utile pour évaluer l'utilisation des connaissances préalables et des prédictions qui en découlent dans des populations et des troubles spécifiques. Par exemple, des troubles tels que les troubles du spectre autistique ou la schizophrénie ont été expliqués par une influence insuffisante ou excessive des prédictions sur la perception, respectivement (e.g., Pellicano & Burr, 2012; Powers et al., 2017; Sinha et al., 2014; Teufel et al., 2015; Van de Cruys et al., 2014). Dans le contexte de cette thèse, on pourrait par exemple s'attendre à ce que les participants autistes ne perçoivent pas nécessairement les stimuli prédictibles comme plus nets que les stimuli non prédictibles, et donc à ce qu'ils n'ajoutent pas plus de flou aux stimuli prédictibles que non prédictibles pour les percevoir comme aussi flous l'un que l'autre. Au contraire, par rapport à des participants témoins neurotypiques, les patients schizophrènes pourraient percevoir les stimuli prédictibles comme encore plus nets que les stimuli non prédictibles. Ils devraient donc avoir besoin d'ajouter encore plus de flou que les participants témoins à un stimulus prédictible pour le percevoir aussi flou qu'un stimulus non prédictible. Il a également été démontré que les mécanismes prédictifs évoluent au cours de la vie. Par exemple, il a été proposé que l'accumulation d'expérience sensorielle avec le vieillissement puisse conduire à s'appuyer plus fortement sur les processus prédictifs, ce qui permettrait également de compenser les déficits sensoriels concomitants (Gilbert & Moran, 2016; Lai et al., 2020; Moran et al., 2014). Dans cette hypothèse, on peut s'attendre à ce que l'effet de la prédictibilité sur la netteté perçue des stimuli soit plus important chez les individus âgés que chez les jeunes.

3. Conclusion

Les études comportementales présentées dans ce travail de thèse nous ont permis de préciser les conséquences perceptives des mécanismes prédictifs impliqués dans la perception de stimuli complexes tels que les scènes et les objets. Nous avons tout d'abord montré que les prédictions influencent qualitativement la perception. Plus précisément, les scènes et objets flous sont subjectivement perçus comme plus nets lorsqu'ils peuvent être prédits sur la base d'informations contextuelles disponibles dans le pattern visuel que lorsque ces informations contextuelles sont absentes. Nous avons ensuite mis en évidence que les effets perceptifs des

Chapitre 3 – Discussion générale et perspectives

prédictions variaient en fonction de la qualité des signaux visuels : ces effets sont d'autant plus fort que les stimuli visuels sont bruités (i.e., flous). De plus, ces effets perceptifs sont également influencés par la validité des prédictions (prédictions correctes ou erronées). Nous avons observé que lorsque le signal visuel est bruité (très flou), des objets attendus (présentés dans un contexte congruent conduisant à des prédictions correctes) sont perçus comme plus nets que ces mêmes objets inattendus (présentés dans un contexte incongruent conduisant à des prédictions erronées). Par contre, lorsque le signal visuel était peu bruité (des objets relativement nets), les objets inattendus étaient cette fois perçus comme plus nets que les objets attendus. Nous avons également fait l'hypothèse que ces effets dépendaient de contraintes temporelles, mais nos résultats n'ont pas mis en évidence de modulation de l'effet des prédictions selon le temps de présentation des stimuli.

Dans l'ensemble, les résultats de ces travaux confortent l'idée que les prédictions basées sur l'information contextuelle et l'expérience antérieure influencent qualitativement la perception. Plus largement, ils apportent également un soutien au cadre théorique de la pénétrabilité cognitive de la perception (Lupyan, 2015; O'Callaghan et al., 2017) selon lequel la perception n'affecte pas seulement l'état cognitif ou les connaissances d'une personne mais qu'elle est également contrainte par ceux-ci. Nos résultats soutiennent également les théories du codage prédictif (Feldman & Friston, 2010; Friston, 2005) et viennent tester et conforter directement pour la première fois les hypothèses de l'opposing process theory (Press et al., 2020) qui postule une modulation de l'influence des prédictions selon la fiabilité du signal visuel et la validité des prédictions.

ANNEXE A – Article Rossel et al. (2022)

It Makes Sense, So I See It Better! Contextual Information About the Visual Environment Increases Its Perceived Sharpness

Pauline Rossel^{1, 2}, Carole Peyrin^{1, 2}, Alexia Roux-Sibilon³, and Louise Kauffmann^{1, 2}

¹ Université Grenoble Alpes, CNRS, LPNC, Grenoble, France

² Université Savoie Mont Blanc, CNRS, LPNC, Grenoble, France

³ Psychological Sciences Research Institute (IPSY), UC Louvain

Predictive coding theories of visual perception postulate that expectations based on prior knowledge modulate the processing of information by sharpening the representation of expected features of a stimulus in visual cortex but few studies directly investigated whether expectations qualitatively affect perception. Our study investigated the influence of expectations based on prior experience and contextual information on the perceived sharpness of objects and scenes. In Experiments 1 and 2, we used a perceptual matching task. Participants saw two blurred images depicting the same object or scene and had to adjust the blur level of the right image to match the blur level of the left one. We manipulated the availability of relevant information to form expectations about the image's content: one of the two images contained predictable information while the other one unpredictable. At an equal level of blur, predictable objects and scenes were perceived as sharper than unpredictable ones. Experiment 3 involving explicit sharpness judgments confirmed these results. Our findings support the sharpening account of predictive coding theories by showing that expectations increase the perceived sharpness of the visual signal. Expectations about the visual environment help us understand it more easily, but also makes us perceive it better.

Public Significance Statement

Many studies have shown that what we know or expect based on past experience and contextual information can influence how easily we can see and interpret the visual world: for example, a cow is recognized faster when presented in a field than a kitchen context or in isolation. Here we show that expectations also affect how we see. In three experiments, we found that blurred scenes and objects that could be related to prior knowledge (upright scenes and objects in predictable context) were perceived as sharper than the same blurred scenes and objects that could not (inverted scenes and objects in unpredictable contexts). Prior knowledge and expectations about the visual environment not only help us understand it more easily, but also makes us perceive it better.

Keywords: predictive coding, expectations, visual perception, perceptual matching, sharpening


Supplemental materials: <https://doi.org/10.1037/xhp0000993.supp>

Making sense of the visual world first necessitates solving a perceptual problem. The light casts shadows on objects, which often

overlap and can have an infinity of appearances depending on viewpoints or lighting conditions. Yet, we are able to recognize objects and scenes in a few hundred milliseconds despite their complexity and their variability (Kirchner & Thorpe, 2006; Rousselet et al., 2005; Thorpe et al., 1996). To deal with this problem and achieve recognition, we strongly rely on expectations based on prior experience and regularities learned from the environment. For instance, we always see objects in context in day-to-day experience. We can create associations between objects and their contexts leading to expectations or predictions about which objects could appear in a particular context or in which contexts a specific object is likely to be found (Bar, 2004, 2007; Oliva & Torralba, 2007). By restricting the possible interpretations of the stimulus we are perceiving, these expectations would facilitate the processing of visual information that fits them (Bar, 2004, 2007; Roux-Sibilon et al., 2019).

Many studies have shown that a predictable stimulus in a given context (e.g., a priest in a church) is recognized more accurately or

Carole Peyrin  <https://orcid.org/0000-0001-7792-092X>

Alexia Roux-Sibilon  <https://orcid.org/0000-0002-3810-0109>

The authors would like to thank all the participants who contributed to the study. We also thank Amaury Velasco for his help with data acquisition of Experiment 3. This study was not preregistered but stimuli, executable files of the experiment, data and analysis scripts are available in the Open Science Framework repository at <https://osf.io/d65zs/>. We have no known conflict of interest to disclose.

Correspondence concerning this article should be addressed to Louise Kauffmann, Université Grenoble Alpes, CNRS, LPNC, 1251 Avenue Centrale CS40700, Michel Dubois Building, 38058 Grenoble Cedex 9, France. Email: louise.kauffmann@univ-grenoble-alpes.fr

more quickly than a nonpredictable one (e.g., a priest on a football field; (Brandman & Peelen, 2017; Davenport, 2007; Davenport & Potter, 2004; Greene et al., 2015; Joubert et al., 2008; Lai et al., 2020; Palmer, 1975; Roux-Sibilon et al., 2019). In that sense, expectations about the visual environment influence *how easily* we perceive it. Expectations would be particularly useful to enable the recognition of noisy visual stimuli, for which the mere analysis of the input's physical attributes may be inefficient. For example, two blurry light points on a foggy road can only be interpreted as a car coming from the opposite direction if one relies on prior knowledge and learned associations about the objects that can be found in such a context. Expectations based on contextual information can even influence *what we perceive* when the visual signal is ambiguous. For example, a dark elliptical shape appearing in a blurry street context can be either interpreted as a car or as a pedestrian depending on its orientation (Oliva & Torralba, 2007; Torralba, 2003; see also Bar, 2004; Roux-Sibilon et al., 2019). Current models of visual perception consider visual processing as a proactive process, where conscious perception is not only determined by the characteristics of sensory inputs, but also strongly depends on expectations about them (Bar, 2007; Clark, 2013; de Lange et al., 2018; Friston, 2005; Summerfield & Egner, 2009; Teufel & Fletcher, 2020). This assumption more broadly falls within the cognitive penetrability of perception theoretical framework, according to which perception influences cognition, but is also constrained by it (Lupyan, 2015; O'Callaghan et al., 2017).

Yet, evidence that expectations qualitatively and subjectively affect perception is still scarce. Indeed, past studies investigating the effects of expectations mainly focused on implicit measures such as reaction times during recognition tasks that may not be sufficient to address subjective perceptual effects. For example, the fact that an expected stimulus is recognized faster than an unexpected one suggests an influence of expectations on processing speed but it does not necessarily follow that this is also associated with a qualitative change in perception. Similarly, higher recognition rate of an expected compared with an unexpected stimulus could be achieved by guessing the identity of the expected stimulus and does not necessarily reflect its enhanced perception. Behavioral tasks involving subjective perceptual reports are needed to address whether expectations influence subjective perception.

Another important question is how expectations do so. In this respect, the predictive coding theoretical framework (Rao & Ballard, 1999; see also Friston, 2005; Spratling, 2017; for reviews see de Lange et al., 2018; Kok & de Lange, 2015) provides useful insights. This model assumes that perception results from a continuous exchange of *Prediction* and *Prediction Error* signals between adjacent levels of the cortical hierarchy. *Prediction* signals are transmitted via descending connections and represent the current hypotheses that best explain the input one area receives. *Prediction error* signals carry the mismatch between these predictions and the actual input information. They are sent via ascending connections to the next upper cortical area allowing to adjust the predictions so that they better fit the input. Moreover, the relative weight of *Prediction* and *Prediction error* signals would vary according to the reliability of the visual signal: *Prediction* signals would weight more when the stimulus is noisy or ambiguous while *Prediction error* signals would weight more when the stimulus is unambiguous (e.g., Feldman & Friston, 2010; Press et al., 2020).

Different theories have been proposed regarding the neural implementation of predictive coding. The dampening account (also referred to as the Cancellation theory) posits that prediction signals are used to filter out or explain-away the predicted features of a stimulus, by inhibiting neurons tuned to these features, resulting in a relatively higher activity of neurons tuned to the unexpected features (e.g., Blakemore et al., 1998; Murray et al., 2004). Alternatively, the sharpening account (also referred to as the Bayesian theory) assumes that prediction signals increase the sensitivity of neurons tuned to the expected features while suppressing the response of neurons tuned to unexpected ones, resulting in a sharper cortical representation of expected stimuli (e.g., Kaiser et al., 2019; Lee & Mumford, 2003). To date, results from neuroimaging experiments have provided contradictory findings arguing in favor of both accounts (see de Lange et al., 2018, for a review). For example, Kok et al. (2012) presented participants with gratings whose orientation could be predicted or not based on auditory cues in an functional magnetic resonance imaging (fMRI) experiment. By examining the pattern of activity (rather than the mean activation) in the primary visual cortex using multivoxel pattern analyses, they found that the orientation of the grating could be decoded with higher accuracy from V1's activity pattern when it was expected than unexpected. These results supported the view that expectations sharpen representations in the visual cortex. Using a similar decoding method in the macaque inferotemporal cortex, Kumar et al. (2017) found the opposite. Expected stimuli were decoded less accurately than unexpected ones, arguing in favor of the dampening account. It has also been proposed that both mechanisms can coexist in separate neural populations or take place at different times over the course of perception (de Lange et al., 2018; Friston, 2005; Press et al., 2020).

While perception may not truly reflect the neural computations occurring in sensory cortices, hypotheses can be derived from these two mechanisms about the perceptual consequences of expectations. The dampening account should result in increasing the sensitivity to unexpected features that should be better detected than expected ones. On the contrary, the sharpening account should result in a higher sensitivity to expected than unexpected features. Recent behavioral studies tend to support the latter hypothesis (Ailović et al., 2021; Cheadle et al., 2015; Han & VanRullen, 2016; Lupyan, 2017; Perez et al., 2020; Stein & Peelen, 2015; Teufel et al., 2018; Wyart et al., 2012) by showing that expectations, or prior knowledge about a visual stimulus (e.g., familiar objects or meaningful shapes), increase the sensitivity to features matching these expectations by enhancing their detection. For example, Teufel et al. (2018) asked participants to detect the contrast or orientation of Gabor patches embedded in meaningful two-tone ("Mooney") images. Participants were then shown the original (meaningful) images used to build the two-tone stimuli and completed the task again. The authors observed that semantic knowledge about the images improved orientation and contrast detection during the second session, only when the border of Gabor patches matched the location of contours in the original images that were no longer present in the two-tone images. These results indicated that, even if irrelevant to the task, object knowledge enhances the detection of consistent low-level features. However, only a few studies used explicit perceptual reports to directly address whether expected stimuli actually *look* sharper (Han & VanRullen, 2016; Lupyan, 2017; Perez et al., 2020). Of particular

interest, a study by Lupyan (2017) showed that prior semantic knowledge has a direct impact on perceived sharpness of the visual signal. In this experiment, participants completed a perceptual matching task in which they were presented with two blurred letter strings, one being the Target and the other one the Sample. Their task was to adjust the blur level of the Sample to match the blur level of the Target. To address the role of prior knowledge on the perceived sharpness of stimuli, the meaningfulness of the Target/Sample letter strings was manipulated so that one of the strings was a meaningful word (prior semantic knowledge, e.g., “much”) and the other one a meaningless pseudoword obtained via shuffling the order of the meaningful word’s letters (no prior semantic knowledge, e.g., “mchu”). The Target could be a word and the Sample a pseudoword or the Target could be a pseudoword and the Sample a word. Results showed that participants added more blur than necessary to match a Sample word to a Target pseudoword and not enough blur to match a Sample pseudoword to a Target word. This indicated that at an objectively equal level of blur, words were subjectively perceived to be sharper than pseudowords. These results suggested that expectations based on prior semantic knowledge influenced the processing of blurred meaningful words by making them look sharper, and that expectations also affect *how* we perceive.

Such a mechanism may be particularly useful for the processing of visual stimuli such as words requiring a detailed processing, and therefore, a sharp percept. However, does it also apply to the processing of visual stimuli such as objects in context or even whole scenes that need less details to be identified (Greene & Oliva, 2009; Kauffmann et al., 2014; Schyns & Oliva, 1994)? This question is crucial to make it a more general principle of visual perception. The aim of the present study was to determine whether prior knowledge about our visual environment can affect the appearance of objects and scenes.

To this end, we used a perceptual matching task inspired by Lupyan’s paradigm (Lupyan, 2017). This task appears particularly relevant to address the consequences of expectations on subjective perception because it allows to directly compare participants’ own perceptual judgments to the known “true state” of the signal. Any resulting difference is more likely to reflect qualitative perceptual effects, in contrast to measures such as reaction times or recognition accuracy. In Experiments 1 and 2, two blurred images depicting a scene, a Target and a Sample, were displayed side by side on the screen. Participants had to adjust the blur level of the Sample to match the blur level of the Target. As in Lupyan (2017), we manipulated the availability of prior experience or relevant information to form expectations about the image content. In Experiment 1, we addressed the effect of expectations on the perceived sharpness of an object by manipulating its scene context. The object was either surrounded by its intact scene context (predictable object) or by a phase-scramble of its scene context (unpredictable object). The Sample and the Target always contained the same object but had a different context. In Experiment 2, we addressed the effect of expectations on the perceived sharpness of a whole scene by manipulating the predictability of scene layout. The scene was either presented upright (predictable scene layout) or upside-down (unpredictable scene layout). The Sample and the Target were images of the same scene but with a different orientation. Based on Lupyan’s results (Lupyan, 2017), we expected stimuli to be perceived as sharper in predictable than unpredictable

conditions. Therefore, at an objectively equal blur level, predictable stimuli should be subjectively perceived to be sharper than unpredictable ones in both experiments. As a consequence, participants should match the blur level of the Target and the Sample stimuli by attributing to a predictable Sample matched to an unpredictable Target a higher blur level than to an unpredictable Sample matched to a predictable Target.

Experiment 1

Method

Participants and Design

Thirty-two undergraduate students of Psychology from University Grenoble Alpes (28 women, $M_{\text{age}} = 22.38$, $SD_{\text{age}} = 4.84$) with normal or corrected to normal vision participated in the experiment. The sample size was set based on an estimated effect size of $d_z = .62$ from Experiment 1B of Lupyan (2017) to achieve a power (i.e., corresponding to $1 - \beta$, with β being the risk of Type-2 error) of .8 with an α level of .05. To estimate this effect size, we calculated the Cohen’s d_z from the t statistic of the interaction between the predictability of the Target and the type of trial. Then we calculated the confidence interval from this Cohen’s d_z and we selected the lower bound. The study was approved by the ethics committee of the University Grenoble Alpes (CER-Grenoble Alpes, COMUE University Grenoble Alpes, CERGA-Avis-2020-4) and was conducted in agreement with the Declaration of Helsinki for experiments involving human subjects. All participants involved in the study gave their informed consent before taking part in the experiment. They received course credits for their participation.

Stimuli

The stimuli were constructed from 20 colored photographs of real-world scenes (10 indoor scenes and 10 outdoor scenes) downloaded from the Pixabay website (<https://pixabay.com/fr/>), a photo sharing site under Creative Commons Zero license. The stimuli were created using MATLAB R2019b (The Mathworks, Natick, MA). Original images were cropped to a square format and resized to 350×350 pixels. They were converted to 256-level gray-scale images by averaging the values from the three color channels for each pixel and the mean luminance of all images was equalized to .5 (for pixel intensity values between 0 and 1).

Each scene contained an object semantically related to the scene category (e.g., a car for an outdoor scene, a lamp for an indoor scene) that was delineated by an ellipse encompassing 176×166 pixels of vertical and horizontal diameters. The object could be at different locations according to the scene image. For each scene, we created two versions: one in which the scene context around the object was left intact (predictable object based on contextual information, Intact condition) and one in which it was phase-scrambled and did not contain semantic information (unpredictable object based on contextual information, Scrambled condition). The phase-scrambled version of the scene was built by combining the amplitude spectrum of the scene with the phase spectrum of a white noise in Fourier space and the resulting image was obtained via inverse Fourier Transform.

The ellipse containing the object was then blurred using 2-D Gaussian filters (*imgaussfilt* Function in Matlab) of linearly increasing size (i.e., increasing standard deviation) to obtain 31 versions of the object (see Figure 1) ranging from a sharp object (close to intact; blur Level 1, $SD = 1$) to a very blurred and hardly recognizable one (blur Level 31, $SD = 10$). Blurring of stimuli had no effect on the mean luminance of stimuli which was set to .5 (for pixels intensities between 0 and 1) before applying the Gaussian filters. It only resulted in slightly reducing the root mean squared (RMS) contrast of stimuli (from .189 at blur Level 1 to .185 at blur Level 31). For each blur level, the ellipse containing the object was then added on the intact and scrambled scene backgrounds and was superimposed on a larger gray elliptical frame of average luminance (.5 for pixel intensity values between 0 and 1) and 30 pixels wide. This gray frame allowed us to isolate the object from the background and avoid local contrast differences between the object and the intact or scrambled backgrounds.

The integration of the object and gray frame on intact and scrambled backgrounds was done as follows. To integrate the gray elliptical frame to the image background, we first created for each stimulus a binary mask of the elliptical frame. This mask consisted of an image of 350×350 pixels (same resolution as the scenes) containing an ellipse of 236×226 pixels of vertical and horizontal diameters centered on the object location. Pixel values within the ellipse were set to 1 while pixel values outside of the ellipse were set to 0. This binary mask was then spatially filtered by a Gaussian with a standard deviation of 2 to smooth the border of the elliptical frame. The resulting image was multiplied with average luminance values of .5 to fill the ellipse with a uniform gray. We then computed the inverse of the smoothed binary mask of the ellipse to obtain an image with pixel values of 0 within the ellipse and 1 in the background. This inverse mask was multiplied with the intact or scrambled scene resulting in an image containing the scene background and pixel values set to 0 within the ellipse. Finally, this image was added to the image containing the gray elliptical frame. A similar procedure was used to blend the smaller ellipse containing the object with the background and the gray elliptical frame. We first created a binary mask of an ellipse of 176×166 pixels of vertical and horizontal diameters centered at the object location which was then spatially smoothed by a Gaussian with a standard deviation of 2. The resulting image was multiplied with the original scene to obtain an image containing only the object (pixel values outside of the ellipse were set to 0). We then

created the inverse of the smoothed binary mask and the resulting image was multiplied with the scene containing the gray elliptical frame. The resulting image was added to the image containing the elliptical object to obtain the final stimulus.

Procedure

The experiment was performed in the context of the coronavirus disease 2019 (COVID-19) lockdown periods in France and was conducted online. It was constructed using E-Prime 3.0 software (E-Prime Psychology Software Tools Inc., Pittsburgh, PA) and stimuli were displayed using an E-Prime Go 1.0 executable file on the participant's personal computer. Before starting the experiment, the experimenter met each participant via videoconference to explain the task and ensured that basic computer criteria were met to perform the experiment (e.g., an operating system supporting E-prime Go, screen resolution of 1920×1080 pixels ensuring the correct display of stimuli).

On each trial, two images containing the same object were displayed on the left and right of the participant's computer screen against a gray background (average luminance of .5). The object in the left image was the Target (T) and the object in the right image was the Sample (S). The Target object was presented with a randomly selected blur level between Level 6 and 25 that remained constant throughout the trial. The Sample object always had an initial blur level of 16 (median blur level between Levels 1 and 31) at the beginning of the trial. Participants were instructed to adjust the blur level of the Sample object to match the blur of the Target object (see Figure 2) by increasing or decreasing the blur of the Sample object in 1-level steps using the "P" and "O" keys on their keyboard, respectively. The Sample could be adjusted up to a maximum blur corresponding to Level 31 and a minimum blur corresponding to Level 1. Participants had to press the "Y" key to validate their response once they judged the Sample and Target objects as equally blurred. Participants were encouraged to rely on their subjective judgment to match the blur level of objects.

For each trial, the backgrounds of each object could be either intact (predictable object) or phase-scrambled (unpredictable object). Furthermore, the Target (T) and Sample (S) objects either had different (different Target/Sample pair-types) or the same background (identical Target/Sample pair-types) resulting in four conditions (see Figure 2A). The different Target/Sample pair-types allowed us to assess the effect of contextual information on the

Figure 1
Examples of an Object at Different Blur Levels Between Blur Levels 1 and 31 in the Intact and Scrambled Conditions

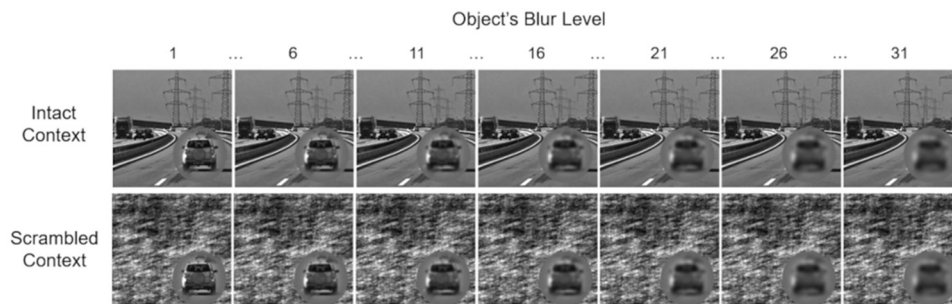
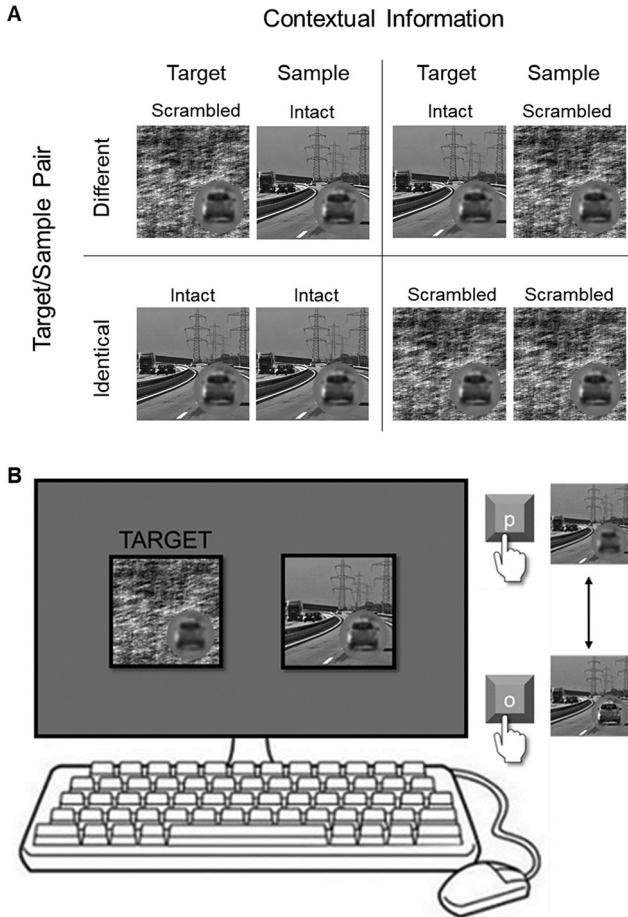


Figure 2
Illustration of the Experimental Conditions and the Display of Experiment 1



Note. (A) Example of stimuli presented in each experimental condition. For each trial, the Target and Sample objects were the same. The backgrounds could be either intact (predictable contextual information) or phase-scrambled (unpredictable contextual information). They either had a different scene background (different Target/Sample pair-types) or the same background (identical Target/Sample pair-types), resulting in four conditions: (a) both the Target and Sample were intact (“T-intact/S-intact”), (b) both the Target and Sample were scrambled (“T-scrambled/S-scrambled”), (c) the Target was intact and the Sample was scrambled (“T-intact/S-scrambled”), and (d) the Target was scrambled and the Sample was intact (“T-scrambled/S-intact”). (B) Illustration of the display in a trial with a scrambled Target and an intact Sample (different Target/Sample pair-types). Two scenes containing a blurred object appeared simultaneously, the Target on the left and the Sample on the right. Participants were asked to adjust the Sample object’s blur level, by pressing the O and P keys on the keyboard, until they judged both objects as equally blurred.

perceived sharpness of objects. The identical Target/Sample pair-types were used as control conditions to assess any general tendency in over- or underestimating the blur of the Sample relative to the Target object irrespective of its predictability, or to detect a more systematic bias in underestimating the blur of objects embedded in intact contexts irrespective of what they are being matched

to. The predictable and unpredictable conditions manipulated the presence and absence of semantic information to generate expectations about the object, but the context (either intact or scrambled) was never relevant to carry out the task that consisted in the blur level adjustment of the object. This clarification is important to dissociate the effects of expectations (i.e., what is likely to be present in the signal) from effects of attention (i.e., what is relevant for the task at hand) as attention has been shown to enhance perception (Rungratsameetaweemana & Serences, 2019; Summerfield & de Lange, 2014).

An experimental session contained 160 trials appearing in a random order (20 objects presented twice in each of the four conditions) with a break in approximately the first and second third of the experiment. For all conditions, each of the 20 Target objects appeared with a blur level varying from 6 to 25 (i.e., a total of 20 blur levels) so that each blur level was presented once per condition for a randomly selected object. The experiment lasted about 30–40 min, depending on the speed at which each participant made their judgment ($M \pm SD$ duration of a trial: 1858 ± 2133 ms). For each trial, we recorded the blur level attributed to the Sample to match the Target (Matched Blur Level, MBL). Before the experimental session, participants performed a training session (four trials) using stimuli that were not subsequently included in the main experiment. The stimuli and executable files of the experiment are available in the Open Science Framework repository: <https://osf.io/d65zs/>.

Data Analysis

Data were analyzed using R (R Core Team, 2019) and *lme4* package (Bates, Mächler, et al., 2015) to perform a linear mixed effects analysis of the Target/Sample pair-type (identical vs. different) and the Sample’s contextual information (intact vs. scrambled) in a 2×2 within-subjects design. Contrary to classical analyses of variance (ANOVA; Judd et al., 2012), mixed-effects analyses maximize the generalizability of our results to other participants, but also to other stimuli. We used the MBL as the dependent variable. Target/Sample pair-type, Sample’s contextual information and their interaction term were entered into the model as fixed effects (i.e., as the effects of the variables of interest). Intercepts for subjects and items, as well as subject-wise and item-wise random slopes for the effect of Target/Sample pair-type, Sample’s contextual information and their interaction were specified as random effects (i.e., as the effects of the variables to which we want to generalize our results).

We conducted the minimum number of analyses needed to test our hypotheses to avoid increasing the risk of Type-1 error. Based on Lupyan (2017), we expected that at an (objectively) equal level of blur, objects embedded in intact contexts would be subjectively perceived as sharper when compared with objects embedded in scrambled contexts. Therefore, for Target/Sample pairs of different contexts, objects in intact contexts should be blurrier than objects in scrambled contexts to be subjectively perceived as equally blurred. Consequently, for Target/Sample pairs of different contexts, the MBL of Sample objects in the T-scrambled/S-intact condition should be higher than the MBL of the Sample objects in the T-intact/S-scrambled condition. Alternatively, this effect could be explained by a systematic bias in underestimating the blur of objects embedded in intact contexts irrespective of the

This document is copyrighted by the American Psychological Association or one of its allied publishers. This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

context of the Target object. If the difference between MBL of Sample objects in intact and scrambled contexts was due to such a systematic bias, it should be as large regardless of the Target/Sample pair-type (different or identical). On the contrary, a larger difference in MBL for different than identical pair-types would lead to the conclusion that the difference in MBL in the different pair-types is unlikely to be explained by a bias based on the Sample's contextual information. To test these hypotheses, we tested the interaction between the contextual information of the Sample and the Target/Sample pair-type, and then the simple effect of contextual information of the Sample object for the Target/Sample pairs of different contexts.

We used the method proposed by Bates, Kliegl, et al. (2015) to construct parsimonious mixed models preventing convergence problems. Visual inspection of residual plots did not reveal any obvious deviations from homoscedasticity or normality. The p -values were obtained by Satterthwaite approximation with the *lmerTest* package (Kuznetsova et al., 2017). The significance threshold was set at .05. Effect sizes were estimated using Cohen's d_z (Lakens, 2013). Data and analysis scripts are available in the Open Science Framework repository: <https://osf.io/d65zs/>.

Results

Mean MBLs for each experimental condition are shown in Figure 3A. Given that the Target object could be displayed at 20 different blur levels ranging from 6 to 25, a perfect match between the Target and Sample objects would result in a mean MBL of $(6 + 25)/2 = 15.5$. A mean MBL with a value above 15.5 indicates an overestimation of the Sample blur relative to the Target, while a mean MBL below 15.5 indicates an underestimation of the Sample blur relative to the Target.

Results revealed neither a main effect of the Target/Sample pair-type, $t(18.99) = .21, p = .838, d_z = .037$ (different pair-type: $M = 15.40, SE = 1.02$; identical pair-type: $M = 15.35, SE = 1$), nor a main effect of the Sample's contextual information, $t(5.078) = 1.53, p = .127, d_z = .270$ (intact condition: $M = 15.47, SE = 1.03$; scrambled condition: $M = 15.28, SE = .99$). However, the interaction between the Target/Sample pair-type and Sample's contextual information was significant, $t(5.078) = 2.06, p = .039, d_z = .364$. In line with our hypotheses, the difference in MBL of Sample objects between intact and scrambled conditions was larger when the Sample and the Target were different (T-scrambled/S-intact: $M = 15.63, SE = 1.05$; T-intact/S-scrambled: $M = 15.17, SE = .98$) than when they were identical (T-intact/S-intact: $M = 15.31, SE = 1$; T-scrambled/S-scrambled: $M = 15.38, SE = 1$). Critically, when the Target and the Sample were different, the MBL was significantly higher for a Sample object in the intact than in the scrambled condition, $t(5079.36) = 2.53, p = .011, d_z = .448$. Overall, these results support our hypotheses. Participants added more blur to match a Sample object in an intact context to a Target object in a scrambled context than to match a Sample object in a scrambled context to a Target object in an intact context. This suggests that at an objectively equal blur level, objects in intact contexts were subjectively perceived as sharper than objects in scrambled contexts.

Figure 3B provides another representation of our results in which the MBL of the Sample object is plotted as a function of the actual blur level of the Target object for the Target/Sample pairs of different contexts only. The dashed line corresponds to a perfect

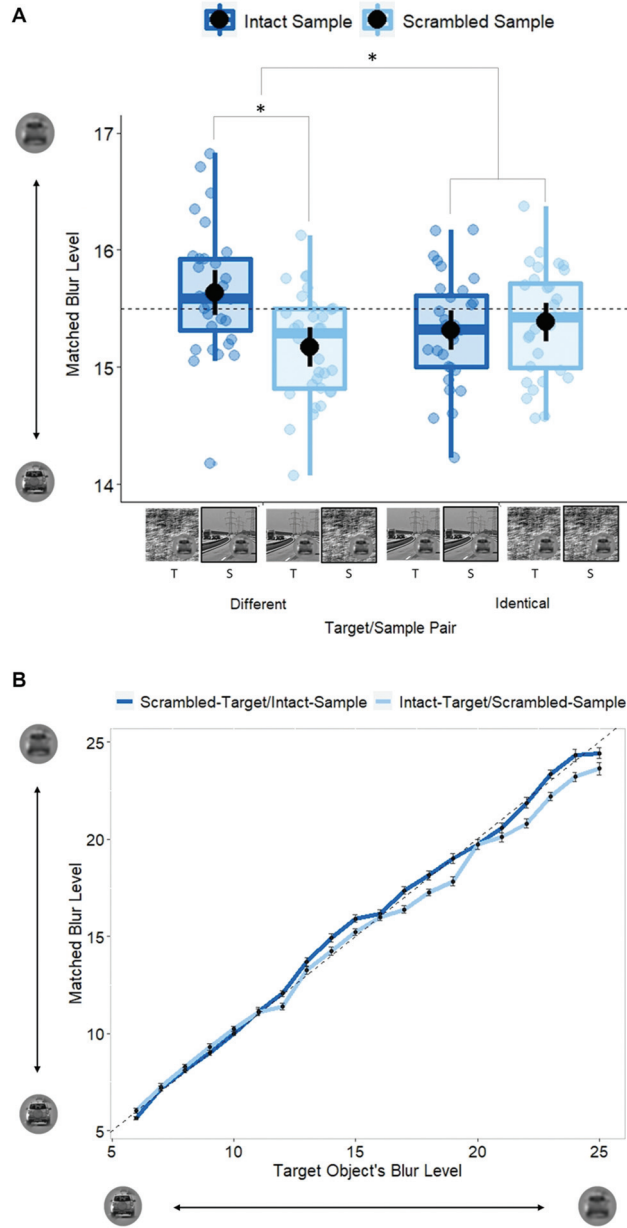
match between the blur of the Sample and Target object. Data points above or below this line indicate an over- or underestimation of the Sample's blur, respectively. Visual inspection of this graph suggests that the effect of contextual information on MBL varied according to the blur level of the Target object. This effect seems to be null for Target objects with a low blur level below 11 and to be the strongest for Target objects with a high blur level above 16. Because there was only a limited number of trials per blur level of the Target object (i.e., each Target object was not displayed at all blur levels throughout the experiment, see Method), the effect of blur level of the Target object was not included as an effect of interest in our model. To get an insight on the significance of this effect, we grouped together trials in which the blur of the Target object was low (below 16) and trials in which the blur of the Target object was high (above 16) and we performed a post hoc mixed-model analysis to examine whether the effect of contextual information could vary depending on the low versus high blur level of the Target object. We calculated as a fixed effect the interaction between the Sample's contextual information (intact vs. scrambled) for the Target/Sample pairs of different contexts only and the blur level of the Target object (low vs. high). The interaction was significant, $t(2378.01) = 3.15, p = .002, d_z = .557$. This interaction suggests that the Sample's contextual information effect was larger for a high blur level of the Target (above 16; T-scrambled/S-intact: $M = 20.98, SE = .57$; T-intact/S-scrambled: $M = 20.13, SE = .55$) than a low blur level (below 16; T-scrambled/S-intact: $M = 10.77, SE = .61$; T-intact/S-scrambled: $M = 10.63, SE = .56$).

Discussion of Experiment 1

The purpose of this first experiment was to determine whether expectations based on scene contextual information improves the perceived sharpness of objects. Participants performed a perceptual matching task in which they had to adjust the blur level of an object (Sample) embedded in a predictable intact or an unpredictable scrambled context to match the blur level of the same object (Target) embedded in a predictable intact or an unpredictable scrambled context.

In line with our hypotheses, results showed that when object pairs had a different context (different Target/Sample pair-type), participants matched the blur level of the Target and the Sample objects by attributing to the Sample objects with an intact context (matched to Target objects with a scrambled context) a higher blur level than to the Sample objects with a scrambled context (matched to Target objects with an intact context). This means that at an objectively equal level of blur, objects in an intact context were subjectively perceived as sharper than objects in a scrambled context. To control that this effect of contextual information in Target/Sample pairs with different contexts was not due to (a) a systematic bias in underestimating the blur of the Sample objects embedded in an intact context irrespective of the context of the Target object or (b) a systematic bias in underestimating the blur of the Sample relative to the Target object, we also included in our experiment Target/Sample pair-types with an identical context. We observed that the difference between the MBL for a Sample object embedded in an intact context and the MBL for the same Sample object embedded in a scrambled context was actually greater when the Sample and the Target object had a different than

Figure 3
 Mean Matched Blur Level for Each Condition of Experiment 1 and for the Different Target/Sample Pair-Types According to the Target Object's Blur Level



Note. (A) Mean Matched Blur Level (MBL) for intact and scrambled Sample object's contextual information according to the Target/Sample pair-type. A high blur level corresponds to a very blurry object whereas a low blur level corresponds to a relatively sharp object. The dashed line represents the perfect match between the blur of the Target (T) and Sample (S) objects. A mean matched blur with a value above or below 15.5 indicates an overestimation or underestimation of the Sample's blur relative to the Target, respectively. Black dots and error bars indicate mean and standard error over all individuals and trials, respectively. Color dots correspond to individual observations. (B) Mean MBL for T-scrambled/S-intact and T-intact/S-scrambled conditions according to the Target object's blur level. The dashed line represents the correct Sample object's blur level for each Target object's blur level. Error bars represent standard error over all individuals and trials. See the online article for the color version of this figure.

* $p < .05$.

an identical context. This indicates that the tendency to underestimate the blur of Sample objects in intact context was larger when they were compared with objects in scrambled context. This argues against an explanation of our results by a systematic bias and rather supports the idea that expectations based on contextual information result in increasing the perceived sharpness of objects. It can also be noted that mean MBL for the identical Target/Sample pairs was below 15.5 suggesting a general tendency to underestimate the level of blur required to match the Sample to the Target objects when they shared a similar context, specifically for very blurry objects with blur level above 16 (see online Supplemental Materials Figure S2a for MBLs according to the Target blur level in identical pair-types). This could be explained by a response bias in favor of a sharp perception of the Sample during the task whereby participants adapted the Sample to be too sharp than necessary rather than too blurry than necessary. Critically, such a bias strengthens our conclusions in the conditions where Target/Sample pairs were different. Participants added more blur than necessary to Sample objects in intact contexts *despite* a general tendency to make the Sample sharper than the Target.

Post hoc analyses suggested that the effect of expectations based on contextual information (i.e., effect of contextual information in the different pairs) was significantly larger when the Target object was very blurry (i.e., a blur level above 16) than when it was relatively sharp (i.e., blur level below 16). This result is also in line with predictive coding theories that assume that the influence of expectations on perception increases as signal precision decreases (Friston, 2005; Rao & Ballard, 1999).

Overall, results of this experiment are consistent with Lupyan's (2017) previous findings that expectations based on prior knowledge increases the perceived sharpness of the visual signal. Our results suggest that this sharpening mechanism not only applies to reading, which requires processing of details, but also to objects perception. Consequently, we wondered whether these findings could be extended more broadly to the perception of whole scenes that need less details to be identified. This would support the sharpening mechanism as a more general principle of vision. Furthermore, although the object was exactly the same in the Target and the Sample, we cannot exclude that the perceived greater sharpness for objects embedded in intact than scrambled contextual information results from a disruption of object processing in a

scrambled than intact context rather than to expectations based on contextual information. Experiment 2 aimed to confirm the present findings and extend them to whole scene perception, while better controlling for signal differences between the Target and Sample stimuli. We used the same perceptual matching task but this time, the Target and Sample stimuli had the exact same signal properties and only differed in terms of orientation: a whole predictable upright scene versus an unpredictable inverted scene.

Experiment 2

Method

Participants and Design

Thirty undergraduate students of Psychology from University Grenoble Alpes (25 women, $M_{\text{age}} = 21.43$, $SD_{\text{age}} = 4.07$) with normal or corrected to normal vision participated in the experiment. This experiment was conducted in parallel to Experiment 1 and the Sample size was estimated similarly as in Experiment 1. It was also performed within the same ethical framework as in Experiment 1.

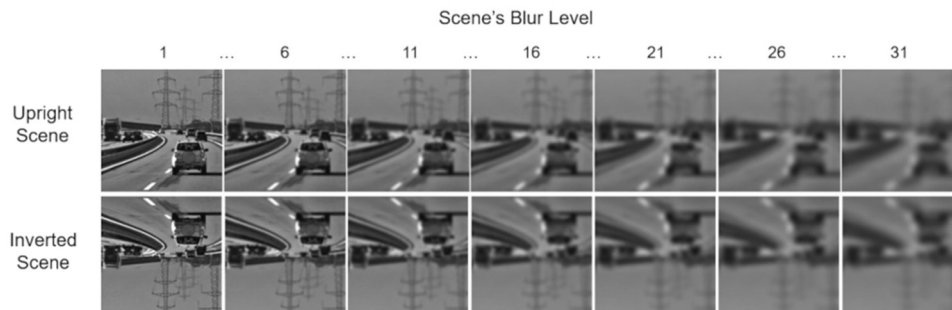
Stimuli

The stimuli were constructed from the same 20 photographs of indoor and outdoor scenes (350×350 pixels, greyscale, and with an equalized mean luminance of .5 as in Experiment 1, using MATLAB R2019b). For each scene, we created two versions in which the scenes were exactly the same but with a different orientation: one in which the scene was left upright (predictable scene layout - upright condition) and the other one in which the scene was vertically flipped around the horizontal axis (unpredictable scene layout - inverted condition). Upright and inverted versions of each scene were then blurred using the same procedure as in Experiment 1 for objects to obtain 31 versions of the image (see Figure 4) ranging from a sharp (close to intact) image (blur Level 1, $SD = 1$) to a very blurred and hardly recognizable one (blur Level 31, $SD = 10$).

Procedure

The experimental settings were similar to Experiment 1 (i.e., stimuli displayed using E-Prime Go 1.0 software on the participant's personal computer; stimuli and executable files are available

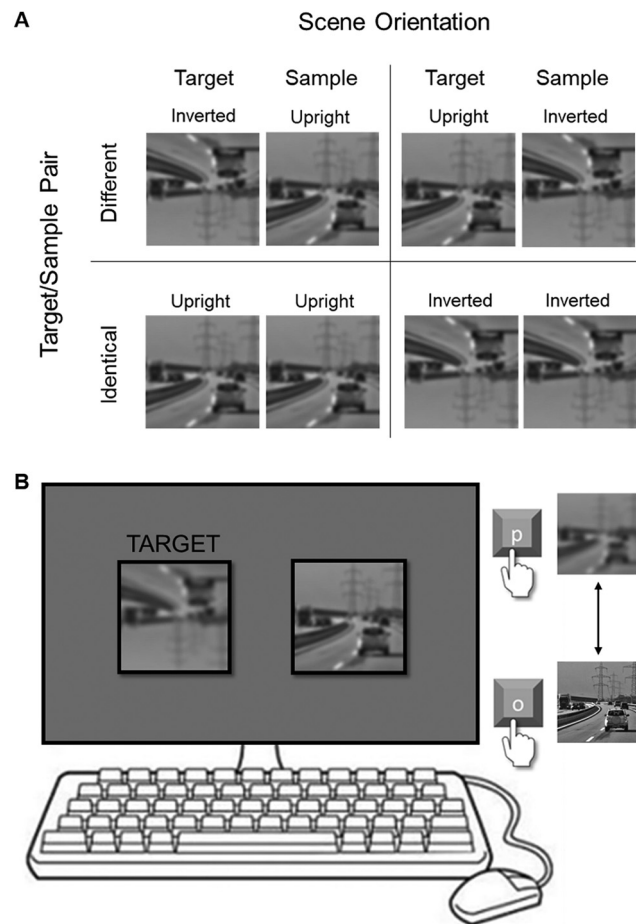
Figure 4
Examples of a Scene at Different Blur Levels Between Blur Levels 1 and 31 in the Upright and Inverted Conditions



in the Open Science Framework repository: <https://osf.io/d65zs/>. The procedure was also similar to Experiment 1 with the difference that the task consisted in matching the blur level of whole scenes instead of objects.

For each trial, the Target and the Sample scenes could be either upright (consistent with prior experience) or inverted (no prior experience). Furthermore, they either had the same (different Target/Sample pair-type) or a different orientation (identical Target/Sample pair-type), resulting in four conditions (see Figure 5A). As in

Figure 5
Illustration of the Experimental Conditions and the Display for Experiment 2



Note. (A) Example of stimuli presented in each experimental condition. For each trial, the Target and the Sample scenes could be either upright (consistent with prior experience) or inverted (no prior experience). They either had the same (different Target/Sample pair-type) or a different orientation (identical Target/Sample pair-type), resulting in four conditions: (a) both the Target and Sample scenes were upright, (b) both the Target and Sample scenes were inverted, (c) the Target scene was upright and the Sample scene was inverted, and (d) the Target scene was upright and the Sample scene was upright. (B) Illustration of the display in a trial with an inverted Target and an upright Sample (different Target/Sample pair-type). Two blurred scenes appeared simultaneously, the Target on the left and the Sample on the right. Participants were asked to adjust the Sample scene’s blur level, by pressing the O and P keys on the keyboard, until they judged both scenes as equally blurred.

Experiment 1, the scene orientation was not relevant to the task, which consisted in the blur level adjustment. Again, blurring of stimuli had no effect on the mean luminance of stimuli which was set to .5 (for pixels intensities between 0 and 1) before applying the Gaussian filters. Blurring of stimuli only resulted in slightly reducing the RMS contrast of stimuli (from .189 at blur Level 1 to .155 at blur Level 31).

An experimental session contained 160 trials appearing in a random order (20 scenes presented twice in each of the four conditions) with a break in approximately the first and second third of the experiment. For all conditions, each of the 20 Target scenes appeared with a blur level varying from 6 to 25 (i.e., a total of 20 blur levels) so that each blur level was presented once per condition for a randomly selected scene. The experiment lasted about 30–40 min, depending on the speed at which each participant made his or her judgment ($M \pm SD$ duration of a trial: 2134 ± 2192 ms). For each trial, we recorded the blur level attributed to the Sample to match the Target (MBL). Before the experimental session, participants performed a training session (four trials) using stimuli that were not subsequently included in the main experiment.

Data Analysis

Data analysis and software were similar to Experiment 1. We performed a linear mixed effects analysis of the Target/Sample pair-type (identical vs. different) and the Sample scene’s orientation (upright vs. inverted) in a 2×2 within-subjects design, with the MBL as a dependent variable. As in Experiment 1, Target/Sample pair-type, Sample scene’s orientation, and their interaction term were entered into the model as fixed effects. Intercepts for subjects and items, as well as subject-wise and item-wise random slopes for the effect of Target/Sample pair-type, Sample scene’s orientation and their interaction were specified as random effects.

Following the same rationale as in Experiment 1, we tested the interaction between the Sample scene’s orientation and the Target/Sample pair-type, as well as the simple effect of the Sample scene’s orientation for the Target/Sample pair of different orientations. We used the same method as in Experiment 1 to construct parsimonious mixed models preventing convergence problems and check for deviations from homoscedasticity or normality. The p -values were obtained by Satterthwaite approximation with the *lmerTest* package (Kuznetsova et al., 2017). Data from one participant who did not adjust the blur level of the Sample scene in any trial were discarded from the analysis. Data and analysis scripts are available in the Open Science Framework repository: <https://osf.io/d65zs/>.

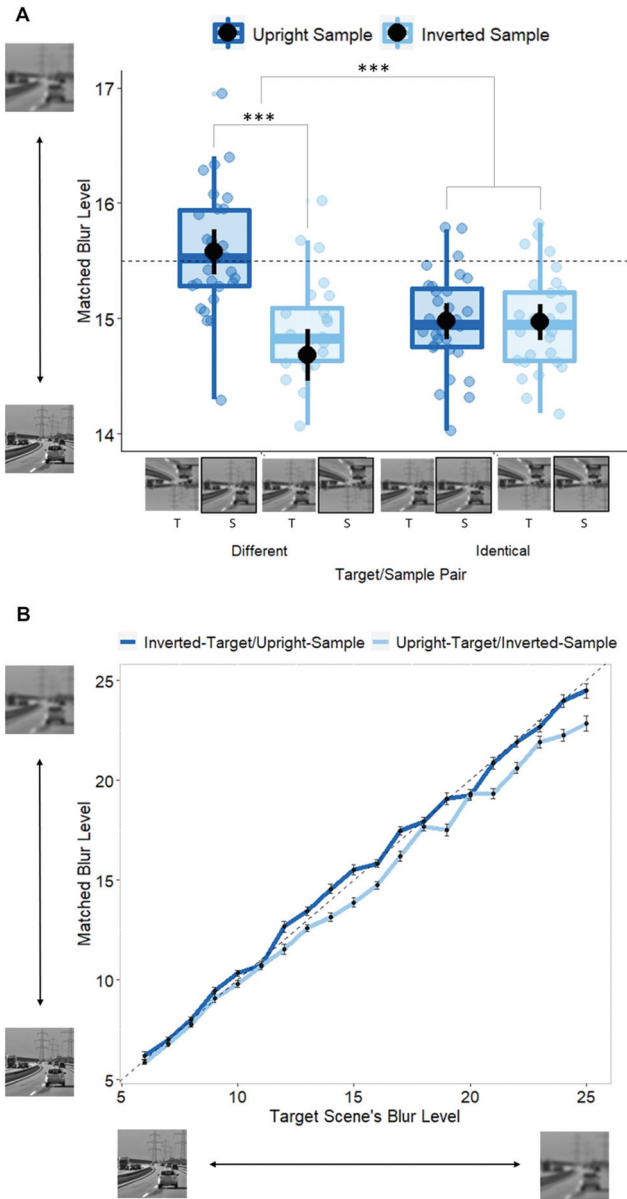
Results

Mean MBLs for each experimental condition are shown in Figure 6A. As in Experiment 1, a mean MBL of 15.5 corresponds to a perfect match between the Target and Sample scenes while a mean MBL with a value above/below 15.5 indicates an over/underestimation of the Sample blur relative to the Target.

Results revealed no main effect of the Target/Sample pair-type, $t(18.99) = .14, p = .892, d_z = .025$ (different Target/Sample pair-type: $M = 15.13, SE = 1.05$; identical Target/Sample pair-type: $M = 14.97, SE = 1.01$), but we observed a main effect of the Sample scene orientation, $t(4,758) = 3.63, p < .001, d_z = .662$, with

This document is copyrighted by the American Psychological Association or one of its allied publishers. This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Figure 6
Mean Matched Blur Level for Each Condition of Experiment 2 and for Different Target/Sample Pair-Types According to the Target Scene's Blur Level



Note. (A) Mean Matched Blur Level (MBL) for upright and inverted Sample scenes according to the Target/Sample pair-type. A high blur level corresponds to a very blurry scene whereas a low blur level corresponds to a relatively sharp scene. The dashed line represents the perfect match between the blur of the Target (T) and Sample (S) scenes. A mean matched blur with a value above or below 15.5 indicates an overestimation or underestimation of the Sample's blur relative to the Target, respectively. Black dots and error bars indicate mean and standard error over all individuals and trials, respectively. Color dots correspond to individual observations. (B) Mean MBL for T-inverted/S-upright and T-upright/S-inverted conditions according to the Target scene's blur level. The dashed line represents the correct Sample scene's blur level for each Target scene's blur level. Error bars represent standard error over all individuals and trials. See the online article for the color version of this figure.

*** $p < .001$.

participants attributing a higher MBL to the Sample scene when it was upright ($M = 15.28$, $SE = 1.05$) than inverted ($M = 14.83$, $SE = 1$). Moreover, the interaction between the Target/Sample pair-type and the Sample scene orientation was significant, $t(4,758) = 3.55$, $p < .001$, $d_z = .649$. In line with our hypotheses, the difference in MBL of Sample scenes between upright and inverted conditions was larger when the Sample and the Target scenes had a different orientation (T-inverted/S-upright: $M = 15.58$, $SE = 1.07$; T-upright/S-inverted: $M = 14.68$, $SE = 1.01$) than an identical orientation (T-upright/S-upright: $M = 14.98$, $SE = 1.02$; T-inverted/S-inverted: $M = 14.97$, $SE = .99$). Critically, when the Target and the Sample scenes had a different orientation, the MBL was significantly higher for a Sample scene in the upright than in the inverted condition, $t(4,758) = 5.08$, $p < .001$, $d_z = .927$. Overall, these results support our hypotheses. Participants added more blur to match an upright Sample scene to an inverted Target scene than to match an inverted Sample scene to an upright Target scene.

As in Experiment 1, Figure 6B provides another representation of our results in which the MBL of the Sample scene is plotted as a function of the blur level of the Target scene for Target/Sample pairs of different orientations only. The dashed line corresponds to a perfect match between the blur of the Sample and Target scene and data points above or below this line indicate an over- or under-estimation of the Sample's blur, respectively. Once again, visual inspection of this graph suggests that the effect of the Sample scene orientation on MBL varied according to the blur level of the Target scene. As in Experiment 1, we calculated in a post hoc analysis the interaction between the Sample scene's orientation for Target/Sample pairs of different orientations only and the blur level of the Target scene (low vs. high). However, the interaction was not significant, $t(2227.56) = 1.86$, $p = .063$, $d_z = .339$, suggesting that the Sample scene's orientation effect was not significantly larger for a high blur level of the Target (above 16; T-inverted/S-upright: $M = 20.86$, $SE = .59$; T-upright/S-inverted: $M = 19.74$, $SE = .56$) than a low blur level (below 16; T-inverted/S-upright: $M = 10.80$, $SE = .62$; T-upright/S-inverted: $M = 10.12$, $SE = .54$).

Discussion of Experiment 2

The purpose of this second experiment was to determine whether expectations based on prior experience increase the perceived sharpness of whole scenes. It also allowed us to ensure that the results from Experiment 1 could not be explained by visual signal differences between the Target and Sample images. We used the same perceptual matching task as in Experiment 1, in which participants had to adjust the blur level of a predictable upright or unpredictable inverted scene (Sample) to match the blur level of the same predictable upright or unpredictable inverted scene (Target).

We replicated the results of Experiment 1 by showing that when Target/Sample pairs had a different orientation (different Target/Sample pair-types), participants matched the blur level of the Target and the Sample scenes by attributing to the upright Sample scene a higher blur level than to the inverted Sample scene. This means that at an objectively equal level of blur, upright scenes would be subjectively perceived as sharper than inverted ones. As in Experiment 1, the significant interaction between the Sample scene orientation and the Target/Sample pair-type suggested that the effect of Sample scene orientation was larger in different than

identical pair-types, allowing to rule out systematic biases (tendency to underestimate the blur of the Sample relative to the Target; tendency to underestimate the blur of upright scenes irrespective of what they are being matched to) that could explain our results when the Target and the Sample had a different orientation. Moreover, as in Experiment 1, participants had a general tendency to underestimate the level of blur required to match the Sample to the Target scenes when they shared a similar context, specifically for very blurry scenes with a blur level above 16 (see online Supplemental Materials Figure S2b for MBLs according to the Target blur level in identical pair-types). Overall, results of this experiment confirm results of Experiment 1. These results further allow us to extend these findings to whole scene perception suggesting that the sharpening mechanism is a more general principle of vision that applies to many stimuli irrespective of the level of details they require to be identified.

The post hoc analysis on the effect of the Target blur level failed to replicate results of Experiment 1. Although the scene orientation effect when the Target and the Sample were different tended to increase with the blur level of the Target, this difference did not reach significance in the present experiment. As in Experiment 1, these exploratory analyses were performed to refine our interpretation of the results. However, Experiments 1 and 2 were not designed to test this hypothesis. In particular, to reduce the amount of trials and the duration of both experiments, each of the 20 Target stimuli used was only presented twice per condition as the Target, and they were associated with only two different blur levels. Thus, the blur level of the Target and the Target stimulus identity were not independent (i.e., a Target stimulus seen with a high level of blur was not necessarily seen with a low level of blur). Therefore, we cannot ensure that the observed tendencies are due to the signal precision (i.e., Target blur level) and not to stimulus identity. For this reason we conducted a control experiment (see Experiment 2bis in the online supplemental materials) that replicated Experiment 2 for the different Target/Sample pair-types in which all Target stimuli could be seen at four blur levels (ranging from low to high), so that the Target blur level was manipulated independently of the stimulus identity. The results of Experiment 2b is supported post hoc analyses of Experiments 1 and 2 by showing that the effect of expectations increased as the signal precision decreased. However, it increased up to a certain blur threshold where the influence of expectations on the perceived sharpness remained constant although the signal precision continued to decrease.

The perceptual matching task used in Experiments 1 and 2 allowed us to implicitly assess the relative perceived sharpness of predictable and unpredictable stimuli. The fact that participants made predictable stimuli blurrier than unpredictable ones to judge both stimuli as equally blurred led us to the assumption that at an equal level of blur, predictable stimuli were perceived as sharper than unpredictable ones. However, this was not explicitly tested in our experiments. We conducted a third experiment to confirm results of Experiments 1 and 2 using a task involving explicit judgments on the sharpness of visual stimuli. We presented participants with pairs of upright and inverted scenes used in Experiment 2 but this time participants were required to indicate whether they perceived a blur difference between the two scenes and if so, which of the two scenes was perceived as the blurriest or as the sharpest. Critically, the two scenes had the same objective blur level in most trials. In line with results of Experiments 1 and 2, we

expected participants to explicitly judge upright scenes as sharper than inverted scenes in these trials. In this experiment, we additionally recorded participants' eye movements to assess the extent to which predictable and unpredictable stimuli elicited differential amounts of visual attention as indexed by fixation duration. Indeed, visual attention has been found to enhance visual perception, for example, resulting in increased perceived contrast or spatial resolution (Carrasco et al., 2004; Yeshurun & Carrasco, 1998). It is possible that increased perceived sharpness for predictable stimuli can be explained by the fact that predictable stimuli elicited more visual attention than unpredictable ones.

Experiment 3

Method

Participants and Design

Thirty-five undergraduate students of Psychology from University Grenoble Alpes (31 women, $M_{\text{age}} = 19.63$, $SD_{\text{age}} = 2.06$) with normal or corrected to normal vision participated in the experiment. The sample size was set based on an estimated effect size of $d_z = .93$ from the observed effect of predictability in the different pair condition from Experiment 2 to achieve a power (i.e., $1 - \beta$, with β being the risk of Type-2 error) of .8 with an α level of .025 after Bonferroni correction for the two planned nonorthogonal analyses on test trials. The present experiment was performed within the same ethical framework as in Experiments 1 and 2.

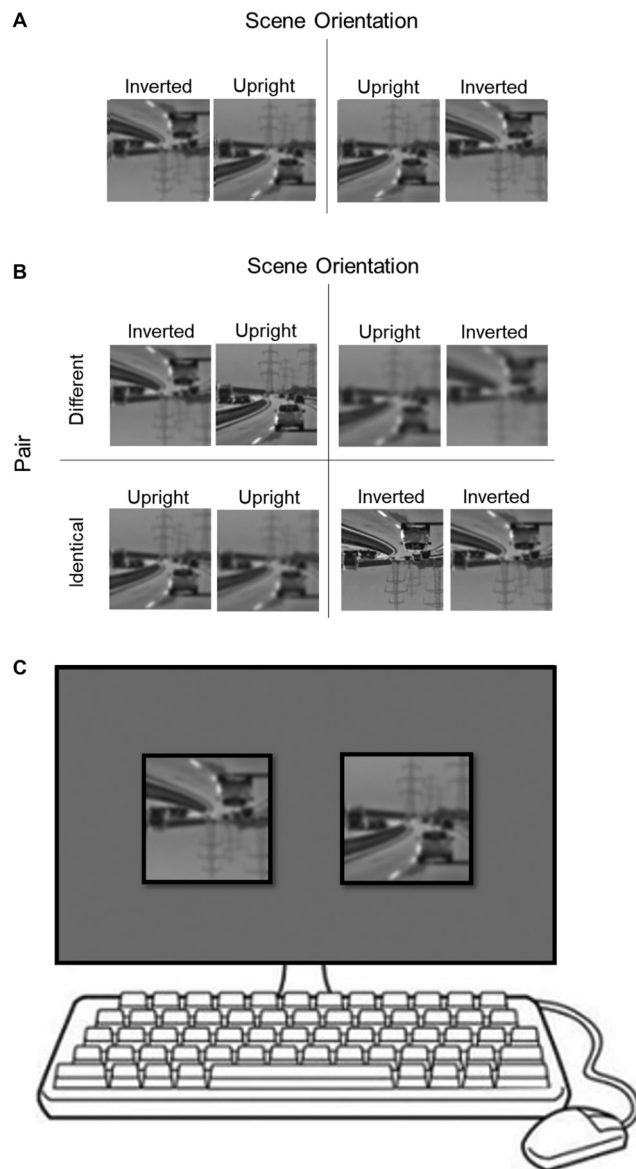
Stimuli and Procedure

The stimuli were exactly the same blurred upright and inverted scenes as those used in Experiment 2. Contrary to Experiments 1 and 2, this experiment was conducted after the COVID-19 lockdown periods in a laboratory allowing us to couple our behavioral measures to eye movement recordings. The experiment was constructed using the Psychtoolbox (Brainard, 1997; Pelli, 1997) implemented in MATLAB R2019b (MathWorks). Stimuli were displayed on a 30" monitor (DELL ULTRASHARP) with a resolution of 1920×1080 pixels (experimental script available in the Open Science Framework repository: <https://osf.io/d65zs/>). Participants put their head on a chin rest 70 cm away from the screen to keep their head fixed for the recording of eye movements. Eye movements were recorded using an Eyelink 1000 eye-tracker (SR Research, Ontario, Canada) with a sampling rate of 1,000 Hz and a nominal spatial resolution of $.01^\circ$ of visual angle. For each participant, we recorded only the left eye using the "pupil-corneal reflection" mode. The Eyelink software automatically detected saccades with the following thresholds: speed $>30^\circ/\text{s}$, acceleration $>8,000^\circ/\text{s}^2$, and saccadic displacement $>.15^\circ$. Fixations were detected when the pupil was visible and no saccade was in progress. Blinks were detected during partial or total occlusion of the pupil. Before starting the experiment and every 50 trials, we set up a calibration procedure in which participants had to orient their gaze toward five separate dots appearing sequentially at the center and at each edge of the screen. A drift correction was performed every 10 trials.

In this experiment, a trial began with a fixation cross appearing at the center of the screen. Then two versions of the same scene were displayed on the left and right of the screen on a gray

background (average luminance of .5) on each trial. The experimental session contained two types of trials: “test” and “catch” trials. In test trials (Figure 7A), the two scenes had the same blur level which could be of 5, 15, or 25 (i.e., low, medium, and high blur level). In this experiment, we only used three blur levels to

Figure 7
Illustration of the Experimental Conditions and the Display for Experiment 3



Note. (A) Example of stimuli presented in test trials. Both scenes of a pair always had a different orientation (one upright and one inverted scene), but the same blur level. (B) Example of stimuli presented in test trials. Both scenes of a pair had either the same or a different orientation, and they had different blur levels. (C) Illustration of the display. Two upright or inverted blurred scenes appeared simultaneously either with the same (test trials) or with different (catch trials) blur levels. Participants were asked to find which of the two scenes was the blurriest or the sharpest depending on their experimental group.

ensure that, contrary to Experiments 1 and 2, all stimuli would be seen at the same blur levels in each experimental condition while keeping the experiment relatively short. For test trials, the two scenes always had a different orientation (one upright predictable scene and one inverted unpredictable scene). The upright scene was displayed on the right and the inverted scene on the left on half of the test trials and the opposite was done in the other half. Participants were told that the blur difference between the scenes could be subtle and were instructed to find which of the two scenes was the blurriest or the sharpest (Figure 7C). It is possible that when participants have doubts about whether there is a difference in blur between the two scenes, they could be prone to systematically select the more familiar (the upright scene) one (Han & VanRullen, 2016). To cancel out any effect of such response bias, half of the participants were instructed to report the scene they judged as the blurriest while the other half had to report the scene they judged as the sharpest. They also had the option to answer that they did not perceive any blur difference between the two scenes. Test trials were the trials of interest allowing us to test our hypothesis that at an objectively equal level of blur, participants would perceive inverted scenes as blurrier than upright scenes. We also included catch trials (Figure 7B) to ensure that participants performed the task properly when the blur of the two scenes was objectively different and to prevent them from realizing that both scenes had often the same blur level. In these catch trials, the two scenes had different blur levels. One of the two scenes had a blur level of 5, 15, or 25. The other one had a blur of 3 to 7 levels above or below. The blurriest scene was on the right for half of the catch trials and on the left for the other half. Furthermore, for catch trials, the two scenes either had a different orientation (i.e., one upright predictable scene and one inverted unpredictable scene) or an identical orientation (i.e., both scenes were upright or inverted). Catch trials in which the two scenes had an identical orientation were added to prevent participants from adopting a strategy consisting of systematically selecting the upright or inverted scene. Given that Experiment 2 revealed that participants made upright scenes blurrier (by up to 1.73 blur levels on average) than inverted scenes to perceive both scenes as equally blurred, the minimal difference between the blur levels of the two scenes in catch trials was set to 3 to ensure that the two scenes would be perceived as having a different blur level. Participants had to give their response (Figure 7C) by pressing the “UpArrow” key if they did not perceive a difference between the two scenes, the “LeftArrow” key if they judged the left scene as the blurriest (or as the sharpest) and the “RightArrow” key if they judged the right scene as the blurriest (or as the sharpest).

The experimental session contained 120 test trials (20 scenes pairs \times 2 orientation combinations [left scene upright - right scene inverted vs. left scene inverted - right scene upright] \times 3 blur levels [5 vs. 15 vs. 25]) and 80 catch trials (20 scenes pairs \times 2 pairs [both scenes with identical vs. different orientations] \times 2 orientations [upright scene vs. inverted scene]). This resulted in a total of 200 randomly displayed trials with a break every 50 trials. For each trial, we recorded whether participants perceived a blur difference between the scenes and if so which scene they chose as the blurriest or as the sharpest depending on the experimental group. We also recorded the fixation durations on each scene. The experiment lasted about 25 min, depending on the speed at which

participants made their judgment ($M \pm SD$ duration of a trial: 3124 ± 2924 ms).

Data Analysis

Behavioral Data. Data were analyzed using the same software and package as in the previous experiments. To analyze test trials (i.e., on which both scenes had the same blur level), we set up two mixed effects logistic regression models. The first logistic regression aimed to evaluate whether or not participants perceived a blur difference between the two images of test trials (i.e., the dependent variable was perceived difference vs. no perceived difference). The second logistic regression allowed us to evaluate whether participants judged more often the predictable or unpredictable scene as the blurriest in test trials (i.e., the dependent variable was predictable scene judged as the blurriest vs. unpredictable scene judged as the blurriest). For both logistic regressions, the intercept and the blur level of the scenes were entered as fixed effects. Intercepts for subjects and items, as well as subject-wise and item-wise random slopes for the effect of the blur level were specified as random effects. Based on results from Experiment 2, we expected that at an objectively equal level of blur, upright scenes would be explicitly perceived as sharper when compared with inverted scenes. To test this hypothesis, we tested (a) if the proportion of perceived blur difference in test trials was significantly greater than chance and (b) if the proportion of inverted scenes judged as the blurriest was significantly greater than chance. As in Experiment 2, we also tested whether this effect was modulated by the signal precision (i.e., blur level) of the scenes in test trials. If the influence of expectations on perceived sharpness increases as the signal precision decreases, we expected the tendency to select the inverted scene as the blurriest to increase with increasing blur level of scene pairs (i.e., interaction between the effect of predictability and blur level). To test this hypothesis, we calculated the effect of the blur level on the proportion of inverted scenes judged as the blurriest and we also performed pairwise comparisons between each blur level.

To analyze catch trials (i.e., when the two scenes had objectively different blur levels), we also set up a mixed effects logistic regression model with the accuracy in detecting blur differences as a dependent variable. Pair-type (i.e., both scenes having or not the same orientation) was entered into the model as fixed effect. Intercepts for subjects and items, as well as subject-wise and item-wise random slopes for the effect of pair-type were entered as random effects. Based on this model, we calculated whether the accuracy in each condition differed significantly from chance to ensure that participants perceived the blur level difference between the two images and performed the task properly.

For all analyses, the significance of the effects was tested with a Wald test. The significance threshold was set at .025 for main effects of blur level on test trials after Bonferroni correction for two planned nonorthogonal logistic regressions ($.05/2 = .025$), at 0.17 for three a posteriori pairwise comparisons on test trials analyses after Bonferroni correction for three nonorthogonal tests ($.05/3 = .017$) and at .017 for catch trials analyses after Bonferroni correction for three nonorthogonal tests on the accuracy. Data and analysis scripts are available in the Open Science Framework repository: <https://osf.io/d65zs/>.

Eye Movements Data. For eye movements analyses, we defined for each trial two areas of interest (AOI) corresponding to the location of the two scenes on the screen. We then extracted the total fixation duration in each of these AOIs for each trial. We ran eye movements analyses on test trials only, to compare the fixation durations as an index of visual attention between the upright and inverted scenes when both scenes had an objectively equal blur level and were differently oriented. This was done to test whether increased perceived sharpness of upright relative to inverted scenes in test trials could be explained by increased visual attention to upright relative to inverted scenes, as visual attention has been shown to enhance visual perception (Yeshurun & Carrasco, 1998). We performed a linear mixed effects analysis of the Scene orientation (upright vs. inverted) with the total fixation duration of the scene as the dependent variable. Intercept and Scene orientation terms were entered into the model as fixed effects. Intercepts for subjects and items, as well as subject-wise and item-wise random slopes for the effect of Scene orientation were specified as random effects. We calculated the main effect of Scene orientation (upright vs. inverted) on fixation duration to assess differences in visual attention between upright and inverted scenes.

Mixed models were constructed using the same method as in Experiments 1 and 2 (i.e., using Bates, Kliegl, et al., 2015, method for building parsimonious models). Data and analysis scripts are available in the Open Science Framework repository: <https://osf.io/d65zs/>.

Results

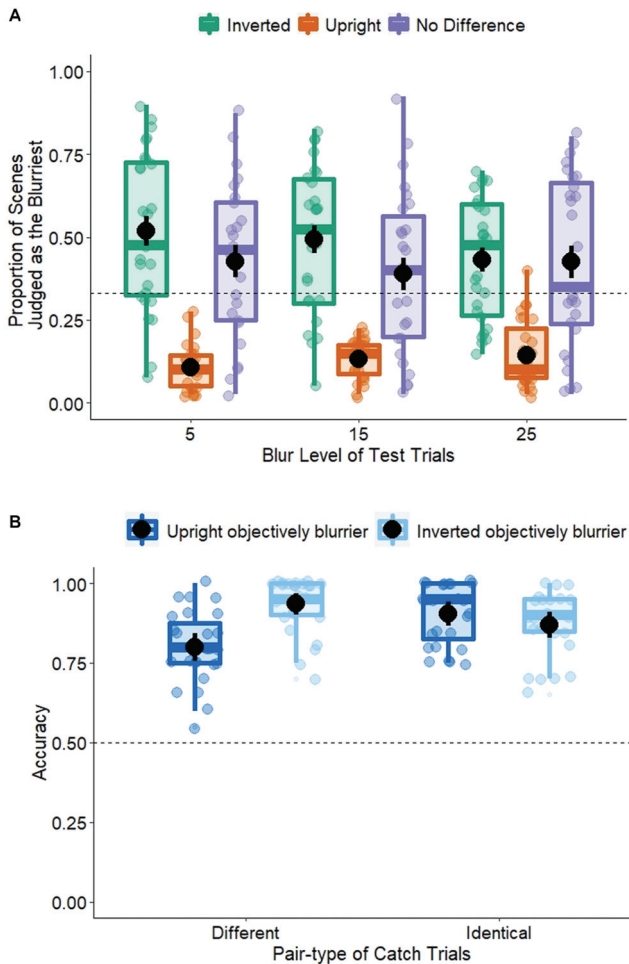
Behavioral Results

Mean proportions of responses for each blur level of test trials are shown in Figure 8A. Mean accuracy for each experimental condition of catch trials is shown in Figure 8B. Eight participants were removed from the analyses due to an accuracy equal or lower than chance in catch trials, indicating that they did not get involved in the task (in fact, these participants did not perform the task and systematically reported “no difference”), resulting in a final sample of 27 participants.

First, the analyses conducted on catch trials revealed that when the two scenes appeared with an objectively different blur level, the accuracy in selecting the blurriest one was high for each condition (different pair-type: $M = .87$, $SE = .07$; identical pair-type: $M = .89$, $SE = .06$) and significantly above chance (different pair-type: $\beta = 2.37$, $z = 7.92$, $p < .001$; identical pair-type: $\beta = 2.58$, $z = 8.57$, $p < .001$). Although the minimal difference between the blur levels of the two scenes was set to 3 to ensure that the two scenes were perceived as having a different blur level, we wondered whether the accuracy decreased when the upright scene was objectively blurrier than the inverted scene relative to the reverse. According to our hypotheses, the upright scene should be perceived as sharper than the inverted one. However, when the upright scene was objectively blurrier than the inverted scene, the increased perceived sharpness of the upright one may have compensated for the difference in blur levels between the two scenes, resulting in a reduced perceived difference. To this end, we performed a mixed effects logistic regression on the catch trials where the two scenes were differently oriented. The intercept and the orientation of the objectively blurriest scene (i.e., upright or inverted)

Figure 8

Mean Proportion of Scenes Judged as the Blurriest According to the Blur Level for Test Trials and Mean Accuracy According to the Type of Pair and the Orientation of the Objectively Blurriest Scene for Catch Trials of Experiment 3



Note. (A) Mean proportion of scenes judged as the blurriest according to the blur level for the test trials (i.e., trials in which both scenes have the same objective blur level). The two scenes in test trials always had a different orientation. Black dots and error bars indicate mean and standard error over all individuals and trials, respectively. Color dots correspond to individual observations. The dashed line represents chance. (B) Mean accuracy for the catch trials (i.e., trials in which the two scenes have an objectively different blur level) according to the orientation of the objectively blurriest scene and the type of pair. The two scenes in catch trials either had a different orientation or an identical orientation. Black dots and error bars indicate mean and standard error over all individuals and trials, respectively. Color dots correspond to individual observations. The dashed line represents chance. See the online article for the color version of this figure.

were entered as fixed effects. Intercepts and random slopes for the effect of the blurriest scene orientation were specified as random effects for the subjects and stimuli. The results showed that the accuracy was significantly lower when the upright scene was blurrier than the inverted one ($M = .80$, $SE = .08$) relative to when the

inverted scene was blurrier than the upright one ($M = .94$, $SE = .05$), $\beta = 2.28$, $z = 5.45$, $p < .001$, arguing in favor of the sharpening mechanism during scene perception.

Regarding test trials, results showed that participants perceived a blur difference ($M = .61$, $SE = .09$) significantly more frequently than no blur difference ($M = .39$, $SE = .09$) even though both scenes had the same objective blur level, $\beta = .68$, $z = 2.60$, $p = .009$, supporting an influence of predictability on blur perception. Moreover, when they perceived a blur difference, participants chose the inverted scene as the blurriest ($M = .79$, $SE = .08$) more frequently than the upright scene ($M = .21$, $SE = .08$), $\beta = 1.42$, $z = 11.19$, $p < .001$. Thus, participants perceived upright scenes as sharper than inverted scenes at an objectively equal level of blur. However, contrary to what was expected, the proportion of inverted scene choice as the blurriest decreased as the blur level of the scene pairs increased (for a low blur of 5: $M = .62$, $SE = .09$; for a medium blur of 15: $M = .63$, $SE = .09$; for a high blur of 25: $M = .57$, $SE = .10$), $\beta = -.02$, $z = -2.32$, $p = .020$. After Bonferroni correction, pairwise comparisons failed to demonstrate a significant difference on the proportion of inverted scene choice as the blurriest between the different blur levels, neither from a blur level of 5 to 15, $\beta = .29$, $z = 1.75$, $p = .081$, nor from a blur level of 15 to 25, $\beta = .17$, $z = 1.13$, $p = .257$, nor from a blur level of 5 to 25, $\beta = .47$, $z = 2.32$, $p = .020$.

Eye Movements Results

The analyses conducted on test trials revealed that participants fixated inverted scenes ($M = 1,849$ ms, $SE = 323$ ms) significantly longer than upright scenes ($M = 1,693$ ms, $SE = 286$ ms), $t(6433.00) = 5.32$, $p < .001$, $d_z = 1.024$, suggesting that more visual attention was allocated to inverted than upright scenes.

Discussion of Experiment 3

This third experiment aimed to confirm results of increased perceived sharpness of predictable relative to unpredictable stimuli implicitly assessed in Experiments 1 and 2 by using an explicit sharpness judgment task. To this end, participants had to decide which of two scenes (upright or inverted) was the blurriest, while both scenes were presented in most trials at an objectively equal level of blur. First, the analyses showed that most participants engaged in the task as they correctly perceived (more than 85% of the time) which scene was the blurriest when the two scenes had an objectively different blur level. Critically, when the two scenes had an objectively equal blur level, we observed that (a) participants did perceive a blur difference and that (b) they judged upright scenes as sharper than inverted scenes. These results further add support to the findings of Experiments 1 and 2 by showing that predictable scenes explicitly appear sharper than unpredictable ones. The analyses of eye movements additionally showed that more visual attention (as indexed by fixation time) was allocated to inverted than upright scenes. This result will be further discussed in the General Discussion.

Contrary to what was expected based on Experiments 1 and 2, participants did not necessarily judge the inverted scene as the blurriest one more often as the scene pairs became blurrier. We even observed the opposite trend, namely that participants judged the inverted scene as the blurriest one less frequently as the blur level of the scene pairs increased; and therefore, as the precision

decreased. One possible explanation for these results is that when the scenes are too blurry, the expectations would be too diverse and too coarse to sharpen the representation of the stimulus. Alternatively, it is possible that when the blur level increased in the task, scenes were actually too blurred to be related to prior knowledge, even in the upright condition. Both the upright and inverted scenes appeared unpredictable, thereby reducing the effect of predictability on perceived sharpness. These analyses according to the blur level of stimuli were conducted as a follow-up for the post hoc analyses performed in Experiments 1 and 2 involving a perceptual matching task. However, in the context of the explicit judgment task used in Experiment 3, the proportion of upright/inverted scene selection as the blurriest may not be an adequate measure to assess how the effect of predictability on perceived sharpness of stimuli is influenced by their precision. Indeed, while this measure reflects the tendency to perceive one scene as sharper or blurrier than the other, it does not allow to quantify this effect (i.e., how big is the perceived difference in sharpness between the two scenes). For example, if the effect of predictability of stimuli on perceived sharpness is greater when stimuli have a medium/low than high precision, but that predictable stimuli are still perceived as sharper than unpredictable ones in all precision conditions, the proportion of selection of unpredictable stimuli as the blurriest may not necessarily differ across precision conditions.

General Discussion

The present study investigated the influence of expectations based on prior experience and contextual information on scene and object perception. Based on prior studies (Alilović et al., 2021; Cheadle et al., 2015; de Lange et al., 2018; Han & VanRullen, 2016; Kok et al., 2012; Lupyan, 2017; Perez et al., 2020; Stein & Peelen, 2015; Teufel et al., 2018; Wyart et al., 2012) suggesting that expectations (or prediction signals) modulate the processing of information in visual areas by sharpening the representation of expected features of a stimulus, we hypothesized that predictable visual stimuli would be subjectively perceived as sharper than unpredictable ones. In Experiments 1 and 2, we used a perceptual blur matching task inspired by Lupyan (2017) to, respectively, address the effect of expectations on the perceived sharpness of objects and whole scenes. In both experiments, results suggested that at an objectively equal level of blur, predictable objects and scenes were subjectively perceived as sharper than unpredictable ones. Experiment 3 confirmed this conclusion using an explicit sharpness discrimination task. Post hoc analyses of Experiment 1 and 2 additionally suggested that the effect of predictability on perceived sharpness of stimuli tended to increase as the precision of the visual signal decreased in line with the hypothesis of predictive coding theories (Friston, 2005; Rao & Ballard, 1999).

Results of the present study replicate previous findings by Lupyan (2017) that semantic knowledge affects the perceived sharpness of letter strings and allow to extend these findings to object and scene perception. Our results are also in agreement with recent findings of Perez et al. (2020) on object perception. In this study, the authors asked participants to perform comparative or equality judgments on the blurriness of familiar and novel objects. The novel objects were made from the same basic shapes composing familiar objects but assembled in a different configuration. Similarly to the present findings, they found that at an equal level of blur, participants judged

familiar objects as sharper than novel ones. Overall, our and past studies support the hypothesis that expectations based on semantic knowledge, but also prior experience and contextual information affect perception by making expected stimuli appear “sharper” and influence “how we perceive.”

Our sharpness adjustment measurements were used to address qualitative perceptual effects. How can we ascertain that they truly reflect the actual sharpness *perception* of participants instead of other nonperceptual effects? For example, Firestone & Scholl (2016; see also Firestone & Scholl, 2014) pointed out that the processing of any visual object not only involves the perception of its visual properties (e.g., its size, its color) but also making higher-level judgments about it (e.g., its cost) that can bias perceptual reports. In the context of our study, it could be argued that a predictable scene or object that can be more easily identified could be more readily judged as “sharp” while unpredictable stimuli would be judged as “blurry.” In this case, the difference in reported perceived sharpness between predictable and unpredictable stimuli would reflect high-level judgments about stimuli rather than perceptual effects. If this was the case, such bias should also be observed in the identical pair-types conditions. Predictable Samples should be more readily judged as sharp and they should be attributed a higher blur level than unpredictable Samples. Critically, this is not what we observed. The predictability effect was reduced in the identical pair-types trials compared with the different pair-type trials, arguing against an explanation of our results simply in terms of higher-level judgments. It has also been argued that expectations do not affect perception or early sensory processing but postperceptual decisional processes involved in perceptual tasks (Bang & Rahnev, 2017). In that sense, the tendency to perceive predictable stimuli as sharper than unpredictable ones would reflect a change in the decision criterion (i.e., response bias) rather than a perceptual bias. For example, when presented with blurred predictable Sample objects, participants could more easily “guess” the presence of sharp edges in the object based on contextual information and adapt their blur adjustment in consequence without it reflects an actual sharpened perception (e.g., they would add more blur than necessary because they assume there are sharp edges without really perceiving these as sharp). However, such a bias should have also been present for predictable Samples in the identical pair-type trials but this is not what we observed. The results observed in these trials rule out an explanation by such a response bias. These considerations make us feel confident that our findings are likely to reflect perceptual effects of expectations.

One could argue that this effect is due to attention rather than prior knowledge. Indeed, attention has been found to enhance the perception of attended stimuli, for example, resulting in increasing their perceived contrast (Carrasco et al., 2004) or spatial resolution (Yeshurun & Carrasco, 1998). Therefore, increased perceived sharpness of predictable stimuli in our study could be explained by the fact that more attention was devoted to predictable than unpredictable stimuli. However, we believe this explanation is unlikely for several reasons. First, the tasks used in our experiments required participants to attend both the Target and Sample stimuli that were predictable or unpredictable in an equal number of trials. Therefore, if more attention was, for example, allocated to the Sample relative to the Target (or vice versa), the effects should be similar for predictable and unpredictable stimuli. It could also be argued that the effect is due to the difference in the task relevance

between a predictable context and unpredictable context; as task relevance can be understood in terms of attention, it is possible that it influenced blur perception. However, participants' task was to focus on the blur adjustment (Experiments 1 and 2) or blur comparison (Experiment 3). The predictability was irrelevant to the task. Furthermore, several studies have shown that when it comes to object and scene perception, attention tends to be captured by novelty (Johnston et al., 1990; Park et al., 2010). In the context of our study, more attention should have been allocated to unpredictable stimuli, which can be considered as novel since they were not related to a prior experience, than to predictable stimuli that were more familiar. This was supported by the analyses of fixation duration in Experiment 3, which showed that unpredictable stimuli elicited longer fixation durations than predictable stimuli. Therefore, more attention to unpredictable stimuli cannot explain the increased perceived sharpness of predictable stimuli observed in the present study. Finally, although Experiments 1 and 2 were run online on participants' personal computers and in variable viewing conditions (e.g., in terms of viewing distance, screen size, luminance, contrast, etc.), we strongly believe that differences in terms of interindividual settings could not affect or explain our results. Indeed, the task required to compare and adjust the blur level of the Sample stimulus to match a Target stimulus simultaneously displayed on the same horizontal plane so that the two stimuli were always viewed and compared under the same conditions at each trial.

Increased perceived sharpness for expected stimuli observed in the present study supports the sharpening account of predictive coding theories (Alilović et al., 2021; Cheadle et al., 2015; de Lange et al., 2018; Han & VanRullen, 2016; Kok et al., 2012; Lupyan, 2017; Perez et al., 2020; Stein & Peelen, 2015; Teufel et al., 2018; Wyart et al., 2012). According to this account, expectations (or prediction signals) sent by high-level areas increase the sensitivity of neurons tuned to expected features in lower-level areas while suppressing the response of neurons tuned to unexpected ones. This results in a sharper representation of expected stimuli in these areas that would in turn lead to a sharper percept (de Lange et al., 2018; Han & VanRullen, 2016; Kok et al., 2012; Teufel et al., 2018; Yon et al., 2018). Under this framework, results of Experiment 1 indicate that, although irrelevant to the task, contextual scene information was used to form prediction signals. These prediction signals could then have back-propagated from high- to low-order areas during object processing. In the case of predictable contextual information, this would, for example, allow to selectively enhance the response of neurons tuned to expected objects or shapes in areas of the ventral stream and of neurons tuned to the expected contours at the object location in early visual areas. This would in turn lead to a sharper percept of the object. In the case of unpredictable contextual information, prediction signals would not be precise enough leading to a noisier representation of the object in low-level cortical areas. Results of Experiments 2 and 3 on whole scene perception additionally suggest that sharpened representations based on contextual information not only apply to the processing of expected objects but more broadly to expected features in the entire scene.

This interpretation is also coherent with recent neuroimaging studies addressing the role of contextual information on visual processing. For example, Brandman and Peelen (2017) asked participants to categorize degraded objects presented alone or in a

context (e.g., a bird flying over a lake, a tractor in a field) during an fMRI experiment. They examined the pattern of activity in object-selective areas of the ventral stream and found that the category of the object could be decoded with higher accuracy in these areas when they were presented in context than in isolation or when the context was presented without the object. This contextual effect in object-selective areas was associated with increased activity in scene-selective areas, suggesting that this area could be the origin of prediction signals about the object. More recently, Heilbron et al. (2020), presented participants with five-letter words and pseudowords in an fMRI experiment. The middle letter of all letter strings was always an "N" or a "U." A decoding analysis of activity patterns in V1 and V2 revealed that the middle letter could be decoded with higher accuracy in these areas when it was embedded in a word than a pseudoword context. Overall, results from these studies are consistent with the view that expectations based on contextual information can enhance the representation of expected features in relatively high (i.e., object-selective areas) and low-level visual areas (i.e., V1 and V2).

While results of the present study support the sharpening account of predictive coding theories, they do not necessarily argue against the dampening account (de Lange et al., 2018; Kumar et al., 2017; Press et al., 2020; Richter et al., 2018). First, this account posits that prediction signals filter out predicted features of stimuli by silencing neurons tuned to these features and increasing the sensitivity of neurons tuned to other features. Under this account, one would expect increased perceived sharpness of stimuli violating expectations. Critically, our experimental manipulations pertained to the *presence or absence* of informative content or prior knowledge allowing to form expectations via scrambling of contextual information or via scene inversion. These manipulations were chosen because they allowed us to control for low-level visual differences between the Target and Sample stimuli. However, they did not allow us to assess the effect of expectations *validity* as we did not include a condition in which stimuli could be unexpected or surprising and violating expectations (e.g., an object in an intact but incongruent context). Such a condition would be needed in further experiments to arbitrate between the sharpening and dampening accounts. Recent works suggest that these two accounts of predictive coding may not be mutually exclusive but could coexist (Press et al., 2020), so that their influence on perception would vary according to temporal constraints and signal precision. Under this framework, the sharpening mechanism would take place first, allowing to confirm expectations. The dampening mechanism would subsequently take over to boost the processing of the remaining unexpected features. This sequence would be further constrained by the signal characteristics. The sharpening mechanism would predominate when we have strong expectations but the visual signal is noisy. Predicted features would weigh more while unpredicted ones would be suppressed. The dampening mechanism would only take over when the signal is reliable but strongly diverges from expectations. Unpredicted features eliciting surprise would then weigh more while predicted features would be suppressed. In this context, increased perceived sharpness of predictable stimuli in our study could reflect a predominance of the sharpening mechanism to process noisy (blurred) stimuli. This assumption is supported by post hoc analyses of Experiments 1 and 2 suggesting that the effect of predictability on perceived sharpness increases with increasing

blur level of the Target. However, results of Experiment 3 suggest the processing of stimuli that are too noisy and for which expectations are too coarse may not benefit from this sharpening mechanism.

Another possibility that has been put forward in the literature is that expectations do not modulate early sensory response but later processes related to decision making and response selection (e.g., Bang & Rahnev, 2017; Rungratsameetaweemana et al., 2018; Rungratsameetaweemana & Serences, 2019; see however, Aitken et al., 2020, for recent arguments against this view). Under this account, sharpened perception of predictable stimuli in the present study could result from a reweighting or selective attention to predicted features of stimuli occurring at later processing stages. Unfortunately, the present study cannot elucidate whether sharpened perception of predictable stimuli results from a modulation of early sensory or later cognitive processes. As previously discussed, our findings nevertheless provide evidence that expectations influence subjective perception, whether at early or late perceptual processing stages.

An important remaining question is whether sharper perception of expected visual features explains the effects of expectations in facilitating visual recognition, for example, in terms of recognition speed or accuracy as reported in many behavioral studies (Brandman & Peelen, 2017; Davenport & Potter, 2004; Greene et al., 2015; Joubert et al., 2007, 2008; Palmer, 1975; Richter et al., 2018; Roux-Sibilon et al., 2019). Could expected stimuli be recognized faster or more accurately *because* they are perceived better? Because the present study did not explicitly involve the recognition of scenes and objects, this assumption is only speculative and to our knowledge, no study has directly addressed this question so far. Further experiments using an explicit recognition task and response time recordings could enrich this point. In the context of our study, it could, for example, be hypothesized that a blurred predictable stimulus would be recognized faster than an unpredictable stimulus with a similar or even lower blur level. In any case, findings from fMRI studies support the assumption that better perception and better recognition performances are correlated. For example, in Kok et al. (2012) sharpened representations of the stimuli in the visual cortex based on expectations were associated with higher behavioral performances (although no causal link could be established based on correlational analyses). A recent study by Gandolfo and Downing (2019) directly addressed the causal relationship between expectations and visual recognition performance, as well as the related cortical activity. In their study, the authors asked participants to categorize images of bodies (heavy vs. slim) and scenes (upright vs. inverted). These stimuli were preceded by a valid or an invalid cue that was orthogonal to the categorization task but could carry useful information to facilitate the categorization. For example, the cue for a body image could be either the “m” (i.e., for male) or the “f” letter (i.e., for female) implicitly associated with a heavy or a slim body, respectively. Results showed that a valid cue improved the categorization performances of both bodies and scenes. Then, during the cue presentation, the brain activity in body-selective or scene-selective cortical areas was disrupted via transcranial magnetic stimulation (TMS). Critically, the facilitation effect by a valid cue disappeared when the TMS was applied on body- or scene-selective areas during the cue presentation for the body or scene task, respectively. These results indicated that expectation-based activity in these

areas is directly related to categorization performance. Overall, these studies support the idea that the expression of expectations in visual areas facilitates visual recognition. Further studies would be needed to address whether higher recognition performances for expected visual stimuli are also associated with their increased perceived sharpness.

Our results also more broadly add support to the “cognitive penetrability of perception” theoretical framework (Lupyan, 2015; O’Callaghan et al., 2017; for arguments against this view, see Firestone & Scholl, 2016; Pylyshyn, 1999) according to which perception not only affects one’s cognitive state or knowledge but is also constrained by it. In this context, our findings are coherent with previous works suggesting that expectations or semantic knowledge influence other aspects of perception such as contrast and color perception (Han & VanRullen, 2016; Lupyan, 2015), perceived duration of stimuli (Pariyadath & Eagleman, 2007; Schindel et al., 2011; Schweitzer et al., 2017), or perceived clarity of speech (Sohoglu et al., 2014). For example, Han and VanRullen (2016) found that line-drawings of objects were perceived as more contrasted with respect to the background than the same line-drawings with random locations (meaningless stimulus). In the context of color perception, Lupyan (2015) found that objects associated with a strong color-prior (e.g., yellow bananas) were perceived as more vivid than objects of the same color without such a strong prior (e.g., a yellow t-shirt). In the same way, Hansen et al. (2006; see also Olkkonen et al., 2008; Witzel et al., 2011) conducted a perceptual matching task in which participants were required to adjust the color of objects with strong color priors (e.g., yellow bananas) and colored discs without any color prior until they appeared gray. They observed that participants adjusted colored objects toward their opponent color to perceive them as gray (e.g., a yellow banana was adjusted to be slightly blue) suggesting that when they were objectively gray, they still appeared in their original color. However, colored discs unrelated to color-prior were accurately adjusted to an objective gray level. These results indicated that prior knowledge can also influence color appearance. In the context of the sharpening account of predictive coding, these results could be explained by an enhanced perception of expected features (e.g., colors) of stimuli. Perceptual matching tasks such as the one used in the present study adapted to different contexts or modalities (e.g., speech processing, tactile or auditory perception) and using other matching criteria (e.g., color or sound intensity, noise level) may be a useful tool to more broadly assess how cognition can modulate perception.

Such perceptual effects may also constitute a useful measure to assess the use of prior knowledge and expectations in specific populations and disorders. For example, disorders such as autism spectrum disorders or schizophrenia have been linked to an insufficient or excessive influence of expectations on perception, respectively (e.g., Pellicano & Burr, 2012; Powers et al., 2017; Sinha et al., 2014; Teufel et al., 2015; Van de Cruys et al., 2014). In the context of our study, one could, for example, expect that participants with autism would not necessarily perceive predictable stimuli as sharper than unpredictable ones. On the contrary, schizophrenic patients should need to add even more blur than control participants to a predictable stimulus to perceive it as equally blurred as an unpredictable one. Predictive mechanisms have also been shown to evolve over the life span. For example, it has been proposed that the accumulation of sensory experience

with aging may lead to rely more strongly on predictive processes, which would also help compensate for concomitant sensory deficits (Gilbert & Moran, 2016; Lai et al., 2020; Moran et al., 2014). Under this assumption, it can be expected that the effect of predictability on the perceived sharpness of stimuli should be greater in aged than young individuals.

Conclusion

In conclusion, the results of the present study showed that expectations based on prior experience and contextual information increase the perceived sharpness of scenes and objects they contain; thus, supporting the sharpening account of predictive coding theories (de Lange et al., 2018; Han & VanRullen, 2016; Kok et al., 2012; Teufel et al., 2018; Yon et al., 2018). Our findings allow us to further extend the previous results of Lupyan (2017) on word reading to the perception of whole scenes and objects, defining the sharpening mechanism as a more general principle of visual perception. Expectations and knowledge about the visual environment not only help us to understand it more easily, but also makes us perceive it better.

References

- Aitken, F., Turner, G., & Kok, P. (2020). Prior expectations of motion direction modulate early sensory processing. *The Journal of Neuroscience*, *40*(33), 6389–6397. <https://doi.org/10.1523/JNEUROSCI.0537-20.2020>
- Alilović, J., Slagter, H. A., & van Gaal, S. (2021). Subjective visibility report is facilitated by conscious predictions only. *Consciousness and Cognition*, *87*, 103048. <https://doi.org/10.1016/j.concog.2020.103048>
- Bang, J. W., & Rahnev, D. (2017). Stimulus expectation alters decision criterion but not sensory signal in perceptual decision making. *Scientific Reports*, *7*(1), 17072. <https://doi.org/10.1038/s41598-017-16885-2>
- Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience*, *5*(8), 617–629. <https://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. <https://doi.org/10.1016/j.tics.2007.05.005>
- Bates, D., Kliegl, R., Vasishth, S., & Baayen, R. H. (2015). *Parsimonious mixed models*. <https://arxiv.org/pdf/1506.04967.pdf>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Blakemore, S.-J., Wolpert, D. M., & Frith, C. D. (1998). Central cancellation of self-produced tickle sensation. *Nature Neuroscience*, *1*(7), 635–640. <https://doi.org/10.1038/2870>
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*(4), 433–436. <https://doi.org/10.1163/156856897X00357>
- Brandman, T., & Peelen, M. V. (2017). Interaction between scene and object processing revealed by human fMRI and MEG decoding. *The Journal of Neuroscience*, *37*(32), 7700–7710. <https://doi.org/10.1523/JNEUROSCI.0582-17.2017>
- Carrasco, M., Ling, S., & Read, S. (2004). Attention alters appearance. *Nature Neuroscience*, *7*(3), 308–313. <https://doi.org/10.1038/nn1194>
- Cheadle, S., Egner, T., Wyart, V., Wu, C., & Summerfield, C. (2015). Feature expectation heightens visual sensitivity during fine orientation discrimination. *Journal of Vision*, *15*(14), 14. <https://doi.org/10.1167/15.14.14>
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204. <https://doi.org/10.1017/S0140525X12000477>
- Davenport, J. L. (2007). Consistency effects between objects in scenes. *Memory & Cognition*, *35*(3), 393–401. <https://doi.org/10.3758/BF03193280>
- Davenport, J. L., & Potter, M. C. (2004). Scene consistency in object and background perception. *Psychological Science*, *15*(8), 559–564. <https://doi.org/10.1111/j.0956-7976.2004.00719.x>
- de Lange, F. P., Heilbron, M., & Kok, P. (2018). How do expectations shape perception? *Trends in Cognitive Sciences*, *22*(9), 764–779. <https://doi.org/10.1016/j.tics.2018.06.002>
- Feldman, H., & Friston, K. J. (2010). Attention, uncertainty, and free-energy. *Frontiers in Human Neuroscience*, *4*, 215. <https://doi.org/10.3389/fnhum.2010.00215>
- Firestone, C., & Scholl, B. J. (2014). “Top-down” effects where none should be found: The El Greco fallacy in perception research. *Psychological Science*, *25*(1), 38–46. <https://doi.org/10.1177/0956797613485092>
- Firestone, C., & Scholl, B. J. (2016). Cognition does not affect perception: Evaluating the evidence for “top-down” effects. *Behavioral and Brain Sciences*, *39*, e229. <https://doi.org/10.1017/S0140525X15000965>
- Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *360*(1456), 815–836. <https://doi.org/10.1098/rstb.2005.1622>
- Gandolfo, M., & Downing, P. E. (2019). Causal evidence for expression of perceptual expectations in category-selective extrastriate regions. *Current Biology*, *29*(15), 2496–2500.e3. <https://doi.org/10.1016/j.cub.2019.06.024>
- Gilbert, J. R., & Moran, R. J. (2016). Inputs to prefrontal cortex support visual recognition in the aging brain. *Scientific Reports*, *6*, 31943. <https://doi.org/10.1038/srep31943>
- Greene, M. R., & Oliva, A. (2009). Recognition of natural scenes from global properties: Seeing the forest without representing the trees. *Cognitive Psychology*, *58*(2), 137–176. <https://doi.org/10.1016/j.cogpsych.2008.06.001>
- Greene, M. R., Botros, A. P., Beck, D. M., & Fei-Fei, L. (2015). What you see is what you expect: Rapid scene understanding benefits from prior experience. *Attention, Perception, & Psychophysics*, *77*(4), 1239–1251. <https://doi.org/10.3758/s13414-015-0859-8>
- Han, B., & VanRullen, R. (2016). Shape perception enhances perceived contrast: Evidence for excitatory predictive feedback? *Scientific Reports*, *6*(1), 22944. <https://doi.org/10.1038/srep22944>
- Hansen, T., Olkkonen, M., Walter, S., & Gegenfurtner, K. R. (2006). Memory modulates color appearance. *Nature Neuroscience*, *9*(11), 1367–1368. <https://doi.org/10.1038/nn1794>
- Heilbron, M., Richter, D., Ekman, M., Hagoort, P., & de Lange, F. P. (2020). Word contexts enhance the neural representation of individual letters in early visual cortex. *Nature Communications*, *11*(1), 321. <https://doi.org/10.1038/s41467-019-13996-4>
- Johnston, W. A., Hawley, K. J., Plewe, S. H., Elliott, J. M. G., & DeWitt, M. J. (1990). Attention capture by novel stimuli. *Journal of Experimental Psychology: General*, *119*(4), 397–411. <https://doi.org/10.1037/0096-3445.119.4.397>
- Joubert, O. R., Fize, D., Rousselet, G. A., & Fabre-Thorpe, M. (2008). Early interference of context congruence on object processing in rapid visual categorization of natural scenes. *Journal of Vision*, *8*(13), 11. <https://doi.org/10.1167/8.13.11>
- Joubert, O. R., Rousselet, G. A., Fize, D., & Fabre-Thorpe, M. (2007). Processing scene context: Fast categorization and object interference. *Vision Research*, *47*(26), 3286–3297. <https://doi.org/10.1016/j.visres.2007.09.013>
- Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology*, *103*(1), 54–69. <https://doi.org/10.1037/a0028347>
- Kaiser, D., Quek, G. L., Cichy, R. M., & Peelen, M. V. (2019). Object vision in a structured world. *Trends in Cognitive Sciences*, *23*(8), 672–685. <https://doi.org/10.1016/j.tics.2019.04.013>

- Kauffmann, L., Ramanoël, S., & Peyrin, C. (2014). The neural bases of spatial frequency processing during scene perception. *Frontiers in Integrative Neuroscience*, 8, 37. <https://doi.org/10.3389/fnint.2014.00037>
- Kirchner, H., & Thorpe, S. J. (2006). Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited. *Vision Research*, 46(11), 1762–1776. <https://doi.org/10.1016/j.visres.2005.10.002>
- Kok, P., & de Lange, F. P. (2015). Predictive coding in sensory cortex. In B. U. Forstmann & E. J. Wagenmakers (Eds.), *An introduction to model-based cognitive neuroscience* (pp. 221–244). Springer. https://doi.org/10.1007/978-1-4939-2236-9_11
- Kok, P., Jehee, J. F. M., & de Lange, F. P. (2012). Less is more: Expectation sharpens representations in the primary visual cortex. *Neuron*, 75(2), 265–270. <https://doi.org/10.1016/j.neuron.2012.04.034>
- Kumar, S., Kaposvari, P., & Vogels, R. (2017). Encoding of predictable and unpredictable stimuli by inferior temporal cortical neurons. *Journal of Cognitive Neuroscience*, 29(8), 1445–1454. https://doi.org/10.1162/jocn_a_01135
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1–26. <https://doi.org/10.18637/jss.v082.i13>
- Lai, L. Y., Frömer, R., Festa, E. K., & Heindel, W. C. (2020). Age-related changes in the neural dynamics of bottom-up and top-down processing during visual object recognition: An electrophysiological investigation. *Neurobiology of Aging*, 94, 38–49. <https://doi.org/10.1016/j.neurobiolaging.2020.05.010>
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, 4, 863. <https://doi.org/10.3389/fpsyg.2013.00863>
- Lee, T. S., & Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America A*, 20(7), 1434–1448. <https://doi.org/10.1364/JOSAA.20.001434>
- Lupyan, G. (2015). Object knowledge changes visual appearance: Semantic effects on color afterimages. *Acta Psychologica*, 161, 117–130. <https://doi.org/10.1016/j.actpsy.2015.08.006>
- Lupyan, G. (2017). Objective effects of knowledge on visual perception. *Journal of Experimental Psychology: Human Perception and Performance*, 43(4), 794–806. <https://doi.org/10.1037/xhp0000343>
- Moran, R. J., Symmonds, M., Dolan, R. J., & Friston, K. J. (2014). The brain ages optimally to model its environment: Evidence from sensory learning over the adult lifespan. *PLoS Computational Biology*, 10(1), e1003422. <https://doi.org/10.1371/journal.pcbi.1003422>
- Murray, S. O., Schrater, P., & Kersten, D. (2004). Perceptual grouping and the interactions between visual cortical areas. *Neural Networks*, 17(5-6), 695–705. <https://doi.org/10.1016/j.neunet.2004.03.010>
- O’Callaghan, C., Kveraga, K., Shine, J. M., Adams, R. B., Jr., & Bar, M. (2017). Predictions penetrate perception: Converging insights from brain, behaviour and disorder. *Consciousness and Cognition*, 47(5), 63–74. <https://doi.org/10.1016/j.concog.2016.05.003>
- Oliva, A., & Torralba, A. (2007). The role of context in object recognition. *Trends in Cognitive Sciences*, 11(12), 520–527. <https://doi.org/10.1016/j.tics.2007.09.009>
- Olkkonen, M., Hansen, T., & Gegenfurtner, K. R. (2008). Color appearance of familiar objects: Effects of object shape, texture, and illumination changes. *Journal of Vision*, 8(5), 13. <https://doi.org/10.1167/8.5.13>
- Palmer, T. E. (1975). The effects of contextual scenes on the identification of objects. *Memory & Cognition*, 3(5), 519–526. <https://doi.org/10.3758/BF03197524>
- Pariyadath, V., & Eagleman, D. (2007). The effect of predictability on subjective duration. *PLoS ONE*, 2(11), e1264. <https://doi.org/10.1371/journal.pone.0001264>
- Park, J., Shimojo, E., & Shimojo, S. (2010). Roles of familiarity and novelty in visual preference judgments are segregated across object categories. *Proceedings of the National Academy of Sciences of the United States of America*, 107(33), 14552–14555. <https://doi.org/10.1073/pnas.1004374107>
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10(4), 437–442. <https://doi.org/10.1163/156856897X00366>
- Pellicano, E., & Burr, D. (2012). When the world becomes ‘too real’: A Bayesian explanation of autistic perception. *Trends in Cognitive Sciences*, 16(10), 504–510. <https://doi.org/10.1016/j.tics.2012.08.009>
- Perez, D. C., Cook, S. M., & Peterson, M. A. (2020). Prior experience alters the appearance of blurry object borders. *Scientific Reports*, 10(1), Article no. 5821. <https://doi.org/10.1038/s41598-020-62728-y>
- Powers, A. R., Mathys, C., & Corlett, P. R. (2017). Pavlovian conditioning-induced hallucinations result from overweighting of perceptual priors. *Science*, 357(6351), 596–600. <https://doi.org/10.1126/science.aan3458>
- Press, C., Kok, P., & Yon, D. (2020). The perceptual prediction paradox. *Trends in Cognitive Sciences*, 24(1), 13–24. <https://doi.org/10.1016/j.tics.2019.11.003>
- Pylyshyn, Z. (1999). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences*, 22, 341–365. <https://doi.org/10.1017/S0140525X99002022>
- R Core Team. (2019). *R: A language and environment for statistical computing* (3.6.1). R Foundation for Statistical Computing.
- 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. <https://doi.org/10.1038/4580>
- Richter, D., Ekman, M., & de Lange, F. P. (2018). Suppressed sensory response to predictable object stimuli throughout the ventral visual stream. *The Journal of Neuroscience*, 38(34), 7452–7461. <https://doi.org/10.1523/JNEUROSCI.3421-17.2018>
- Rousset, 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. <https://doi.org/10.1080/13506280444000553>
- Roux-Sibilon, A., Trouilloud, A., Kauffmann, L., Guyader, N., Mermillod, M., & Peyrin, C. (2019). Influence of peripheral vision on object categorization in central vision. *Journal of Vision*, 19(14), 7. <https://doi.org/10.1167/19.14.7>
- Rungratsameetaweemana, N., & Serences, J. T. (2019). Dissociating the impact of attention and expectation on early sensory processing. *Current Opinion in Psychology*, 29, 181–186. <https://doi.org/10.1016/j.copsyc.2019.03.014>
- Rungratsameetaweemana, N., Itthipuripat, S., Salazar, A., & Serences, J. T. (2018). Expectations do not alter early sensory processing during perceptual decision-making. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 38(24), 5632–5648. <https://doi.org/10.1523/JNEUROSCI.3638-17.2018>
- Schindel, R., Rowlands, J., & Arnold, D. H. (2011). The oddball effect: Perceived duration and predictive coding. *Journal of Vision*, 11(2), 17. <https://doi.org/10.1167/11.2.17>
- Schweitzer, R., Trapp, S., & Bar, M. (2017). Associated information increases subjective perception of duration. *Perception*, 46(8), 1000–1007. <https://doi.org/10.1177/0301006616689579>
- Schyns, P. G., & Oliva, A. (1994). From blobs to boundary edges: Evidence for time- and spatial-scale-dependent scene recognition. *Psychological Science*, 5(4), 195–200. <https://doi.org/10.1111/j.1467-9280.1994.tb00500.x>
- Sinha, P., Kjelgaard, M. M., Gandhi, T. K., Tsourides, K., Cardinaux, A. L., Pantazis, D., Diamond, S. P., & Held, R. M. (2014). Autism as a disorder of prediction. *Proceedings of the National Academy of Sciences of the United States of America*, 111(42), 15220–15225. <https://doi.org/10.1073/pnas.1416797111>
- Sohoglu, E., Peelle, J. E., Carlyon, R. P., & Davis, M. H. (2014). Top-down influences of written text on perceived clarity of degraded speech. *Journal of Experimental Psychology: Human Perception and Performance*, 40(1), 186–199. <https://doi.org/10.1037/a0033206>

- Spratling, M. W. (2017). A hierarchical predictive coding model of object recognition in natural images. *Cognitive Computation*, 9(2), 151–167. <https://doi.org/10.1007/s12559-016-9445-1>
- Stein, T., & Peelen, M. V. (2015). Content-specific expectations enhance stimulus detectability by increasing perceptual sensitivity. *Journal of Experimental Psychology: General*, 144(6), 1089–1104. <https://doi.org/10.1037/xge0000109>
- Summerfield, C., & de Lange, F. P. (2014). Expectation in perceptual decision making: Neural and computational mechanisms. *Nature Reviews Neuroscience*, 15(11), 745–756. <https://doi.org/10.1038/nrn3838>
- Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences*, 13(9), 403–409. <https://doi.org/10.1016/j.tics.2009.06.003>
- Teufel, C., & Fletcher, P. C. (2020). Forms of prediction in the nervous system. *Nature Reviews Neuroscience*, 21(4), 231–242. <https://doi.org/10.1038/s41583-020-0275-5>
- Teufel, C., Dakin, S. C., & Fletcher, P. C. (2018). Prior object-knowledge sharpens properties of early visual feature-detectors. *Scientific Reports*, 8(1), 10853. <https://doi.org/10.1038/s41598-018-28845-5>
- Teufel, C., Subramaniam, N., Dobler, V., Perez, J., Finnemann, J., Mehta, P. R., Goodyear, I. M., & Fletcher, P. C. (2015). Shift toward prior knowledge confers a perceptual advantage in early psychosis and psychosis-prone healthy individuals. *Proceedings of the National Academy of Sciences of the United States of America*, 112(43), 13401–13406. <https://doi.org/10.1073/pnas.1503916112>
- Thorpe, S. J., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, 381, 520–522. <https://doi.org/10.1038/381520a0>
- Torralba, A. (2003). Contextual priming for object detection. *International Journal of Computer Vision*, 53(2), 169–191. <https://doi.org/10.1023/A:1023052124951>
- Van de Cruys, S., Evers, K., Van der Hallen, R., Van Eylen, L., Boets, B., de-Wit, L., & Wagemans, J. (2014). Precise minds in uncertain worlds: Predictive coding in autism. *Psychological Review*, 121(4), 649–675. <https://doi.org/10.1037/a0037665>
- Witzel, C., Valkova, H., Hansen, T., & Gegenfurtner, K. R. (2011). Object knowledge modulates colour appearance. *i-Perception*, 2(1), 13–49. <https://doi.org/10.1068/i0396>
- Wyart, V., Nobre, A. C., & Summerfield, C. (2012). Dissociable prior influences of signal probability and relevance on visual contrast sensitivity. *Proceedings of the National Academy of Sciences of the United States of America*, 109(9), 3593–3598. <https://doi.org/10.1073/pnas.1120118109>
- Yeshurun, Y., & Carrasco, M. (1998). Attention improves or impairs visual performance by enhancing spatial resolution. *Nature*, 396(6706), 72–75. <https://doi.org/10.1038/23936>
- Yon, D., Gilbert, S. J., de Lange, F. P., & Press, C. (2018). Action sharpens sensory representations of expected outcomes. *Nature Communications*, 9(1), 4288. <https://doi.org/10.1038/s41467-018-06752-7>

Received June 22, 2021

Revision received November 8, 2021

Accepted December 15, 2021 ■

ANNEXE B - Prétest sur la linéarité de la perception subjective de flou

Plusieurs études ont mis en évidence que la perception de flou ne suivrait pas une tendance linéaire (Maiello et al., 2017; voir Watson & Ahumada, 2011 pour une revue) mais plutôt une *dipper function* (i.e., en forme de louche). En effet, la différence de flou minimum pour être perçue serait très faible lorsque les stimuli sont relativement nets. Par contre, lorsque les stimuli sont très flous, il faudrait une grande différence de flous entre eux pour juger qu'ils sont effectivement différents.

Il est donc possible que dans les Expériences 1, 2 et 4, dans le cas de l'ajustement du stimulus sample à un stimulus target très flou, les participants n'arrivaient pas à percevoir la différence de flou entre deux niveaux contigus de flou du sample. Ils auraient donc eu la même sensation de flou pour un niveau de flou de 23 et de 24 par exemple. Cela pourrait expliquer que dans ces expériences, les participants avaient plus de difficultés à ajuster correctement le flou du stimulus sample à celui du stimulus target lorsque le stimulus target apparaissait à un niveau de flou élevé (i.e., un stimulus très flou) que lorsqu'il apparaissait à un niveau de flou faible (i.e., un stimulus relativement net).

Pour les expériences ultérieures (Expérience 5) utilisant un paradigme d'ajustement du flou du stimulus sample à celui du stimulus target, nous avons souhaité mieux contrôler que les participants percevaient bien une différence de flou entre chacun des niveaux de flou utilisés. Pour cela, nous avons réalisé un prétest afin de déterminer le seuil de discrimination du flou (i.e., niveau de flou minimum nécessaire entre deux stimuli pour percevoir une différence de flou) spécifiquement pour les stimuli d'objets que nous allons utiliser dans l'Expérience 5.

Méthode

Participants et Design

Neuf étudiants en licence de psychologie de l'Université Grenoble Alpes (7 femmes, $M_{\text{age}} = 20.71$, $SD_{\text{age}} = 1.03$) ayant une vision normale ou corrigée à la normale ont participé à l'expérience.

Stimuli

Les stimuli étaient constitués de 20 photographies en couleurs d'objets (10 animés et 10 non animés). Les photographies ont été téléchargées sur le site Pixabay (<https://pixabay.com/fr>) et sur la base de données Google Images en filtrant pour la licence CC0. Les stimuli ont été créés à l'aide de MATLAB R2019b (MathWorks, Natick, MA, USA).

Les photographies originales des objets ont d'abord été recadrées au format carré et redimensionnées à 120×120 pixels. Les objets ont été isolés de leur arrière-plan en délimitant leur contour sur le logiciel GIMP (The GIMP Development Team, 2019) et en appliquant des pixels transparents à leur arrière-plan. Ils ont ensuite été convertis en images en niveaux de gris en faisant la moyenne des valeurs des trois canaux de couleur pour chaque pixel. La luminance moyenne et le contraste RMS de toutes les images ont également été égalisés (pour les valeurs d'intensité des pixels entre 0 et 1) à 0.5 et la valeur moyenne du contraste RMS de l'ensemble des images (0.2117) a été appliquée à chaque image. Nous avons flouté l'image de l'objet à l'aide de filtres gaussiens 2D (fonction *imgaussfilt* dans Matlab) de taille linéairement croissante (i.e., avec un écart-type croissant - SD) pour obtenir 31 versions de l'objet allant d'un objet net (proche de l'état intact ; niveau de flou 1, SD = 1) à un objet très flou et difficilement reconnaissable (niveau de flou 31, SD = 10). Afin de limiter la durée de l'expérience, nous n'avons pas testé le seuil de discrimination pour chacun des 31 niveaux de flou. Nous avons donc échantillonné cinq niveaux de flou référence (3, 8, 13, 18 et 23) auxquels étaient comparés des stimuli de niveau de flou variable autour de la référence, permettant d'évaluer le seuil de discrimination pour des objets relativement nets (i.e., niveau de flou de 3) jusqu'à des objets très flous (i.e., niveau de flou de 23). Nous avons ensuite déterminé le seuil de discrimination prédit pour chacun de nos cinq niveaux de flou référence à l'aide de l'équation de paramétrisation de la dipper function du flou (Maiello et al., 2017; Mather & Smith, 2002) :

$$\Delta b = \sqrt{\left(1 + 1/S\right) \left(\sigma_i^2 + \sigma_e^2\right) - \sigma_e}$$

Δb correspond au seuil de discrimination prédit, S correspond à la sensibilité au flou du système visuel de l'observateur, σ_i correspond au niveau de flou intrinsèque du système visuel en minutes d'arc et σ_e correspond au niveau de flou référence en minutes d'arc. Pour utiliser cette formule, nous avons donc converti nos cinq niveaux de flou référence en minutes d'arc, ce qui nous donnait respectivement 1.92, 3.72, 5.52, 7.32 et 9.12 minutes d'arc pour σ_e . D'après les résultats de Maiello et al. (2017) sur la perception de stimuli en vision binoculaire et à une excentricité visuelle de 0° d'angle (i.e., en vision centrale), nous avons fixé les paramètres de

sensibilité au flou S et de flou intrinsèque σ_i à 0.9 et 2 respectivement. Le seuil de discrimination prédit arrondi à l'entier supérieur pour les cinq niveaux de flou référence de 3 à 23 était respectivement de 3, 3, 4, 4 et 5.

Procédure

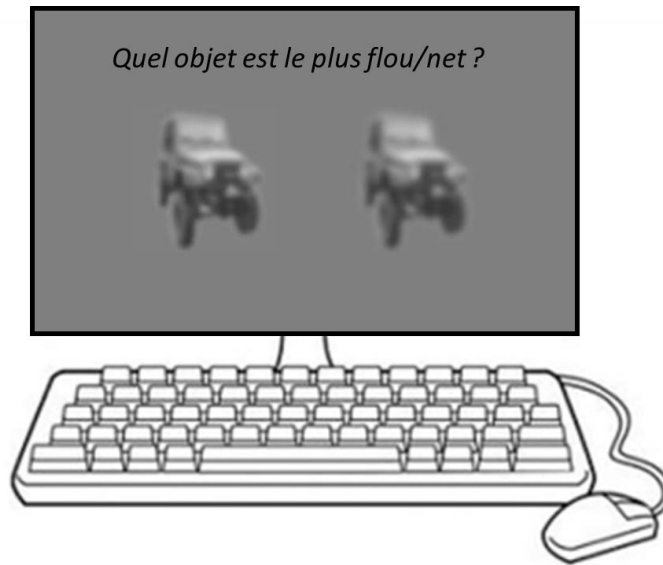
L'expérience a été construite à l'aide de la Psychtoolbox (Brainard, 1997; Pelli, 1997) de MATLAB R2019b (MathWorks, Natick, MA, USA). Les stimuli apparaissaient sur un écran de 30 pouces (DELL ULTRASHARP) avec une résolution de 2560×1600 pixels. Les participants posaient leur tête sur une mentonnière placée à 72 cm de l'écran afin de conserver une taille angulaire des stimuli fixe pour chaque participant. Dans cette expérience, chaque essai débutait par une croix de fixation apparaissant au centre de l'écran pendant 500 ms, qui était suivie par deux images du même objet affichées à gauche et à droite de l'écran sur un fond gris (luminance moyenne de 0.5). Pour la moitié des participants, la tâche consistait à indiquer quel objet était le plus flou et pour l'autre moitié, quel objet était le plus net à l'aide des touches 1 et 2 du clavier d'ordinateur (Figure B1).

Parmi les deux objets présentés, l'un apparaissait à un niveau de flou référence de 3, 8, 13, 18 ou 23 et sa position était contrebalancée entre l'objet de droite et l'objet de gauche. Le niveau de flou de l'autre objet correspondait à la valeur du niveau de flou référence + la valeur du seuil de discrimination prédit pour ce niveau de flou référence ± 3 niveaux de flou. Par exemple pour un niveau de flou référence de 23, le second objet pouvait apparaître à un niveau de flou de $23 + 5 \pm 3$, soit un niveau de flou compris entre 25 et 31. Cela permettait de limiter la durée de l'expérience en échantillonnant plusieurs niveaux de flou autour du seuil de discrimination prédit pour chaque niveau de flou référence.

L'expérience comptait au total 1400 essais présentés aléatoirement avec une pause tous les 70 essais : (20 objets \times 5 niveaux de flou référence \times 7 différences de flou entre l'objet référence et le second objet \times 2 test-retest) et durait environ une heure. Pour chaque essai, nous avons enregistré quel objet était jugé comme le plus flou.

Figure B1

Illustration du Déroulement de l'Expérience.



Note. A chaque essai, deux versions du même objet apparaissait sur un écran gris. L'un des objets avait un niveau de flou référence de 3, 8, 13, 18 ou 23 et le second objet avait un niveau de flou compris entre le niveau de flou référence + seuil de discrimination prédit ± 3 niveaux de flou. La tâche des participants consistait à indiquer quel objet parmi les deux leur paraissait le plus flou ou le plus net.

Analyse des données

Nous avons fixé le seuil de discrimination pour chaque niveau de flou référence à un taux de 70 % de réponses correctes.

Résultats

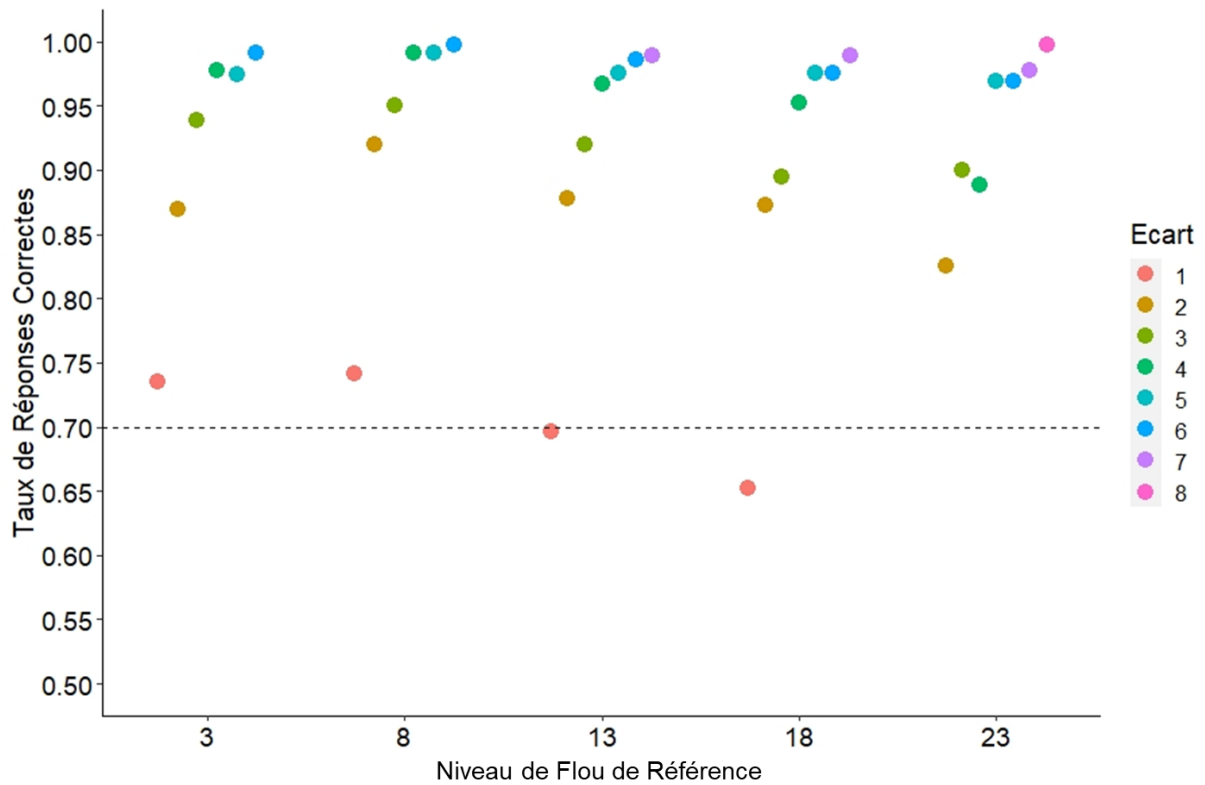
Les taux de réponses correctes moyens pour chaque écart de flou entre l'objet référence et le second objet en fonction du niveau de flou référence sont présentés dans la Figure B2.

Les résultats ont mis en évidence un seuil de discrimination correspondant à 1 niveau de flou pour les niveaux de flou référence de 3 et 8, mais un seuil de discrimination correspondant à 2 niveaux de flou pour les niveaux de flou référence de 13, 18 et 23. Ces résultats suggèrent que pour percevoir une différence de flou entre deux objets, les participants n'ont besoin que d'une différence d'un niveau de flou (i.e., SD du filtre gaussien = 0.3) lorsque les objets apparaissent à un niveau de flou inférieur à 13 tandis qu'ils ont besoin d'une

différence de deux niveaux de flou (i.e., SD du filtre gaussien = 0.6) lorsque les objets apparaissent à un niveau de flou supérieur ou égal à 13.

Figure B2

Taux de Réponses Correctes Moyens pour Juger Quel Objet est le Plus Flou pour Chaque Condition de l'Expérience.



Note. Taux de réponses correctes moyens pour chaque écart de flou entre l'objet ayant le flou de référence et le second objet en fonction du niveau de flou référence. La ligne en pointillés indique le taux de réponses correctes minimum pour le seuil de discrimination.

ANNEXE C – Post-test sur la congruence sémantique et physique des stimuli de l'Expérience 5

Les analyses statistiques de l'Expérience 5 ont mis en évidence que l'interaction entre la congruence du sample et la fiabilité du signal observée dans nos modèles mixtes n'incluant pas les effets aléatoires des stimuli disparaissait lorsque ceux-ci étaient inclus dans nos modèles. Cela suggère que la variabilité liée au stimulus incluse dans ces analyses supplémentaires masque les effets fixes globaux testés car ces effets fixes n'étaient pas assez importants par rapport à la variabilité des stimuli pour atteindre le seuil de significativité. Nous avons donc mené des analyses complémentaires sur les stimuli afin mieux comprendre ce qui pourrait expliquer cette variabilité liée aux stimuli. En effet, il est possible que les paires objets-contexte que nous avons considérées comme congruentes sémantiquement ne soient pas jugées comme telles par les participants. Il est également possible qu'elles ne soient pas suffisamment plausibles physiquement pour être jugées comme congruentes. Par exemple, l'objet dépanneuse qui est placé sur une autoroute pour la condition de contexte congruent (cf. Figure 2.17) n'est pas dans le bon sens de circulation et sa taille n'est pas physiquement plausible dans le contexte de cette scène particulière. Il est donc possible que les participants aient considéré cette paire dépanneuse-autoroute comme un stimulus incongruent au lieu de congruent. Nous avons donc mis en place une expérience contrôle où nous avons demandé à des participants n'ayant jamais vu les stimuli de juger leur congruence sémantique mais également physique. Nous avons ensuite réanalysé les données de l'Expérience 5 en tenant compte de ces jugements pour considérer les stimuli comme congruents ou incongruents.

Méthode

Participants et Design

Dix-neuf étudiants en licence de psychologie de l'Université Grenoble Alpes (11 femmes, $M_{\text{age}} = 24$, $SD_{\text{age}} = 5.20$) ayant une vision normale ou corrigée à la normale ont participé à l'expérience.

Stimuli

Les stimuli étaient les 40 images composées d'un objet en contexte (20 objets en contexte congruent sémantiquement et 20 objets en contexte incongruent sémantiquement) utilisées pour l'Expérience 5.

Procédure

L'expérience a été construite et diffusée en ligne à l'aide de lab.js (Henninger et al., 2020). Les stimuli apparaissaient sur l'écran de l'ordinateur personnel du participant.

À chaque essai, une image d'un objet en contexte congruent ou incongruent sémantiquement apparaissait au centre de l'écran sur un fond gris (luminance moyenne de 0.5). Les participants devaient dans un premier temps évaluer la congruence sémantique entre l'objet et son contexte (« À quel point la présence de cet objet paraît-elle plausible à vos yeux dans ce contexte ? »), puis dans un deuxième temps la congruence physique (« À quel point la taille et la position de l'objet sont-elles plausibles dans ce contexte ? »). Ils pouvaient répondre à l'aide d'un curseur sur une échelle de 10 points : 0 correspondant à une totale incongruence objet-contexte et 10 à une parfaite congruence objet-contexte.

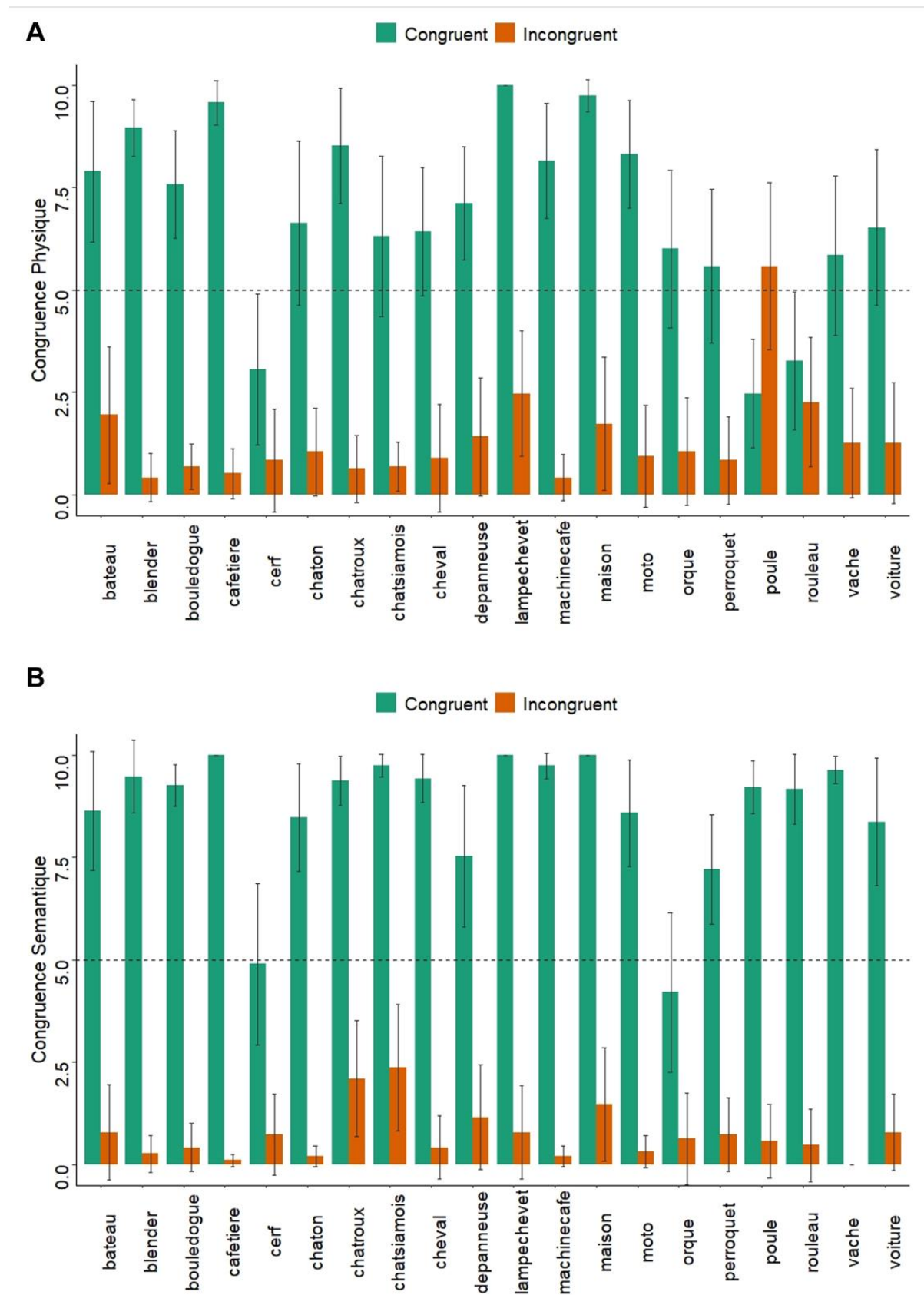
Une session expérimentale comprenait 40 essais (20 objets \times 2 conditions de congruence sémantique du contexte (contexte congruent vs. incongruent) avec une évaluation de la congruence sémantique et physique au sein de chaque essai. L'expérience durait environ 10 minutes. Pour chaque essai, nous avons enregistré le niveau de congruence physique et sémantique de 0 à 10 attribué à chaque stimulus.

Analyse des données

Les données concernant la congruence sémantique et physique évaluée par les participants pour chacune des 40 images d'objet en contexte précédemment classées comme congruentes et incongruentes sémantiquement dans l'Expérience 5 sont présentées dans la Figure C1.

Figure C1

Niveau de Congruence Physique et Sémantique Pour Chaque Stimulus de l'Expérience 5.



Note. A. Niveau de congruence physique attribué à chaque objet classé comme congruent ou incongruent sémantiquement avec son contexte dans l'Expérience 5. Zéro correspond à une totale incongruence et 10 à une totale congruence objet-contexte. La ligne en pointillés correspond à la neutralité, à savoir un stimulus qui n'est ni considéré comme congruent ni comme incongruent. Les barres d'erreurs représentent l'intervalle de confiance à 95% sur tous les individus. B. Niveau de congruence sémantique attribué à chaque objet classé comme congruent ou incongruent sémantiquement avec son contexte dans l'Expérience 5. Zéro correspond à une totale incongruence et 10 à une totale congruence objet-contexte. La ligne en pointillés correspond à la neutralité, à savoir un stimulus qui n'est ni considéré comme congruent ni comme incongruent. Les barres d'erreurs représentent l'intervalle de confiance à 95% sur tous les essais et individus.

Nous avons considéré les stimuli ayant un intervalle de confiance dont les deux bornes étaient strictement supérieures à 5 pour la congruence physique et sémantique comme congruents. Les stimuli ayant un intervalle de confiance dont les deux bornes étaient strictement inférieures à 5 pour la congruence physique et sémantique ont été considérés comme incongruents. Les stimuli dont les deux bornes de l'intervalle de confiance pour la congruence physique ou sémantique n'étaient pas strictement supérieures ou inférieures à 5 (i.e., un stimulus considéré ni comme congruent, ni comme incongruent) ont été définis comme ambigus et ont été écartés des analyses statistiques. Les stimuli ambigus étaient donc le chaton congruent, le chat siamois congruent, le cheval congruent, l'orque congruente, le perroquet congruent, la vache congruente, la voiture congruente, la poule incongruente et le cerf congruent. Certains stimuli considérés comme congruents dans l'Expérience 5 ont été reclassés d'après les résultats de ce post-test car finalement, la congruence sémantique qui leur avait été fixée dans l'Expérience 5 ne correspondait pas à la congruence attribuée par les participants. Sur la base des résultats de la congruence sémantique, la poule congruente et le rouleau congruent ayant un intervalle de confiance strictement inférieur à 5, ils ont été reclassés comme stimuli incongruents.

Nous avons ensuite renouvelé les analyses statistiques de l'Expérience 5 avec les stimuli congruents et incongruents reclassés et les stimuli ambigus supprimés. Nous avons à nouveau utilisé la méthode proposée par Bates, Kliegl, et al. (2015) pour construire des modèles mixtes parcimonieux limitant les problèmes de convergence. L'inspection visuelle des graphiques de résidus n'a pas révélé de déviations évidentes de l'homoscédasticité ou de la normalité. Les p-valeurs ont été obtenues par approximation de Satterthwaite avec le package *lmerTest*

(Kuznetsova et al., 2017). Le seuil de significativité a été fixé à .025 (.05/2 = .025) après correction de Bonferroni pour les deux tests non orthogonaux réalisés. Les tailles d'effet ont été estimées à l'aide du d_z de Cohen (Lakens, 2013).

Résultats

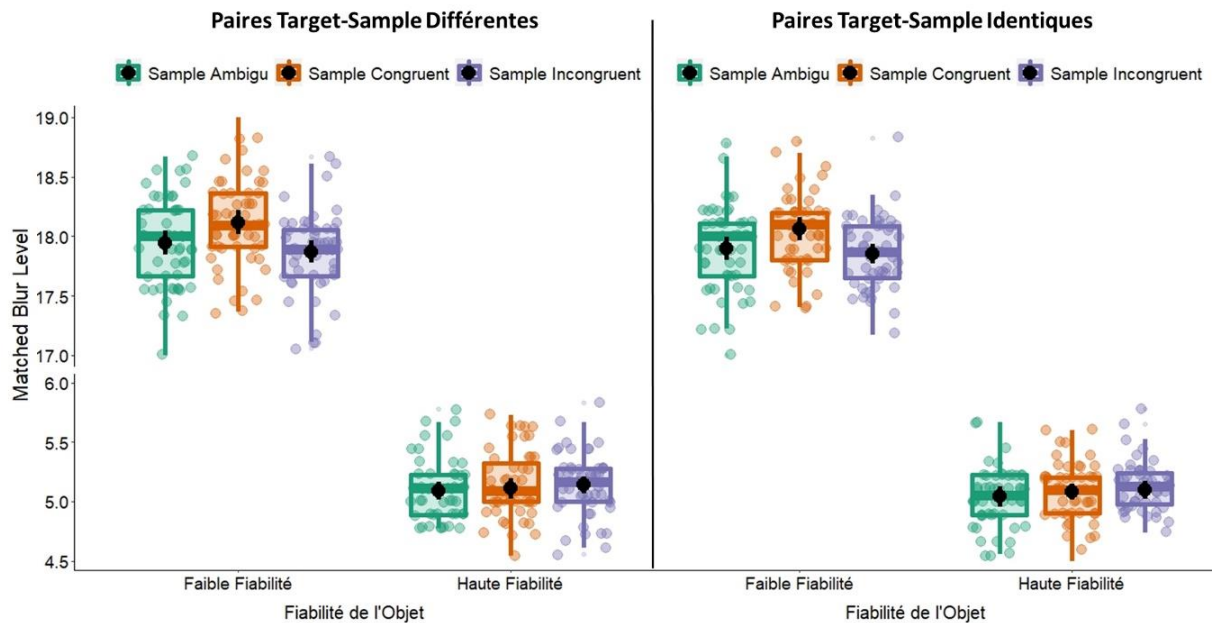
Les résultats n'ont révélé aucun effet de la congruence du sample pour les paires target-sample différentes, ni en condition de haute fiabilité (objet-contexte congruents : $M = 5.10$, $SE = 0.10$; objet-contexte incongruents : $M = 5.12$, $SE = 0.10$), $t(8.99) = -0.61$, $p = .555$, $d_z = 0.084$, ni en condition de faible fiabilité (objet-contexte congruents : $M = 18.10$, $SE = 0.14$; objet-contexte incongruents : $M = 17.86$, $SE = 0.13$), $t(9) = 2.26$, $p = .051$, $d_z = 0.310$. De plus, l'interaction entre la fiabilité du signal de l'objet, le type de paire target-sample et la congruence du sample n'était pas significative, $t(11.05) = 0.34$, $p = .741$, $d_z = 0.047$.

Par ailleurs, les résultats ont montré un effet principal de la fiabilité, $t(34.68) = 301.68$, $p < .001$, $d_z = 41.439$, indiquant que le MBL était plus élevé en condition de faible fiabilité ($M = 17.95$, $SE = 0.14$) que de haute fiabilité ($M = 5.11$, $SE = 0.10$). En revanche, il n'y avait pas d'effet principal significatif de la congruence du sample, $t(12.91) = 2$, $p = .067$, $d_z = 0.275$, ni d'effet principal du type de paire target-sample, $t(9.09) = 0.50$, $p = .630$, $d_z = 0.068$, ni d'interaction entre le type de paire target-sample et la fiabilité, $t(9.21) = -0.41$, $p = .690$, $d_z = 0.056$, ni d'interaction entre le type de paire target-sample et la congruence du sample, $t(10.91) = 0.08$, $p = .941$, $d_z = 0.010$, ni d'interaction entre la fiabilité et la congruence du sample, $t(10.26) = 2.27$, $p = .046$, $d_z = 0.312$.

Bien que nous ayons enlevé les stimuli considérés comme ambigus de nos analyses, nous avons regardé descriptivement leurs patterns de résultats. Les MBL moyens pour chaque condition expérimentale, y compris la condition de stimuli ambigus, sont donc présentés dans la Figure C2. Nous remarquons qu'en condition de haute fiabilité, le MBL des stimuli ambigus se trouve entre le MBL des stimuli congruents et celui des stimuli incongruents. Le fait que les stimuli ambigus ne soient donc ni considérés comme congruent ni comme incongruent se retrouve donc en quelque sorte dans les résultats.

Figure C2

Matched Blur Level Moyen pour Chaque Condition de l'Expérience.



Note. Niveau de flou moyen attribué aux stimuli en sample (Matched Blur Level) reclassés comme congruents, incongruents et ambigus en condition de faible et de haute fiabilité pour les paires target-sample différentes (à gauche) et identiques (à droite). Un niveau de flou élevé correspond à un objet très flou, tandis qu'un niveau de flou faible correspond à un objet relativement net. Les points noirs et les barres d'erreur indiquent respectivement la moyenne et l'erreur standard de tous les participants. Les points en couleur correspondent aux observations individuelles.

ANNEXE D – Article Rossel et al. (2023)



Subjective perception of objects depends on the interaction between the validity of context-based expectations and signal reliability

Pauline Rossel, Carole Peyrin, Louise Kauffmann*

Univ. Grenoble Alpes, CNRS, LPNC, 38000 Grenoble, France

ARTICLE INFO

Keywords:

Object perception
Predictive coding
Expectations
Opposing process theory
Sharpening
Dampening

ABSTRACT

Predictive coding theories of visual perception assume that expectations based on prior knowledge modulate the processing of information. However, the underlying mechanisms remain debated. Some accounts propose that expectations enhance the perception of expected relative to unexpected stimuli while others assume the opposite. Recently, the opposing process theory suggested that enhanced perception of expected vs. unexpected stimuli may occur alternatively depending upon the reliability of the visual signal. When the signal is noisy, perception would be biased toward what is expected since anything else may be too noisy to be resolved. When the signal is unambiguous, perception would be biased toward what diverges from expectations and is more informative. Our study tested this hypothesis, using a perceptual matching task to investigate the influence of expectations on the perceived sharpness of objects in context. Participants saw two blurred images depicting the same object and had to adjust the blur level of one object to match the blur level of the other one. We manipulated the validity of expectations about objects by varying their scene context (congruent or incongruent context leading to valid or invalid expectations about the object). We also manipulated the reliability of the visual signal by varying the initial blur level of object pairs. Results showed that expectations validity differentially affected the perception of objects depending on signal reliability. Perception of validly expected objects was enhanced (sharpened) relative to unexpected objects when visual inputs were unreliable while this effect reversed to the benefit of unexpected objects when the signal was more reliable.

1. Introduction

Our visual system is constantly exposed to abundant streams of information which can often be noisy or ambiguous due to signal constraints (e.g., masking of objects, dim light or foggy weather) or sensory loss (e.g., in ageing or visual pathologies). To rapidly make sense of it and adequately react, current models of visual perception assume that we constantly rely on prior knowledge and expectations built upon regularities learnt from past perceptual experiences. This assumption is supported by a large body of research which for example showed improved performance during the processing of expected than unexpected visual stimuli (see for example Brandman & Peelen, 2017; Davenport, 2007; Davenport & Potter, 2004; Greene et al., 2015; Joubert et al., 2008; Lai et al., 2020; Palmer, 1975; Roux-Sibilon et al., 2019 in the context of object recognition).

Interestingly, recent studies suggest that expectations and prior knowledge can also qualitatively affect the content of visual perception and the appearance of visual stimuli (Han & VanRullen, 2016; Hansen

et al., 2006; Lupyán, 2017; Perez et al., 2020; Rossel et al., 2022). These studies addressed this question using tasks that directly tap into subjective perception while minimising the influence of non-perceptual factors (e.g., decisional or semantic factors). They for example used perceptual matching tasks or equal/same judgments based on different visual properties (e.g., colour, blur or brightness) of visual stimuli that were associated or not with prior knowledge (e.g., indicate whether silhouettes pairs of a meaningful and a meaningless object composed of the same basic shapes are equally blurred or not; Perez et al., 2020). These studies consistently showed that although objectively similar in their manipulated perceptual properties, stimuli that could be expected or associated with prior knowledge were subjectively perceived as different from stimuli that could not. Furthermore, expected stimuli tended to be perceptually enhanced. For example, Lupyán (2017) used a perceptual matching task in which participants had to adjust the blur level of a letter string (a Sample stimulus) in order to match the blur level of another letter string (a Target stimulus). Letter strings could be either meaningful words (e.g., “much”) or meaningless pseudowords

* Corresponding author.

E-mail address: louise.kauffmann@univ-grenoble-alpes.fr (L. Kauffmann).

composed of the same letters in a randomised order (e.g., “mchu”). Results revealed that participants attributed more blur than necessary to match a Sample word to a Target pseudoword and not enough blur to match a Sample pseudoword to a Target word. These results suggest that at an objectively equal level of blur, meaningful stimuli such as words for which we have semantic knowledge, were subjectively perceived to be sharper than meaningless pseudowords, and that expectations based on prior semantic knowledge influenced how we perceive.

We recently adapted this perceptual matching paradigm on the perceived sharpness of objects (Rossel et al., 2022). Two blurred versions of the exact same object (e.g., a car), one being the Target and the other one the Sample, appeared on a screen. Participants had to adjust the blur level of the Sample object to match the blur level of the Target object. We manipulated the predictability of the two objects so that one of them was surrounded by an intact context leading to valid expectations about the object (i.e., a highway context - predictable condition) while the other one was surrounded by a meaningless context (e.g., phase-scrambled version of the highway context - unpredictable condition). Similar to Lupyan (2017)'s findings, results showed that participants attributed more blur to a predictable Sample object matched to an unpredictable Target object than to an unpredictable Sample object matched to a predictable Target object. This means that, at a physically identical blur level, objects that could be predicted based on their intact context were subjectively perceived as sharper than objects that could not (i.e., scrambled context).

Enhanced perception of expected stimuli in these studies is consistent with previous findings of higher perceptual sensitivity to expected than unexpected features of visual stimuli (e.g., Pinto et al., 2015; Stein & Peelen, 2015; Teufel et al., 2018; Wyart et al., 2012), giving rise to the so-called “Sharpening” account of predictive processing theories. According to this account, more weight is attributed to the processing of expected relative to unexpected features of sensory inputs, such that the content of perceptual processing would be biased toward what is consistent with expectations (de Lange et al., 2018; Kaiser et al., 2019; Kok et al., 2012; Lee & Mumford, 2003). Given the highly probabilistic structure of the environment, such a mechanism would enable to form accurate percepts in the vast majority of cases. Enhanced perception of expected stimuli in the above-mentioned studies however appears to contradict an alternative account of the predictive processing framework (so-called “Dampening” account) which posits that more weight is attributed to the processing of – more informative – unexpected features, such that the processing of expected features would be relatively dampened. This account has been mainly supported by research in the action literature showing that expected action's outcomes tend to be perceptually suppressed (e.g., Blakemore et al., 1998). In the vision domain, a few studies did suggest dampened processing of expected relative to unexpected visual stimuli (Hollingworth & Henderson, 2000; LaPointe et al., 2013; Mack et al., 2017; Spaak et al., 2022). For example, Spaak et al. recently showed that object-related changes in visual scenes were better detected when objects were unexpected than expected and unexpected objects were also better identified than expected ones. However, these studies did not directly involve subjective perceptual reports and it remains unclear whether surprising stimuli can be perceptually enhanced.

Although in apparent contradiction, the sharpening and dampening accounts may not necessarily be mutually exclusive. Recent theoretical developments, the opposing process theory (Press et al., 2020), propose that both expected and unexpected stimuli can be perceptually enhanced alternatively according to time and signal constraints. Under this view, perception would be initially biased toward what is expected, allowing to confirm valid expectations. The processing of unexpected features would then be subsequently boosted allowing to update expectations (e.g., Yon & Press, 2017). This sequence would be further constrained by the quality (or reliability) of sensory inputs. Under this view, enhanced processing of expected features would be mostly useful when sensory inputs are noisy or ambiguous (as is the case for blurred

stimuli), allowing to increase the signal-to-noise ratio. Perception would therefore be biased toward what is expected since anything else may be too noisy to be resolved. On the contrary, when sensory inputs are clear and unambiguous, there would be less point in favouring the processing of expected signal features. Indeed, it would be more relevant to favour the processing of what does not conform to expectations and what is therefore more informative. In this case, perception should therefore be biased towards what diverges from expectations. This is also in line with other accounts according to which the relative weight of predictions and sensory signals are weighted by their respective reliability (Feldman & Friston, 2010; Friston, 2005). Within this opposing process framework, enhanced perception of expected visual stimuli reported in past studies could therefore be explained by the fact that the stimuli used were generally noisy (e.g., blurred; Lupyan, 2017; Perez et al., 2020; Rossel et al., 2022) while enhanced processing of unexpected stimuli was observed when visual signals were unambiguous (e.g., Spaak et al., 2022).

However, the interactions between prediction's validity and signal reliability and their consequences for subjective perception have, to our knowledge, never been systematically addressed within the same study. Indeed, the above-mentioned studies investigating perceptual effects of expectations either used ambiguous stimuli (e.g., blurred; Lupyan, 2017; Perez et al., 2020; Rossel et al., 2022) or unambiguous ones (e.g., Spaak et al., 2022) but did not directly compare the two conditions. Furthermore, studies using ambiguous stimuli only manipulated the presence vs. absence of prior knowledge or relevant information allowing to form expectations. However, they did not address the effect of expectations validity by including a condition in which the visual signal could be surprising and go against expectations (i.e., a meaningful but incoherent/surprising stimulus). A comprehensive understanding of these mechanisms is however important to better understand the determinants of subjective perception of expected/unexpected visual stimuli under various constraints and how this may in turn influence behaviour. These questions also more broadly tackle the long-standing debate on whether (and how) cognition affects perception (Firestone & Scholl, 2016; Pylyshyn, 1999).

The present behavioural experiment aimed to address these issues and explicitly test the assumption of the opposing process theory by comparing how the subjective perception of ambiguous and unambiguous stimuli can vary according to their expected versus surprising status. To this end, we adapted the perceptual matching task used in our previous study (Rossel et al., 2022). Two blurred objects (a Target and a Sample) appeared simultaneously on the right and left sides of the screen. The task consisted in adjusting the blur level of the Sample object to match the blur level of the Target object. In contrast to our previous study, both the Target and Sample objects were surrounded by an informative scene context. We manipulated the validity of predictions about the objects by varying the congruence between the objects and their scene contexts. One of the objects was surrounded by a congruent scene context (leading to valid predictions about the object) while the other one was surrounded by an incongruent context (leading to invalid predictions). In the present study, we also manipulated the reliability of the visual signal by varying the initial blur level of object pairs, so that both objects could have either a high signal reliability (i.e., relatively sharp objects) or a low signal reliability (i.e., blurry objects). Based on the opposing process theory (Press et al., 2020), we expected that at a physically equal level of blur, participants should perceive objects in congruent contexts as sharper than the same objects in incongruent contexts (i.e. perceptual enhancement of expected stimuli) when the objects are blurry (low signal reliability) while the opposing effect should be observed when objects are clearer (i.e., objects in incongruent contexts should be perceived as sharper than objects in congruent contexts – perceptual enhancement of unexpected stimuli). As a consequence, participants should compensate this perceived difference between objects by attributing a higher blur level to congruent than incongruent Samples objects in the low reliability condition while they

should attribute a higher blur level to incongruent than congruent Sample objects in the high reliability condition.

Additionally, it has been argued that if high-level factors (such as expectations based on contextual information) influence the subjective perception of objects, such perceptual effects should equally affect all objects in the visual display. For example, if an expected object in a given context is perceived as sharper than it actually is, simultaneously displayed identical objects sharing the same context should equally benefit from this perceptual enhancement, resulting in the absence of perceived blur difference between them (cf. “El Greco fallacy”, Firestone & Scholl, 2014 – but see Dubova & Goldstone, 2022 for recent empirical arguments against this view). A secondary objective of this study was therefore to test this hypothesis. To this end, we also included a condition in which the Target and Sample object shared the same context (either congruent or incongruent with the object). We expected the perceived difference in blur between the Target and Sample objects in these trials in both signal Reliability conditions to be reduced relative to trials in which the Target and Sample object had a different context.

2. Method

2.1. Participants

Fifty-three undergraduate students of Psychology from University Grenoble Alpes (50 women, $M_{\text{age}} = 21.43$, $SD_{\text{age}} = 5.54$) with normal or corrected to normal vision participated in the experiment. The sample size was set based on an estimated effect size of $d_z = 0.45$ from the simple effect of objects predictability on blur perception for Target-Sample pairs with different contextual information (i.e., comparison between -1 – the blur level attributed to a Sample object with uninformative contextual information matched to a Target object with uninformative contextual information and -2 – the blur level attributed to a Sample object with uninformative contextual information matched to a Target object with informative contextual information) in Experiment 1 of Rossel et al. (2022) to achieve a power (i.e., corresponding to $1 - \beta$, with β being the risk of Type-2 error) of 0.8 with an alpha level of 0.025 after Bonferroni correction for the two planned non-orthogonal analyses ($0.05/2 = 0.025$). The study was approved by the ethics committee of the University Grenoble Alpes (CER-Grenoble Alpes, COMUE University Grenoble Alpes, CERGA-Avis-2020-4) and was conducted in agreement with the Declaration of Helsinki for experiments involving human subjects. All participants involved in the study gave their informed written consent before taking part in the experiment. They received course credits for their participation.

2.2. Stimuli

The stimuli consisted of 20 colour photographs of scenes (10 indoor and 10 outdoor scenes) and 20 colour photographs of objects (10 animated and 10 non-animated objects). The photographs were downloaded from Pixabay (<https://pixabay.com/fr/>), a Creative Commons Zero (CC0) licensed photo sharing site, and from the Google Images database by filtering for the CC0 licence. The stimuli were created using MATLAB R2019b (The Mathworks, Natick, MA).

The original photographs of the scenes were first cropped to a square format and resized to 700×700 pixels. They were converted to 256-level grey-scale images by averaging the values of the three colour channels for each pixel. The average luminance of each image was equalised to 0.5 (for pixel intensity values between 0 and 1) and the average Root Mean Square (RMS) contrast value of all images (0.1947) was applied to each image. For each scene image, we then constructed a 700×700 pixels disc containing the scene.

The original photographs of the objects were also cropped to a square format and resized to 120×120 pixels which corresponded to the maximum size of objects (either in width or height). The objects were then isolated from their background by outlining them on GIMP

software (The GIMP Development Team. (2019), 2019) and applying transparent pixels to their background. They were then converted to greyscale images. The average luminance and RMS contrast of each object were also equalised to 0.5 (for pixel intensity values between 0 and 1) and the average RMS contrast value of all images (0.2117) was applied to each image.

Each object was then blurred using 2D Gaussian filters (*imgaussfilt* function in Matlab) of increasing size (i.e., with increasing standard deviation - SD) to obtain 22 versions of the object ranging from a sharp object (close to intact; blur level 1, $SD = 1$) to a very blurred and hardly recognisable object (blur level 22, $SD = 10$, see Rossel et al., 2022 for a similar procedure). We then assembled the object and scene images to construct final stimuli containing a central object surrounded by the circular scene context. To this end, we placed at the centre of each scene a 250×250 pixels grey disc of luminance 0.5 that was spatially filtered by a Gaussian with a standard deviation of 2 to smooth the edge of the disc. We then added the blurred object to the centre of the grey disc. This grey disc allowed us to isolate the object from the background and avoid local contrast differences between the object and the scene in the different conditions.

The object and scene images were paired so that each object appeared either surrounded by a semantically congruent scene (e.g., a boat in a harbour scene), an experimental condition that should lead to valid predictions about the object, or surrounded by a semantically incongruent scene (e.g., a boat in a kitchen scene), an experimental condition that should lead to invalid predictions about the object (Fig. 1).

2.3. Procedure

The experiment was constructed using the Psychtoolbox (Brainard, 1997; Pelli, 1997) implemented in MATLAB R2019b (MathWorks, Natick, MA, USA). Stimuli were displayed on a 30" monitor (DELL UL-TRASHARP) with a resolution of 2560×1600 pixels. Participants put their head on a chin rest 72 cm away from the screen so that the grey disc containing the object corresponded to an angular size of 5° visual angle, the object corresponded to a maximal angular size of 2.4° visual angle and the circular scene corresponded to an angular size of 14° visual angle.

On each trial, two images containing the same object were displayed on the left (the Target) and right (the Sample) sides of the screen against a grey background (average luminance of 0.5). To test the effect of visual signal reliability on the perceived sharpness of objects, the Target object appeared in the first half of the trials at a blur level of 18 (SD of Gaussian filter = 7.6, low signal reliability) and in the second half of the trials at a blur level of 5 (SD of Gaussian filter = 2.2, high signal reliability). In a pilot study, we found that the minimum difference in blur required to perceive a blur difference between two objects was of 0.3 SD when the objects had a blur level of less than 4.6 SD (as in the high reliability condition) and of 0.6 SD when the objects had a blur level of 4.6 SD or more (as in the low signal reliability condition). The size of Gaussian filters used to obtain the blur levels of Target and Sample therefore varied in increments of 0.3 SD in the high signal reliability condition and of 0.6 SD levels in the low signal reliability condition. This resulted in a total of 22 blurred versions of each object used in the experiment. We did not counterbalance the order of the reliability conditions to avoid that the prior perception of sharp objects of the high reliability condition could be later used to disambiguate the signal of blurred objects of the low reliability condition if performed next. Concerning the initial blur level of the Sample object which had to be matched to the Target, it varied randomly between three blur levels above or below the blur level of the Target object. However, the levels varied according to the signal reliability condition of the Target (low vs. high). Indeed, it is possible that in the low reliability condition of the Target object (i.e., when the object is very blurred and therefore ambiguous), having the same object in Sample at a medium blur level and therefore better identifiable,

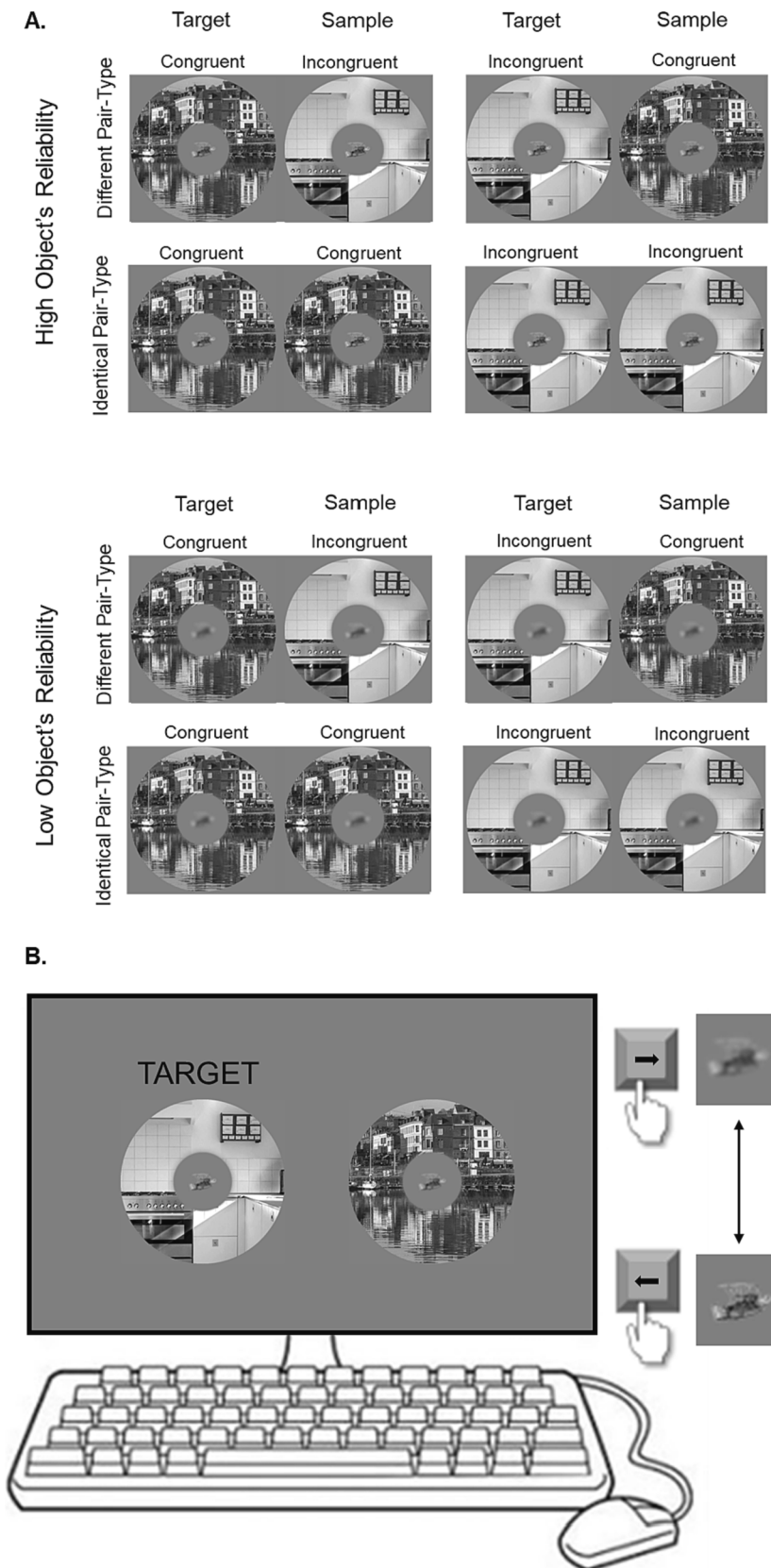


Fig. 1. Illustration of the experimental conditions and the display. **A.** Example of stimuli presented in each experimental condition. For each trial, the Sample and Target objects were the same but appeared either at a low blur level where the object was fairly sharp (high reliability condition) or at a high blur level making the object identity ambiguous (low reliability condition). The scene context could be either congruent with the object (correct/valid predictions) or incongruent with the object (incorrect/invalid predictions). Furthermore, the scene contexts of the Target and the Sample could be either identical or different. **B.** Illustration of the display in a trial with an incongruent Target and a congruent Sample (different Target-Sample pair-type). Two blurred objects appeared simultaneously, the Target on the left and the Sample on the right. Participants were asked to adjust the Sample object's blur level until they judged both objects as equally blurred.

would bias the processing of the Target which would then become less ambiguous, even if very blurred. To avoid this, the initial blur level of the Sample object remained close to the one of the Target object in the low reliability condition. Therefore, the Sample object could have a blur level between 16 (SD = 5.5) and 21 in the low signal reliability condition and a blur level between 3 (SD = 4.6) and 8 (SD = 3.1) in the high signal reliability condition. Participants had to adjust the blur level of the Sample object to match the blur level of the Target object by increasing or decreasing the blur of the Sample object in steps of one blur level in the high reliability condition and in steps of two blur levels in the low reliability condition, using the left and right arrows on the keyboard respectively. In the low reliability condition, the Sample object could be adjusted to a maximum blur corresponding to level 22 (SD = 10) and a minimum blur corresponding to level 15 (SD = 5.8) and in the high reliability condition, the Sample object could be adjusted to a maximum blur corresponding to level 14 (SD = 5.2) and a minimum blur corresponding to level 1 (SD = 1). Participants were asked to press the spacebar to validate their response once they judged that the objects in the Sample and Target were equally blurred. Participants were encouraged to rely on their subjective judgement to match the blur level of the objects and were not limited in time at each trial.

It can be noted that in the present study, blur was used both as the signal to be processed to perform the matching task and as a way to modulate object reliability. Given that our main assumption was that the exact same blurred object can be subjectively perceived as different according to its contextual information, we needed to ensure that any perceived difference in blur between object pairs could not be attributed to other objective differences between them. For example, manipulation of other perceptual variables (e.g., contrast) as a way to modulate signal reliability may have unevenly affected object perception at different blur levels. Furthermore, manipulation of blur was preferred to other perceptual variables (e.g., brightness or orientation) because blur is a fundamental feature of vision that can be easily transposed to ecological conditions (e.g., small or distant objects in our visual field may initially appear blurry before they can be resolved as we get closer or change viewpoint).

For each trial, the objects in Target and Sample had either a different scene context (one congruent and one incongruent with the object, i.e., Different Target-Sample pair-types) or the same scene context (both congruent or incongruent with the object, i.e., Identical Target-Sample pair-types). The Different Target-Sample pair-types allowed us to test our main hypothesis of opposing effects of prediction validity depending on signal reliability on the perceived sharpness of the objects. The Identical Target-Sample pair-types were included to also assess potential differences in perceived sharpness during the matching of identical Sample and Target stimuli.

An experimental session contained 160 trials divided into two blocks of 80 trials: the 80 trials where the object had low reliability always appeared first and were followed by the 80 trials where the object had high reliability. Within each block, we had 20 objects \times 2 conditions of semantic context congruence for the Sample object (congruent vs. incongruent context) \times 2 types of Target-Sample pairs (identical Target-Sample pair-types vs. different Target-Sample pair-types). The experiment lasted approximately 30–40 min, depending on the speed at which participants adjusted the blur of the Sample object, and contained a break every 40 trials. For each trial, we recorded the blur level attributed to the Sample object to match the Target object (Matched Blur Level, MBL).

3. Data analysis

We measured the level of blur that participants attributed to the Sample to judge the Target and the Sample as equally blurred (Matched Blur Level, MBL). Data were analysed using R software (R Core Team, 2019) and *lme4* package (Bates, Mächler et al., 2015) to perform a linear mixed-effects analysis of the effect of signal Reliability (high vs. low),

Target-Sample pair-type (identical vs. different), and Congruence between the Sample object and its context (congruent vs. incongruent) in a $2 \times 2 \times 2$ within-subject design. The signal Reliability, the Target-Sample pair-type, the congruence of the Sample and their interaction terms were introduced into the model as fixed effects. The intercepts for subjects, as well as the random slopes per participant for the effect of signal Reliability, Target-Sample pair-type, Sample congruence and their interactions were specified as random effects.

We conducted the minimum number of analyses needed to test our hypotheses to avoid increasing the risk of Type-1 error. According to opposing process theory (Press et al., 2020), enhanced perception of expected relative to unexpected objects should occur when the stimuli are ambiguous and the visual signal is unreliable while enhanced perception of unexpected relative to expected objects should occur when the signal is more reliable. We therefore expected that in the low reliability condition, congruent objects would be subjectively perceived as sharper than incongruent objects at objectively equal blur levels. Therefore, for Target-Sample (T/S) pairs of different contexts in the low reliability condition, the MBL of Sample objects in the T-incongruent/S-congruent condition should be higher than the MBL of Sample objects in the T-congruent/S-incongruent condition. On the other hand, we expected that in the high reliability condition, incongruent objects would be subjectively perceived as sharper than congruent objects at an objectively equal blur level. Therefore, for Target-Sample pairs of different contexts in the high reliability condition, the MBL of Sample objects in the T-congruent/S-incongruent condition should be higher than the MBL of Sample objects in the T-incongruent/S-congruent condition. To address this main hypothesis, we therefore tested the interaction between the Congruence of Sample objects and the signal Reliability for Different pair-types, as well as the difference between the MBL of congruent and incongruent Sample objects in each signal Reliability conditions (i.e., two non-orthogonal a priori tests).

Furthermore, if expectations based on congruent or incongruent contextual information influence the perceived sharpness of objects (through the relative enhancement of either expected or unexpected features), such perceptual effects should also occur during the matching of identical Target-Sample pairs (i.e., both congruent or incongruent stimuli). However, they should equally affect the Target and Sample objects (Firestone & Scholl, 2014). Therefore, the perceived difference in blur between the Target and Sample objects in the Identical pair-types (and hence the difference in MBL between T-congruent/S-congruent and T-incongruent/S-incongruent conditions) in both objects' Reliability conditions should be reduced (or even absent) relative to Different pair-type trials. To address this hypothesis, we thus also tested the three-way interaction between the object's Reliability, the Target-Sample pair-type and the Congruence of the Sample.

We used the method proposed by Bates, Kliegl et al. (2015) to build parsimonious mixed models limiting convergence problems. Visual inspection of the residual plots did not reveal any obvious deviations from homoscedasticity or normality. The p-values were obtained by Satterthwaite approximation with the *lmerTest* package (Kuznetsova et al., 2017). The significance threshold was set at 0.025 ($0.05/2 = 0.025$) after Bonferroni correction for the two non-orthogonal tests performed (the Sample congruence effect in the low and high reliability conditions for the different Target-Sample pair-types). Effect sizes were estimated using Cohen's d_z (Lakens, 2013).

4. Results

Mean MBLs for each experimental condition are shown in Fig. 2. Given that the blur level of the Target object could be of 5 or 18 in the high and low reliability conditions, respectively, a mean MBL above (below) these values in each signal Reliability condition would indicate an overestimation (underestimation) of the Sample's blur relative to the Target.

Results revealed first a main effect of signal Reliability, $t(51.99) =$

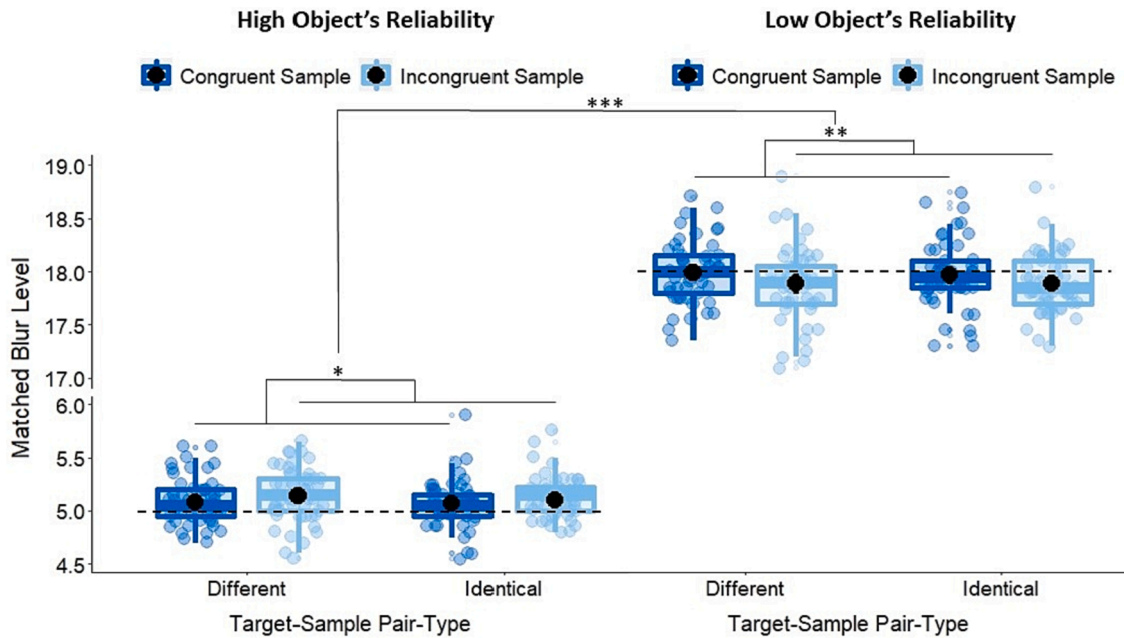


Fig. 2. Mean Matched Blur Level (MBL) for each condition. Mean MBL for congruent and incongruent Samples under different and identical Target-Sample pair-types conditions for high (left) and low (right) reliability. A high blur level corresponds to a very blurred object, while a low blur level corresponds to a relatively sharp object. The y-axis is discontinued for illustrative purposes. Black dots and error bars indicate mean and standard error over all individuals and trials, respectively. The dashed lines correspond to the perfect matches for each condition. Colour dots correspond to individual observations. * $p < .05$. ** $p < .01$. *** $p < .001$.

375.66, $p < .001$, $d_z = 51.601$, indicating that the MBL was higher in the low reliability condition ($M = 17.93$, $SE = 0.13$) than in the high reliability condition ($M = 5.10$, $SE = 0.10$). This was to be expected given that the blur level of the Target was higher in the low than high reliability condition. There was neither a significant main effect of Sample congruence, $t(8368) = 1.26$, $p = .208$, $d_z = 0.173$, nor a significant main effect of Target-Sample pair-type, $t(8368) = 1.13$, $p = .259$, $d_z = 0.155$.

Concerning our main hypothesis, results revealed a significant interaction between Sample congruence and signal Reliability for the different Target-Sample pair-types, $t(4132) = 3.27$, $p = .001$, $d_z = 0.449$. More precisely, in the low reliability condition of different Target-Sample pair-types, we observed a significant effect of Sample congruence, $t(2066) = 2.54$, $p = .011$, $d_z = 0.349$, where participants attributed a higher blur level to congruent Sample objects than incongruent Sample objects (congruent Sample: $M = 17.99$, $SE = 0.14$; incongruent Sample: $M = 17.89$, $SE = 0.14$). On the contrary, in the high reliability condition of different Target-Sample pair-types, participants tended to attribute a higher blur level to incongruent Sample objects than congruent Sample objects but this difference was not significant after Bonferroni correction, $t(2066) = -2.07$, $p = .039$, $d_z = 0.284$ (congruent Sample: $M = 5.08$, $SE = 0.10$; incongruent Sample: $M = 5.14$, $SE = 0.10$).

Concerning our second hypothesis, the three-way interaction between Sample congruence, Target-Sample pair-type and signal Reliability was not significant, $t(8368) = 0.74$, $p = .462$, $d_z = 0.102$, indicating that the interaction between Sample congruence and signal Reliability did not significantly differ in the identical relative to the different Target-Sample pair-types. This pattern of results therefore rather suggests a two-way interaction between the Sample congruence and the signal Reliability irrespective of the Target-Sample pair-type which was indeed significant, $t(8368) = 4.02$, $p < .001$, $d_z = 0.552$. We tested the effect of Sample congruence a posteriori for each of the signal Reliability conditions. In the low reliability condition, we found a significant effect of Sample congruence, $t(4186) = 3.26$, $p = .001$, $d_z = 0.448$, where participants attributed a higher blur level to congruent Sample objects than incongruent Sample objects (congruent Sample: $M = 17.98$, $SE = 0.13$; incongruent Sample: $M = 17.89$, $SE = 0.13$), suggesting perceptual enhancement of expected relative to unexpected

objects. On the contrary, in the high reliability condition, participants attributed a higher blur level to incongruent Sample objects than congruent Sample objects and this time the difference was significant, $t(4186) = -2.34$, $p = .019$, $d_z = 0.321$ (congruent Sample: $M = 5.07$, $SE = 0.10$; incongruent Sample: $M = 5.12$, $SE = 0.10$), suggesting perceptual enhancement of unexpected relative to expected objects.

There was no other significant interaction (interaction between Target-Sample pair-type and signal Reliability: $t(8368) = -0.47$, $p = .637$, $d_z = 0.065$, interaction between Target-Sample pair-type and Sample congruence: $t(8368) = -0.08$, $p = .937$, $d_z = 0.036$).

For completeness, we repeated the same statistical analyses, this time adding the random effects of stimuli. Thus, intercepts for stimuli, as well as the random slopes per stimuli for the effect of signal Reliability, Target-Sample pair-type, Sample congruence and their interactions were specified as random effects in our mixed effect model. Again, visual inspection of the residual plots did not reveal any obvious deviations from homoscedasticity or normality.

After including the random effects of the stimuli, results revealed no significant interaction between Sample congruence and signal Reliability for the different Target-Sample pair-types, $t(18.99) = 1.70$, $p = .106$, $d_z = 0.234$. Furthermore, the three-way interaction between signal Reliability, Target-Sample pair-type and Sample congruence was not significant, $t(19.02) = 0.34$, $p = .738$, $d_z = 0.047$ and there was also no interaction between signal Reliability and Sample congruence, $t(19.26) = 1.74$, $p = .098$, $d_z = 0.239$.

5. Discussion

The present study investigated the effect of predictions validity, based on contextual information, on the subjective perception of objects depending on the reliability of the visual signal. To this end, participants performed a perceptual matching task consisting in adjusting the blur level of an object (Sample) embedded in a congruent or incongruent scene context (leading to valid or invalid predictions about the object, respectively) to match the blur level of the same object (Target) embedded in a congruent or incongruent context. To address the effect of signal reliability, the objects could appear at a low (i.e., high signal

reliability) or high blur level (i.e., low signal reliability). Based on the opposing process theory (Press et al., 2020) postulating that expected stimuli are perceptually enhanced relative to unexpected stimuli when the visual signal is noisy while they should be relatively perceptually dampened when the visual signal is more reliable, we expected that at an objectively equal blur level, congruent objects would be subjectively perceived as sharper than incongruent ones when the blur level was high and that incongruent objects would be perceived as sharper than congruent ones when the blur level was low. However, we expected the perceived difference in blur between the Target and Sample objects in both signal Reliability conditions to be reduced when object pairs had an identical than a different context (Firestone & Scholl, 2014).

Contrary to this assumption, results revealed no significant difference between perceptual effects of expectations observed in the Different and Identical pair-type conditions. This suggests that even when identical, the Target and Sample objects were however perceived as different. The significant two-way interaction between Sample congruence and signal Reliability as well as pairwise post-hoc tests however suggested that the effect of Sample Congruence reversed according to signal Reliability (i.e., congruent Sample objects were perceived as sharper than incongruent ones when signal reliability was low, while the opposite was found when signal reliability was high) irrespective of the pair-type condition - and hence of the congruence of the Target object.

Following Firestone & Scholl's rationale, if "perceptual" effects can still be found when stimuli are identical, such effects may rather reflect the influence of non-perceptual judgements or response biases. In the context of our experiment, objects that make more sense in congruent contexts may for example be more readily judged as sharp or "clear" than objects in incongruent contexts without it being associated with perceptual effects. Such non-perceptual judgements could therefore bias perceptual matching. However, the fact that the effect of Sample congruence was reversed as a function of signal reliability irrespective of the pair-type condition (i.e., crossed interaction between Sample congruence and signal Reliability) makes this interpretation unlikely. Indeed, if our effects were driven by a systematic response bias for example consisting in judging congruent objects as clearer than incongruent ones (and therefore in systematically attributing more blur to congruent than incongruent Sample objects), such a bias should be present in both reliability conditions. However, there is little chance that response biases were also reversed as a function of signal reliability (i.e., that congruent objects were judged as clearer than incongruent ones when signal reliability was low, but were judged as less clear when signal reliability was high) without it being influenced by perceptual effects.

If results in the identical pair-types cannot be attributed to a response bias, how to explain that the pattern of results in this condition was similar to the different pair-type condition? Results of a recent study by Dubova and Goldstone (2022) may provide an interesting lead to answer that question. These authors challenged the assumption that perceptual effects induced by prior knowledge should equally affect all items in the visual field by arguing that during perceptual matching tasks such as ours, the Target and Sample stimuli (even if identical) can hardly be perceived similarly at the same time. In their study, these authors recorded participants' eye movements during a perceptual (colour) matching task such as ours and observed that participants fixated more often on the Sample which was being adjusted than on the Target which therefore mostly appeared peripherally throughout the task. In the theoretical context of the present study, such behaviour could have a critical impact on matching performance and may explain the observed differences even in the Identical pair-type trials. Indeed, our Target and Sample objects had a relatively small size (2.4° of visual angle) and were each surrounded by a large context. When participants fixated one of the objects, the other appeared with an eccentricity of 16.3° degrees and the inner border of its context was at 14.3° . If participants adopted a similar strategy as in Dubova and Goldstone (2022) consisting in fixating on the

Sample object most of the time, this implies that the Target object was perceived with a poorer spatial resolution inherent to peripheral vision. More detrimental, the context of the Target object may also have appeared blurry and therefore less informative or more ambiguous with respect to its congruence with the Target object. Perceptual effects induced by context-based expectations could thus have affected the Target less strongly than the Sample. Our pattern of results does suggest that participants' performance (i.e., matched blur level) varied according to the congruence of the Sample object but not the congruence of the Target object. Another possible explanation for the similar pattern of results in the Identical and Different Target-Sample pair-types may come from the design of our experiment in which the Target's blur level was kept constant throughout the high and low reliability blocks (blur levels of 5 and 18, respectively) while the Sample's blur varied from trial to trial. This could have led participants to create an accurate estimate of the Target blur level in each condition, which became less prone to effects induced by contextual information relative to the Sample's, thus eliminating the Target congruence effect.

On the other hand, the observed effect of the Sample's congruence which was reversed as a function of signal reliability (irrespective of the Target-Sample pair-type) is in line with our hypotheses. In the low reliability condition, participants attributed a higher blur level to expected Sample objects than to unexpected Sample objects. This indicates that at a physically equal blur level between the Sample and Target objects, expected objects subjectively appeared as sharper than unexpected ones. This is also in line with previous works showing (1) increased perceived sharpness of blurred stimuli which can be related to prior knowledge relative to meaningless/unpredictable stimuli (Lupyan, 2017; Perez et al., 2020; Rossel et al., 2022) and (2) that this effect is stronger as the reliability of the visual signal decreases (Rossel et al., 2022). Results of the present study therefore allow us to extend these previous findings by showing that blurred expected stimuli are also perceived as sharper than unexpected ones.

This pattern of results was reversed in the high reliability condition. Participants attributed a higher blur level to the Sample objects with an incongruent scene context than to the Sample objects with a congruent scene context, indicating that at a physically equal blur level between the Sample and Target objects, unexpected objects subjectively appeared as sharper than expected ones. This pattern of results therefore supports the idea of opposed perceptual effects of expectations as a function of signal reliability, in line with the opposing process theory. Interestingly, it can however be noted that these perceptual effects mostly seemed to affect unexpected rather than expected stimuli. Indeed, as can be seen in Fig. 2, the MBL for expected Sample objects was very close to a "perfect match" suggesting an accurate estimation of their blur irrespective of the Signal reliability. On the contrary, the blur of incongruent Sample objects tended to be underestimated when signal reliability was high - which would suggest that the object is perceived as sharper than it actually is, while it was overestimated when signal reliability was low - which would suggest that the object is perceived as blurrier than it actually is. The relatively enhanced (dampened) perception of expected than unexpected stimuli when signal reliability is low (high) put forward by the opposing process theory may therefore actually stem from these effects on unexpected stimuli.

These findings raise the question of the neural mechanisms underpinning the selective perceptual enhancement of stimuli depending on predictions' accuracy and signal reliability. The opposing process theory (Press et al., 2020) proposes that predictions about sensory inputs are signalled via descending connections between adjacent areas along the cortical hierarchy. These prediction signals would "preactivate" the neural representation of expected stimulus features in lower-level areas by enhancing the response of neurons tuned to these features while inhibiting the response of neurons tuned to unexpected ones. If sensory inputs closely match predictions, no further process is required which speeds up stimulus processing and recognition. If sensory inputs however clearly and strongly diverge from expectations, neural

representations in low-level areas should be dominated by the response of neurons tuned away from expected features. The gain of these neurons' response would subsequently increase while the gain of neurons tuned to expected features would decrease. This would favour the processing of unexpected and hence potentially informative content used to update predictions. However, if sensory inputs are noisy, their neural representation should be too. The response of neurons tuned to unexpected features may therefore not differ enough from the response of neurons tuned to expected ones to trigger the dampening mechanism. In which case, enhanced response of neurons tuned to expected features would still predominate. In this context, signal reliability is thus estimated as the variance of the distribution of neural responses to the stimulus (the larger the variance, the less reliable the signal, Ernst & Banks, 2002). Such "sensory" estimation of signal reliability may also be combined to top-down expectations based on prior experience (e.g., knowing that sensory inputs are likely to be (un)reliable, Yon & Frith, 2021). An important implication of these assumptions is therefore that the selective enhancement of responses to expected/unexpected features not only occur alternatively depending on signal reliability, but also over the time course of visual processing - the relative enhancement of responses to expected features preceding that of unexpected features. Although the design of the present study does not allow to empirically address this assumption, further studies addressing the subjective perception of expected and unexpected stimuli while varying the exposure duration of stimuli could be considered to explicitly test this hypothesis (e.g., Yon & Press, 2017).

It should however be noted that the supplementary analyses including the random effects of the stimuli in our models showed no significant interaction between the Sample congruence and the signal Reliability. This indicates that these results, which were significant in the main analyses not including the stimuli variability, cannot be extended to all visual stimuli but rather restrict to the stimuli used for the present experiment. Indeed, the stimulus variability included in the supplementary analyses hides the global fixed effects tested as these fixed effects were not large enough relative to the variability of the stimuli to reach the significant threshold. The fact that we cannot generalise our results to all stimuli is not necessarily problematic as our stimuli were not very ecological and specifically built to address our hypotheses. Therefore, it would however be interesting to replicate this study with more ecological stimuli (e.g., photographs of real scenes containing expected or unexpected objects, see e.g., Öhlschläger & Vö, 2017) to determine whether such findings can be generalised to every visual stimulus.

Investigating the perceptual effects of expectations as a function of signal reliability under more ecological conditions may also have important implications to better understand the determinants of subjective perception of expected/unexpected visual stimuli in various contexts and populations and how this may in turn influence behaviour. For example, in daily situations such as driving in which expectations and anticipation are crucial and information across the visual field becomes noisier with speed, the present findings may suggest that an unexpected stimulus in the periphery could be perceived as less clearly and possibly ignored, undetected and/or misunderstood relative to an expected stimulus. Answering these questions also has societal and clinical implications as it would for example allow to better identify the extent to which sensory deficits related to visual pathologies or ageing can be compensated for or on the contrary amplified by these predictive mechanisms when processing expected relative to unexpected stimuli. For example, in a widespread visual pathology such as cataract which can induce colour-faded or blurry vision making sensory inputs less reliable, perception of unexpected stimuli is likely to be more affected than that of expected stimuli.

6. Conclusions

In conclusion, our results suggest that expectations based on scene

contextual information differentially affect the subjective perception of expected and unexpected objects and that these effects are also modulated by the reliability of visual signals. Visual perception is dominated by what we expect when the environment is uncertain but tends to be biased toward what surprises us as sensory inputs become more reliable. The present findings therefore provide empirical support for the recently developed opposing process theory which may help to reconcile a priori contradictory findings in the literature on how expectations shape perception.

CRedit authorship contribution statement

Pauline Rossel: Conceptualization, Methodology, Formal analysis, Investigation, Writing – original draft, Visualization. **Carole Peyrin:** Conceptualization, Writing – review & editing, Supervision. **Louise Kauffmann:** Conceptualization, Methodology, Writing – original draft, Supervision, Project administration.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Stimuli, executable files of the experiment, data and analysis scripts are available in the Open Science Framework repository. The link is specified in the author note.

Acknowledgments

We warmly thank Clare Press as well as another anonymous Reviewer for their helpful and constructive comments on earlier versions of the manuscript.

Author Note.

Declarations of interest: none.

This study was not preregistered but stimuli, executable files of the experiment, data and analysis scripts are available in the Open Science Framework repository at <https://osf.io/mcv8p/>.

References

- Bates, D., Kliegl, R., Vasishth, S., & Baayen, R. H. (2015). *Parsimonious mixed models*.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Blakemore, S.-J., Wolpert, D. M., & Frith, C. D. (1998). Central cancellation of self-produced tickle sensation. *Nature Neuroscience*, 1(7), 635–640. <https://doi.org/10.1038/2870>
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, 10(4), 433–436. <https://doi.org/10.1163/156856897X00357>
- Brandman, T., & Peelen, M. V. (2017). Interaction between scene and object processing revealed by human fMRI and MEG decoding. *The Journal of Neuroscience*, 37(32), 7700–7710. <https://doi.org/10.1523/JNEUROSCI.0582-17.2017>
- Davenport, J. L. (2007). Consistency effects between objects in scenes. *Memory & Cognition*, 35(3), 393–401. <https://doi.org/10.3758/BF03193280>
- Davenport, J. L., & Potter, M. C. (2004). Scene consistency in object and background perception. *Psychological Science*, 15(8), 559–564. <https://doi.org/10.1111/j.0956-7976.2004.00719.x>
- de Lange, F. P., Heilbron, M., & Kok, P. (2018). How do expectations shape perception? *Trends in Cognitive Sciences*, 22(9), 764–779. <https://doi.org/10.1016/j.tics.2018.06.002>
- Dubova, M., & Goldstone, R. L. (2022). Categorical perception meets El Greco: Categories unequally influence color perception of simultaneously present objects. *Cognition*, 223, Article 105025. <https://doi.org/10.1016/j.cognition.2022.105025>
- Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415, 429–433. <https://doi.org/10.1038/415429a>
- Feldman, H., & Friston, K. J. (2010). Attention, uncertainty, and free-energy. *Frontiers in Human Neuroscience*, 4, 1–23. <https://doi.org/10.3389/fnhum.2010.00215>
- Firestone, C., & Scholl, B. J. (2014). "Top-Down" effects where none should be found. *Psychological Science*, 25(1), 38–46. <https://doi.org/10.1177/0956797613485092>

- Firestone, C., & Scholl, B. J. (2016). Cognition does not affect perception: Evaluating the evidence for “top-down” effects. *Behavioral and Brain Sciences*, 39, e229.
- Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1456), 815–836. <https://doi.org/10.1098/rstb.2005.1622>
- Greene, M. R., Botros, A. P., Beck, D. M., & Fei-Fei, L. (2015). What you see is what you expect: Rapid scene understanding benefits from prior experience. *Attention, Perception, & Psychophysics*, 77(4), 1239–1251. <https://doi.org/10.3758/s13414-015-0859-8>
- Han, B., & VanRullen, R. (2016). Shape perception enhances perceived contrast: Evidence for excitatory predictive feedback? *Scientific Reports*, 6(1), 22944. <https://doi.org/10.1038/srep22944>
- Hansen, T., Olkkonen, M., Walter, S., & Gegenfurtner, K. R. (2006). Memory modulates color appearance. *Nature Neuroscience*, 9(11), 1367–1368. <https://doi.org/10.1038/nn1794>
- Hollingworth, A., & Henderson, J. M. (2000). Semantic informativeness mediates the detection of changes in natural scenes. *Visual Cognition*, 7(1–3), 213–235. <https://doi.org/10.1080/135062800394775>
- Joubert, O. R., Fize, D., Rousselet, G. A., & Fabre-Thorpe, M. (2008). Early interference of context congruence on object processing in rapid visual categorization of natural scenes. *Journal of Vision*, 8(13). <https://doi.org/10.1167/8.13.11>
- Kaiser, D., Quek, G. L., Cichy, R. M., & Peelen, M. V. (2019). Object vision in a structured world. *Trends in Cognitive Sciences*, 23(8), 672–685. <https://doi.org/10.1016/j.tics.2019.04.013>
- Kok, P., Jehee, J. F. M., & de Lange, F. P. (2012). Less is more: Expectation sharpens representations in the primary visual cortex. *Neuron*, 75(2), 265–270. <https://doi.org/10.1016/j.neuron.2012.04.034>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1–26. <https://doi.org/10.18637/jss.v082.i13>
- Lai, L. Y., Frömer, R., Festa, E. K., & Heindel, W. C. (2020). Age-related changes in the neural dynamics of bottom-up and top-down processing during visual object recognition: An electrophysiological investigation. *Neurobiology of Aging*, 94, 38–49. <https://doi.org/10.1016/j.neurobiolaging.2020.05.010>
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, 4, 1–12. <https://doi.org/10.3389/fpsyg.2013.00863>
- LaPointe, M. R. P., Lupianez, J., & Milliken, B. (2013). Context congruency effects in change detection: Opposing effects on detection and identification. *Visual Cognition*, 21(1), 99–122. <https://doi.org/10.1080/13506285.2013.787133>
- Lee, T. S., & Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America A*, 20(7), 1434. <https://doi.org/10.1364/JOSAA.20.001434>
- Lupyan, G. (2017). Objective effects of knowledge on visual perception. *Journal of Experimental Psychology: Human Perception and Performance*, 43(4), 794–806. <https://doi.org/10.1037/xhp0000343>
- Mack, A., Clarke, J., Erol, M., & Bert, J. (2017). Scene incongruity and attention. *Consciousness and Cognition*, 48, 87–103. <https://doi.org/10.1016/j.concog.2016.10.010>
- Öhlschläger, S., & Vö, M. L. (2017). SCEGRAM: An image database for semantic and syntactic inconsistencies in scenes. *Behavior Research Methods*, 49, 1780–1791. <https://doi.org/10.3758/s13428-016-0820-3>
- Palmer, T. E. (1975). The effects of contextual scenes on the identification of objects. *Memory & Cognition*, 3(5), 519–526. <https://doi.org/10.3758/BF03197524>
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10(4), 437–442. <https://doi.org/10.1163/156856897X00366>
- Perez, D. C., Cook, S. M., & Peterson, M. A. (2020). Prior experience alters the appearance of blurry object borders. *Scientific Reports*, 10(1), 1–13. <https://doi.org/10.1038/s41598-020-62728-y>
- Pinto, Y., van Gaal, S., de Lange, F. P., Lamme, V. A. F., & Seth, A. K. (2015). Expectations accelerate entry of visual stimuli into awareness. *Journal of Vision*, 15(8), 13. <https://doi.org/10.1167/15.8.13>
- Press, C., Kok, P., & Yon, D. (2020). The perceptual prediction paradox. *Trends in Cognitive Sciences*, 24(1), 13–24. <https://doi.org/10.1016/j.tics.2019.11.003>
- Pylyshyn, Z. (1999). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences*, 22, 341–423. <https://doi.org/10.1017/s0140525x99002022>
- R Core Team. (2019). *R: A language and environment for statistical computing* (3.6.1). R Foundation for Statistical Computing.
- Rossel, P., Peyrin, C., Roux-Sibilon, A., & Kauffmann, L. (2022). It makes sense, so i see it better! contextual information about the visual environment increases its perceived sharpness. *Journal of Experimental Psychology: Human Perception and Performance*, 48(4), 331–350. <https://doi.org/10.1037/xhp0000993>
- Roux-Sibilon, A., Trouilloud, A., Kauffmann, L., Guyader, N., Mermillod, M., & Peyrin, C. (2019). Influence of peripheral vision on object categorization in central vision. *Journal of Vision*, 19(14), 1–16. <https://doi.org/10.1167/19.14.7>
- Spaak, E., Peelen, M. V., & de Lange, F. P. (2022). Scene context impairs perception of semantically congruent objects. *Psychological Science*, 33(2), 299–313. <https://doi.org/10.1177/09567976211032676>
- Stein, T., & Peelen, M. V. (2015). Content-specific expectations enhance stimulus detectability by increasing perceptual sensitivity. *Journal of Experimental Psychology: General*, 144(6), 1089–1104. <https://doi.org/10.1037/xge0000109>
- Teufel, C., Dakin, S. C., & Fletcher, P. C. (2018). Prior object-knowledge sharpens properties of early visual feature-detectors. *Scientific Reports*, 8, 1–12. <https://doi.org/10.1038/s41598-018-28845-5>
- The GIMP Development Team. (2019). *GIMP*.
- Wyart, V., Nobre, A. C., & Summerfield, C. (2012). Dissociable prior influences of signal probability and relevance on visual contrast sensitivity. *Proceedings of the National Academy of Sciences of the United States of America*, 109(9), 3593–3598. <https://doi.org/10.1073/pnas.1204601109>
- Yon, D., & Frith, C. D. (2021). Precision and the Bayesian brain. *Current Biology*, 31(17), R1026–R1032. <https://doi.org/10.1016/j.cub.2021.07.044>
- Yon, D., & Press, C. (2017). Predicted action consequences are perceptually facilitated before cancellation. *Journal of Experimental Psychology: Human Perception and Performance*, 43(6), 1073–1083.

Bibliographie

- Aggleton, J. P., & Vann, S. D. (2004). Testing the importance of the retrosplenial navigation system : lesion size but not strain matters : a reply to Harker and Whishaw. *Neuroscience and Biobehavioral Reviews*, 28(5), 525–531.
<https://doi.org/https://doi.org/10.1016/j.neubiorev.2004.08.003>
- Aitken, F., Turner, G., & Kok, P. (2020). Prior Expectations of Motion Direction Modulate Early Sensory Processing. *The Journal of Neuroscience*, 40(33), 6389–6397.
<https://doi.org/10.1523/JNEUROSCI.0537-20.2020>
- Alink, A., Schwiedrzik, C. M., Kohler, A., Singer, W., & Muckli, L. (2010). Stimulus predictability reduces responses in primary visual cortex. *Journal of Neuroscience*, 30(8), 2960–2966. <https://doi.org/10.1523/JNEUROSCI.3730-10.2010>
- American Psychological Association. (2017). American Psychological Association’s Ethics Code. *Ethical Principles of Psychologists and Code of Conduct*.
<https://doi.org/10.1525/california/9780520234277.003.0001>
- Angelucci, A., Levitt, J. B., & Lund, J. S. (2002). Anatomical origins of the classical receptive field and modulatory surround field of single neurons in macaque visual cortical area V1. *Progress in Brain Research*, 136, 373–388.
[https://doi.org/https://doi.org/10.1016/S0079-6123\(02\)36031-X](https://doi.org/https://doi.org/10.1016/S0079-6123(02)36031-X)
- Bang, J. W., & Rahnev, D. (2017). Stimulus expectation alters decision criterion but not sensory signal in perceptual decision making. *Scientific Reports*, 7(1), 17072.
<https://doi.org/10.1038/s41598-017-16885-2>
- Bar, M. (2003). A cortical mechanism for triggering top-down facilitation in visual object recognition. *Journal of Cognitive Neuroscience*, 15(4), 600–609.
<https://doi.org/10.1162/089892903321662976>
- Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience*, 5(8), 617–629.
<https://doi.org/10.1038/nrn1476>
- Bar, M. (2007). The proactive brain: using analogies and associations to generate predictions.

Bibliographie

- Trends in Cognitive Sciences*, 11(7), 280–289. <https://doi.org/10.1016/j.tics.2007.05.005>
- Bar, M. (2009). The proactive brain: Memory for predictions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1521), 1235–1243.
<https://doi.org/10.1098/rstb.2008.0310>
- Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., Hamalainen, M. S., Marinkovic, K., Schacter, D. L., Rosen, B. R., & Halgren, E. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences*, 103(2), 449–454. <https://doi.org/10.1073/pnas.0507062103>
- Bates, D., Kliegl, R., Vasishth, S., & Baayen, R. H. (2015). *Parsimonious Mixed Models*.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48.
<https://doi.org/10.18637/jss.v067.i01>
- Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive Psychology*, 14(2), 143–177.
[https://doi.org/10.1016/0010-0285\(82\)90007-X](https://doi.org/10.1016/0010-0285(82)90007-X)
- Blank, H., & Davis, M. H. (2016). Prediction Errors but Not Sharpened Signals Simulate Multivoxel fMRI Patterns during Speech Perception. *PLoS Biology*, 14(11), 1–32.
<https://doi.org/10.1371/journal.pbio.1002577>
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10(4), 433–436.
<https://doi.org/10.1163/156856897X00357>
- Brandman, T., & Peelen, M. V. (2017). Interaction between scene and object processing revealed by human fMRI and MEG decoding. *The Journal of Neuroscience*, 37(32), 7700–7710. <https://doi.org/10.1523/JNEUROSCI.0582-17.2017>
- Brandman, T., & Peelen, M. V. (2019). Signposts in the Fog: Objects Facilitate Scene Representations in Left Scene-selective Cortex. *Journal of Cognitive Neuroscience*, 31(3), 390–400. https://doi.org/10.1162/jocn_a_01258
- Brandman, T., & Peelen, M. V. (2023). Objects sharpen visual scene representations : evidence from MEG decoding. *BioRxiv*, 1–24.

Bibliographie

<https://doi.org/https://doi.org/10.1101/2023.04.06.535903>

- Bullier, J. (2001). Integrated model of visual processing. *Brain Research Reviews*, 36(2–3), 96–107. [https://doi.org/https://doi.org/10.1016/S0165-0173\(01\)00085-6](https://doi.org/https://doi.org/10.1016/S0165-0173(01)00085-6)
- Buonocore, A., Dimigen, O., & Melcher, D. (2020). Post-saccadic face processing is modulated by pre-saccadic preview: Evidence from fixation-related potentials. *Journal of Neuroscience*, 40(11), 2305–2313. <https://doi.org/10.1523/JNEUROSCI.0861-19.2020>
- Carrasco, M., Ling, S., & Read, S. (2004). Attention alters appearance. *Nature Neuroscience*, 7(3), 308–313. <https://doi.org/10.1038/nn1194>
- Chalk, M., Seitz, A. R., & Seriès, P. (2010). Rapidly learned stimulus expectations alter perception of motion. *Journal of Vision*, 10(8), 1–18. <https://doi.org/10.1167/10.8.2>
- Cheadle, S., Egnér, T., Wyart, V., Wu, C., & Summerfield, C. (2015). Feature expectation heightens visual sensitivity during fine orientation discrimination. *Journal of Vision*, 15(14), 14. <https://doi.org/10.1167/15.14.14>
- Christensen, J. H., Bex, P. J., & Fiser, J. (2015). Prior implicit knowledge shapes human threshold for orientation noise. *Journal of Vision*, 15(9), 24. <https://doi.org/10.1167/15.9.24>
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(3), 181–204. <https://doi.org/10.1017/S0140525X12000477>
- Coco, M. I., Nuthmann, A., & Dimigen, O. (2020). Fixation-related brain potentials during semantic integration of object–scene information. *Journal of Cognitive Neuroscience*, 32(4), 571–589. https://doi.org/10.1162/jocn_a_01504
- Collins, T. (2019). The perceptual continuity field is retinotopic. *Scientific Reports*, 9, 18841. <https://doi.org/10.1038/s41598-019-55134-6>
- Collins, T. (2020). Serial dependence alters perceived object appearance. *Journal of Vision*, 20(13), 1–8. <https://doi.org/10.1167/JOV.20.13.9>

Bibliographie

- Collins, T. (2022). Serial dependence tracks objects and scenes in parallel and independently. *Journal of Vision*, 22(7), 1–7. <https://doi.org/10.1167/jov.22.7.4>
- Curcio, C. A., Sloan, K. R., Kalina, R. E., & Hendrickson, A. E. (1990). What is Perception ? In *Journal of comparative neurology* (Vol. 292, Issue 4, pp. 497–523). <https://doi.org/https://doi.org/10.1002/cne.902920402>
- Daniel, P. M., & Whitteridge, D. (1961). The representation of the visual field on the cerebral cortex in monkeys. *The Journal of Physiology*, 159(2), 203–221. <https://doi.org/10.1113/jphysiol.1961.sp006803>
- Davenport, J. L. (2007). Consistency effects between objects in scenes. *Memory & Cognition*, 35(3), 393–401. <https://doi.org/10.3758/BF03193280>
- Davenport, J. L., & Potter, M. C. (2004). Scene Consistency in Object and Background Perception. *Psychological Science*, 15(8), 559–564. <https://doi.org/10.1111/j.0956-7976.2004.00719.x>
- de Haan, E. H. F., & Cowey, A. (2011). On the usefulness of “what” and “where” pathways in vision. *Trends in Cognitive Sciences*, 15(10), 460–466. <https://doi.org/10.1016/j.tics.2011.08.005>
- de Lange, F. P., Heilbron, M., & Kok, P. (2018). How do expectations shape perception? *Trends in Cognitive Sciences*, 22(9), 764–779. <https://doi.org/10.1016/j.tics.2018.06.002>
- De Valois, R. L., Albrecht, D. G., & Thorell, L. G. (1982). Spatial frequency selectivity of cells in macaque visual cortex. *Vision Research*, 22(5), 545–559. [https://doi.org/10.1016/0042-6989\(82\)90113-4](https://doi.org/10.1016/0042-6989(82)90113-4)
- Denison, R. N., Sheynin, J., & Silver, M. A. (2016). Perceptual suppression of predicted natural images. *Journal of Vision*, 16(13), 1–15. <https://doi.org/10.1167/16.13.6>
- Dijkstra, N., Kok, P., & Fleming, S. M. (2022). Imagery adds stimulus-specific sensory evidence to perceptual detection. *Journal of Vision*, 22(2), 1–8. <https://doi.org/10.1167/jov.22.2.11>
- Dilks, D. D., Julian, J. B., Paunov, A. M., & Kanwisher, N. (2013). The Occipital Place Area Is Causally and Selectively Involved in Scene Perception. *Journal of Neuroscience*,

Bibliographie

- 33(4), 1331–1336. <https://doi.org/10.1523/JNEUROSCI.4081-12.2013>
- Dubova, M., & Goldstone, R. L. (2022). Categorical perception meets El Greco: Categories unequally influence color perception of simultaneously present objects. *Cognition*, 223, 105025. <https://doi.org/10.1016/j.cognition.2022.105025>
- Eaton, J. W., Bateman, D., Hauberg, S., & Wehbring, R. (2019). *GNU Octave version 5.1.0 manual: a high-level interactive language for numerical computations*.
- Edwards, G., Vanrullen, R., & Cavanagh, P. (2018). Decoding trans-saccadic memory. *Journal of Neuroscience*, 38(5), 1114–1123. <https://doi.org/10.1523/JNEUROSCI.0854-17.2017>
- Eger, E., Henson, R. N., Driver, J., & Dolan, R. J. (2007). Mechanisms of top-down facilitation in perception of visual objects studied by fMRI. *Cerebral Cortex*, 17(9), 2123–2133. <https://doi.org/10.1093/cercor/bhl119>
- Enroth-cugell, C., & Robson, J. G. (1966). The contrast sensitivity of retinal ganglion cells of the cat. *Journal of Physiology*, 187(3), 517–552. <https://doi.org/https://doi.org/10.1113/jphysiol.1966.sp008107>
- Enroth-cugell, C., & Robson, J. G. (1984). Functional Characteristics and Diversity of Cat Retinal Ganglion Cells Basic Characteristics and Quantitative Description. *Investigative Ophthalmology & Visual Science*, 25(3), 250–267.
- Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, 392, 598–601. <https://doi.org/https://doi.org/10.1038/33402>
- Ergenzinger, E. R., Glasier, M. M., Hahn, J. O., & Pons, T. P. (1998). Cortically induced thalamic plasticity in the primate somatosensory system. *Nature Neuroscience*, 1, 226–229. <https://doi.org/https://doi.org/10.1038/673>
- Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415, 429–433. <https://doi.org/https://doi.org/10.1038/415429a>
- Esterman, M., & Yantis, S. (2010). Category expectation modulates object-selective cortical activity. *Journal of Vision*, 8(6), 555–555. <https://doi.org/10.1167/8.6.555>

Bibliographie

- Feldman, H., & Friston, K. J. (2010). Attention, uncertainty, and free-energy. *Frontiers in Human Neuroscience*, 4, 1–23. <https://doi.org/10.3389/fnhum.2010.00215>
- Felleman, D. J., & Van Essen, D. C. (1991). Distributed Hierarchical Processing in the Primate Cerebral Cortex. *Cerebral Cortex*, 1(1), 1–47. <https://doi.org/10.1093/cercor/1.1.1-a>
- Firestone, C., & Scholl, B. J. (2014). “Top-Down” Effects Where None Should Be Found. *Psychological Science*, 25(1), 38–46. <https://doi.org/10.1177/0956797613485092>
- Firestone, C., & Scholl, B. J. (2015). When do ratings implicate perception versus judgment? The “overgeneralization test” for top-down effects. *Visual Cognition*, 23(9–10), 1217–1226. <https://doi.org/10.1080/13506285.2016.1160171>
- Firestone, C., & Scholl, B. J. (2016). Cognition does not affect perception: Evaluating the evidence for “top-down” effects. *Behavioral and Brain Sciences*, 39, e229. <https://doi.org/10.1017/S0140525X15000965>
- Fischer, J., & Whitney, D. (2014). Serial dependence in visual perception. *Nature Neuroscience*, 17(5), 738–743. <https://doi.org/10.1038/nn.3689>
- Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1456), 815–836. <https://doi.org/10.1098/rstb.2005.1622>
- Gallant, J. L., Connor, C. E., Rakshit, S., Lewis, J. W., & Van Essen, D. C. (1996). Neural responses to polar, hyperbolic, and Cartesian gratings in area V4 of the macaque monkey. *Journal of Neurophysiology*, 76(4), 2718–2739. <https://doi.org/10.1152/jn.1996.76.4.2718>
- Gandolfo, M., & Downing, P. E. (2019). Causal evidence for expression of perceptual expectations in category-selective extrastriate regions. *Current Biology*, 29(15), 2496–2500.e3. <https://doi.org/10.1016/j.cub.2019.06.024>
- Gilbert, J. R., & Moran, R. J. (2016). Inputs to prefrontal cortex support visual recognition in the aging brain. *Scientific Reports*, 6, 1–9. <https://doi.org/10.1038/srep31943>
- Girshick, A. R., Landy, M. S., & Simoncelli, E. P. (2011). Cardinal rules: Visual orientation

Bibliographie

- perception reflects knowledge of environmental statistics. *Nature Neuroscience*, *14*, 926–932. <https://doi.org/10.1038/nn.2831>
- Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, *15*(1), 20–25. [https://doi.org/https://doi.org/10.1016/0166-2236\(92\)90344-8](https://doi.org/https://doi.org/10.1016/0166-2236(92)90344-8)
- Greene, M. R., Botros, A. P., Beck, D. M., & Fei-Fei, L. (2015). What you see is what you expect: rapid scene understanding benefits from prior experience. *Attention, Perception, & Psychophysics*, *77*(4), 1239–1251. <https://doi.org/10.3758/s13414-015-0859-8>
- Grill-Spector, K. (2003). The neural basis of object perception. *Current Opinion in Neurobiology*, *13*(2), 159–166. [https://doi.org/10.1016/S0959-4388\(03\)00040-0](https://doi.org/10.1016/S0959-4388(03)00040-0)
- Han, B., & VanRullen, R. (2016). Shape perception enhances perceived contrast: evidence for excitatory predictive feedback? *Scientific Reports*, *6*(1), 22944. <https://doi.org/10.1038/srep22944>
- Hansen, T., Olkkonen, M., Walter, S., & Gegenfurtner, K. R. (2006). Memory modulates color appearance. *Nature Neuroscience*, *9*(11), 1367–1368. <https://doi.org/10.1038/nn1794>
- Heilbron, M., Richter, D., Ekman, M., Hagoort, P., & de Lange, F. P. (2020). Word contexts enhance the neural representation of individual letters in early visual cortex. *Nature Communications*, *11*(1), 321. <https://doi.org/10.1038/s41467-019-13996-4>
- Henderson, J. M. (2017). Gaze Control as Prediction. *Trends in Cognitive Sciences*, *21*(1), 15–23. <https://doi.org/10.1016/j.tics.2016.11.003>
- Henderson, J. M., & Hollingworth, A. (1999). High-level scene perception. *Annual Review of Psychology*, *50*, 243–271. <https://doi.org/10.1146/annurev.psych.50.1.243>
- Henninger, F., Shevchenko, Y., Mertens, U. K., Kieslich, P. J., & Hilbig, B. E. (2020). *lab.js: A free, open, online study builder*. <https://doi.org/https://doi.org/10.31234/osf.io/fqr49>
- Hollingworth, A., & Henderson, J. M. (2000). Semantic Informativeness Mediates the Detection of Changes in Natural Scenes. *Visual Cognition*, *7*(1–3), 213–235. <https://doi.org/10.1080/135062800394775>

Bibliographie

- Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, *160*(1), 106–154. <https://doi.org/10.1113/jphysiol.1962.sp006837>
- Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology*, *195*(1), 215–243. <https://doi.org/10.1113/jphysiol.1968.sp008455>
- Hubel, D. H., & Wiesel, T. N. (1977). Ferrier lecture - Functional architecture of macaque monkey visual cortex. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, *198*(1130), 1–59. <https://doi.org/10.1098/rspb.1977.0085>
- Jiang, L. P., & Rao, R. P. N. (2022). Dynamic Predictive Coding : A New Model of Hierarchical Sequence Learning and Prediction in the Cortex. *BioRxiv*. <https://doi.org/https://doi.org/10.1101/2022.06.23.497415>
- Johnston, W. A., Hawley, K. J., Plewe, S. H., Elliott, J. M. G., & DeWitt, M. J. (1990). Attention capture by novel stimuli. *Journal of Experimental Psychology: General*, *119*(4), 397–411. <https://doi.org/10.1037/0096-3445.119.4.397>
- Joubert, O. R., Fize, D., Rousselet, G. A., & Fabre-Thorpe, M. (2008). Early interference of context congruence on object processing in rapid visual categorization of natural scenes. *Journal of Vision*, *8*(13), 11–11. <https://doi.org/10.1167/8.13.11>
- Joubert, O. R., Rousselet, G. a, Fize, D., & Fabre-Thorpe, M. (2007). Processing scene context: Fast categorization and object interference. *Vision Research*, *47*(26), 3286–3297. <https://doi.org/10.1016/j.visres.2007.09.013>
- Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology*, *103*(1), 54–69. <https://doi.org/10.1037/a0028347>
- Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The Fusiform Face Area : A Module in Human Extrastriate Cortex Specialized for Face Perception. *Journal of Neuroscience*, *17*(11), 4302–4311. <https://doi.org/https://doi.org/10.1523/JNEUROSCI.17-11-04302>

Bibliographie

- Kaplan, E. (2004). The M, P and K pathways of the Primate Visual System. *The Visual Neurosciences, 1*, 481–493. <https://doi.org/10.7551/mitpress/7131.003.0036>
- Kaplan, E., & Shapley, R. M. (1986). The primate retina contains two types of ganglion cells, with high and low contrast sensitivity. *Proceedings of the National Academy of Sciences, 83*(8), 2755–2757. <https://doi.org/https://doi.org/10.1073/pnas.83.8.2755>
- Kauffmann, L., Bourgin, J., Guyader, N., & Peyrin, C. (2015). The Neural Bases of the Semantic Interference of Spatial Frequency-based Information in Scenes. *Journal of Cognitive Neuroscience, 27*(12), 2394–2405. https://doi.org/10.1162/jocn_a_00861
- Kauffmann, L., Chauvin, A., Pichat, C., & Peyrin, C. (2015). Effective connectivity in the neural network underlying coarse-to-fine categorization of visual scenes. A dynamic causal modeling study. *Brain and Cognition, 99*, 46–56. <https://doi.org/10.1016/j.bandc.2015.07.004>
- Kauffmann, L., Ramanoël, S., Peyrin, C., & Greene, M. R. (2014). The neural bases of spatial frequency processing during scene perception. *Frontiers in Integrative Neuroscience, 8*, 37. <https://doi.org/10.3389/fnint.2014.00037>
- Kok, P., & de Lange, F. P. (2014). Shape perception simultaneously up- and downregulates neural activity in the primary visual cortex. *Current Biology, 24*(13), 1531–1535. <https://doi.org/10.1016/j.cub.2014.05.042>
- Kok, P., & de Lange, F. P. (2015). Predictive coding in sensory cortex. In B. U. Forstmann & E. J. Wagenmakers (Eds.), *An Introduction to Model-Based Cognitive Neuroscience* (pp. 221–244). Springer New York. https://doi.org/10.1007/978-1-4939-2236-9_11
- Kok, P., Jehee, J. F. M., & de Lange, F. P. (2012). Less is more: Expectation sharpens representations in the primary visual cortex. *Neuron, 75*(2), 265–270. <https://doi.org/10.1016/j.neuron.2012.04.034>
- Kok, P., Rahnev, D., Jehee, J. F. M., Lau, H. C., & De Lange, F. P. (2012). Attention reverses the effect of prediction in silencing sensory signals. *Cerebral Cortex, 22*(9), 2197–2206. <https://doi.org/10.1093/cercor/bhr310>
- Kumar, S., Kaposvari, P., & Vogels, R. (2017). Encoding of predictable and unpredictable

Bibliographie

- stimuli by inferior temporal cortical neurons. *Journal of Cognitive Neuroscience*, 29(8), 1445–1454. https://doi.org/10.1162/jocn_a_01135
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1–26. <https://doi.org/10.18637/jss.v082.i13>
- Lai, L. Y., Frömer, R., Festa, E. K., & Heindel, W. C. (2020). Age-related changes in the neural dynamics of bottom-up and top-down processing during visual object recognition: an electrophysiological investigation. *Neurobiology of Aging*, 94, 38–49. <https://doi.org/10.1016/j.neurobiolaging.2020.05.010>
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, 4, 1–12. <https://doi.org/10.3389/fpsyg.2013.00863>
- Lamme, V. A., Super, H., & Spekreijse, H. (1998). Feedforward, horizontal, and feedback processing in the visual cortex. *Current Opinion in Neurobiology*, 8(4), 529–535. [https://doi.org/https://doi.org/10.1016/S0959-4388\(98\)80042-1](https://doi.org/https://doi.org/10.1016/S0959-4388(98)80042-1)
- LaPointe, M. R. P., Lupianez, J., & Milliken, B. (2013). Context congruency effects in change detection: Opposing effects on detection and identification. *Visual Cognition*, 21(1), 99–122. <https://doi.org/10.1080/13506285.2013.787133>
- Lee, T. S., & Nguyen, M. (2001). Dynamics of subjective contour formation in the early visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 98(4), 1907–1911. <https://doi.org/10.1073/pnas.98.4.1907>
- Lupyan, G. (2015). Object knowledge changes visual appearance: Semantic effects on color afterimages. *Acta Psychologica*, 161, 117–130. <https://doi.org/10.1016/j.actpsy.2015.08.006>
- Lupyan, G. (2017). Objective effects of knowledge on visual perception. *Journal of Experimental Psychology: Human Perception and Performance*, 43(4), 794–806. <https://doi.org/10.1037/xhp0000343>
- Mack, A., Clarke, J., Erol, M., & Bert, J. (2017). Scene incongruity and attention.

Bibliographie

- Consciousness and Cognition*, 48, 87–103. <https://doi.org/10.1016/j.concog.2016.10.010>
- Maiello, G., Walker, L., Bex, P. J., & Vera-diaz, F. A. (2017). Blur perception throughout the visual field in myopia and emmetropia. *Journal of Vision*, 17(5), 1–13.
<https://doi.org/10.1167/17.5.3>
- Malach, R., Reppas, J. B., Benson, R. R., Kwong, K. K., Jlang, H., Kennedy, W. A., Ledden, P. J., Brady, T. J., Rosen, B. R., & Tootell, R. B. H. (1995). Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proceedings of the National Academy of Sciences*, 92(18), 8135–8139.
<https://doi.org/10.1073/pnas.92.18.8135>
- Marr, D., & Vaina, L. (1982). Representation and recognition of the movements of shapes. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 214(1197), 501–524. <https://doi.org/https://doi.org/10.1098/rspb.1982.0024>
- Mather, G., & Smith, D. R. R. (2002). Blur discrimination and its relation to blur-mediated depth perception. *Perceptio*, 31(10), 1211–1220. <https://doi.org/10.1068/p3254>
- McGuinness, D. (1976). Away from a Unisex Psychology: Individual Differences in Visual Sensory and Perceptual Processes. *Perception*, 5(3), 279–294.
<https://doi.org/10.1068/p050279>
- Merigan, W. H., & Maunsell, J. H. R. (1993). How parallel are the primate visual pathways? *Annual Reviews of Neuroscience*, 16, 369–402.
<https://doi.org/https://doi.org/10.1146/annurev.ne.16.030193.002101>
- Millidge, B., Seth, A., & Buckley, C. L. (2021). Predictive Coding: a Theoretical and Experimental Review. *ArXiv Preprint*, 12979.
- Moran, R. J., Symmonds, M., Dolan, R. J., & Friston, K. J. (2014). The brain ages optimally to model its environment: Evidence from sensory learning over the adult lifespan. *PLoS Computational Biology*, 10(1), e1003422. <https://doi.org/10.1371/journal.pcbi.1003422>
- Mumford, D. (1992). On the computational architecture of the neocortex - II The role of cortico-cortical loops. *Biological Cybernetics*, 66, 241–251.
<https://doi.org/10.1007/BF00198477>

Bibliographie

- Murray, S. O., Kersten, D., Olshausen, B. A., Schrater, P., & Woods, D. L. (2002). Shape perception reduces activity in human primary visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(23), 15164–15169.
<https://doi.org/10.1073/pnas.192579399>
- O’Callaghan, C., Kveraga, K., Shine, J. M., Adams, R. B., & Bar, M. (2017a). Predictions penetrate perception : Converging insights from brain , behaviour and disorder. *Consciousness and Cognition*, *47*, 63–74. <https://doi.org/10.1016/j.concog.2016.05.003>
- O’Callaghan, C., Kveraga, K., Shine, J. M., Adams, R. B., & Bar, M. (2017b). Predictions penetrate perception: Converging insights from brain, behaviour and disorder. *Consciousness and Cognition*, *47*(5), 63–74.
<https://doi.org/10.1016/j.concog.2016.05.003>
- Öhlschläger, S., & Vö, M. L. (2017). SCEGRAM : An image database for semantic and syntactic inconsistencies in scenes. *Behavior Research Methods*, *49*, 1780–1791.
<https://doi.org/10.3758/s13428-016-0820-3>
- Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: a holistic representation of the spatial envelope. *International Journal of Computer Vision*, *42*, 145–175.
<https://doi.org/https://doi.org/10.1023/A:1011139631724>
- Oliva, A., & Torralba, A. (2007). The role of context in object recognition. *Trends in Cognitive Sciences*, *11*(12), 520–527. <https://doi.org/10.1016/j.tics.2007.09.009>
- Olkkonen, M., Hansen, T., & Gegenfurtner, K. R. (2008). Color appearance of familiar objects: Effects of object shape, texture, and illumination changes. *Journal of Vision*, *8*(5), 13. <https://doi.org/10.1167/8.5.13>
- Owsley, C. (2011). Aging and vision. *Vision Research*, *51*(13), 1610–1622.
<https://doi.org/10.1016/j.visres.2010.10.020>
- Palmer, T. E. (1975). The effects of contextual scenes on the identification of objects. *Memory & Cognition*, *3*(5), 519–526. <https://doi.org/10.3758/BF03197524>
- Panichello, M. F., Cheung, O. S., & Bar, M. (2013). Predictive feedback and conscious visual experience. *Frontiers in Psychology*, *3*, 1–8. <https://doi.org/10.3389/fpsyg.2012.00620>

Bibliographie

- Park, J., Shimojo, E., & Shimojo, S. (2010). Roles of familiarity and novelty in visual preference judgments are segregated across object categories. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(33), 14552–14555. <https://doi.org/10.1073/pnas.1004374107/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1004374107>
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442. <https://doi.org/10.1163/156856897X00366>
- Pellicano, E., & Burr, D. (2012). When the world becomes too real : a Bayesian explanation of autistic perception. *Trends in Cognitive Sciences*, *16*(10), 504–510. <https://doi.org/10.1016/j.tics.2012.08.009>
- Perez, D. C., Cook, S. M., & Peterson, M. A. (2020). Prior experience alters the appearance of blurry object borders. *Scientific Reports*, *10*(1), 1–13. <https://doi.org/10.1038/s41598-020-62728-y>
- Peyrin, C., Michel, C. M., Schwartz, S., Thut, G., Seghier, M., Landis, T., Marendaz, C., & 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. <https://doi.org/10.1162/jocn.2010.21424>
- Peyrin, C., Roux-sibilon, A., Trouilloud, A., Khazaz, S., Joly, M., Pichat, C., Boucart, M., Krainik, A., & Kauffmann, L. (2021). *Semantic and Physical Properties of Peripheral Vision Are Used for Scene Categorization in Central Vision*. *1994*, 1–15.
- Pinto, Y., van Gaal, S., de Lange, F. P., Lamme, V. A. F., & Seth, A. K. (2015). Expectations accelerate entry of visual stimuli into awareness. *Journal of Vision*, *15*(8), 13. <https://doi.org/10.1167/15.8.13>
- Powers, A. R., Mathys, C., & Corlett, P. R. (2017). Pavlovian conditioning – induced hallucinations result from overweighting of perceptual priors. *Science*, *357*(6351), 596–600. <https://doi.org/10.1126/science.aan3458>
- Press, C., Gherri, E., Heyes, C., & Eimer, M. (2010). Action preparation helps and hinders perception of action. *Journal of Cognitive Neuroscience*, *22*(10), 2198–2211.

Bibliographie

<https://doi.org/https://doi.org/10.1162/jocn.2009.21409>

Press, C., Kok, P., & Yon, D. (2020). The perceptual prediction paradox. *Trends in Cognitive Sciences*, 24(1), 13–24. <https://doi.org/10.1016/j.tics.2019.11.003>

Pylyshyn, Z. (1999). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences*, 22, 341–423. <https://doi.org/10.1017/s0140525x99002022>

R Core Team. (2019). *R: A language and environment for statistical computing* (3.6.1). R Foundation for Statistical Computing.

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, 79–87. <https://doi.org/10.1038/4580>

Rémy, F., Saint-Aubert, L., Bacon-Macé, N., Vayssière, N., Barbeau, E., & Fabre-Thorpe, M. (2013). Object recognition in congruent and incongruent natural scenes: A life-span study. *Vision Research*, 91, 36–44. <https://doi.org/10.1016/j.visres.2013.07.006>

Richter, D., Ekman, M., & de Lange, F. P. (2018). Suppressed sensory response to predictable object stimuli throughout the ventral visual stream. *Journal of Neuroscience*, 38(34), 7452–7461. <https://doi.org/10.1523/JNEUROSCI.3421-17.2018>

Rossel, P., Peyrin, C., & Kauffmann, L. (2023). Subjective perception of objects depends on the interaction between the validity of context-based expectations and signal reliability. *Vision Research*, 206, 108191. <https://doi.org/10.1016/j.visres.2023.108191>

Rossel, P., Peyrin, C., Roux-Sibilon, A., & Kauffmann, L. (2022). It Makes Sense, So I See It Better! Contextual Information About the Visual Environment Increases Its Perceived Sharpness. *Journal of Experimental Psychology: Human Perception and Performance*, 48(4), 331–350. <https://doi.org/10.1037/xhp0000993>

Roux-sibilon, A. (2020). *Reconnaissance visuelle : Étude de l'influence de la vision périphérique dans le cadre d'un modèle fréquentiel prédictif*. Université Grenoble Alpes.

Roux-Sibilon, A., Trouilloud, A., Kauffmann, L., Guyader, N., Mermillod, M., & Peyrin, C. (2019). Influence of peripheral vision on object categorization in central vision. *Journal*

Bibliographie

- of Vision*, 19(14), 1–16. <https://doi.org/10.1167/19.14.7>
- Rungratsameetaweemana, N., Itthipuripat, S., Salazar, A., & Serences, J. T. (2018). Expectations Do Not Alter Early Sensory Processing during Perceptual Decision-Making. *The Journal of Neuroscience*, 38(24), 5632–5648. <https://doi.org/10.1523/JNEUROSCI.3638-17.2018>
- Rungratsameetaweemana, N., & Serences, J. T. (2019). Dissociating the impact of attention and expectation on early sensory processing. *Current Opinion in Psychology*, 29, 181–186. <https://doi.org/10.1016/j.copsyc.2019.03.014>
- Sabatinelli, D., Flaisch, C. A. T., Bradley, M. M., Fitzsimmons, J. R., & Lang, P. J. (2004). Affective picture perception: gender differences in visual cortex? *NeuroReport*, 15(7), 1109–1112. <https://doi.org/10.1097/01.wnr.0000126558.41468.3d>
- Serences, J. T., Saproo, S., Scolari, M., Ho, T., & Muftuler, L. T. (2009). Estimating the influence of attention on population codes in human visual cortex using voxel-based tuning functions. *NeuroImage*, 44(1), 223–231. <https://doi.org/10.1016/j.neuroimage.2008.07.043>
- Seriès, P., & Seitz, A. R. (2013). Learning what to expect (in visual perception). *Frontiers in Human Neuroscience*, 7, 1–14. <https://doi.org/10.3389/fnhum.2013.00668>
- Sinha, P., Kjelgaard, M. M., Gandhi, T. K., Tsourides, K., & Cardinaux, A. L. (2014). Autism as a disorder of prediction. *Proceedings of the National Academy of Sciences of the United States of America*, 111(42), 15220–15225. <https://doi.org/10.1073/pnas.1416797111>
- Spaak, E., Peelen, M. V., & de Lange, F. P. (2022). Scene Context Impairs Perception of Semantically Congruent Objects. *Psychological Science*, 33(2), 299–313. <https://doi.org/10.1177/09567976211032676>
- Stein, T., & Peelen, M. V. (2015). Content-specific expectations enhance stimulus detectability by increasing perceptual sensitivity. *Journal of Experimental Psychology: General*, 144(6), 1089–1104. <https://doi.org/10.1037/xge0000109>
- Sterzer, P., Adams, R. A., Fletcher, P., Frith, C., Lawrie, S. M., Muckli, L., Petrovic, P.,

Bibliographie

- Uhlhaas, P., Voss, M., & Corlett, P. R. (2018). The Predictive Coding Account of Psychosis. *Biological Psychiatry*, *84*(9), 634–643.
<https://doi.org/10.1016/j.biopsych.2018.05.015>
- Summerfield, C., & de Lange, F. P. (2014). Expectation in perceptual decision making: neural and computational mechanisms. *Nature Reviews Neuroscience*, *15*(11), 745–756.
<https://doi.org/10.1038/nrn3838>
- Summerfield, C., & Eger, T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences*, *13*(9), 403–409. <https://doi.org/10.1016/j.tics.2009.06.003>
- Teufel, C., Dakin, S. C., & Fletcher, P. C. (2018). Prior object-knowledge sharpens properties of early visual feature-detectors. *Scientific Reports*, *8*, 1–12.
<https://doi.org/10.1038/s41598-018-28845-5>
- Teufel, C., & Nanay, B. (2017). How to (and how not to) think about top-down influences on visual perception. *Consciousness and Cognition*, *47*, 17–25.
<https://doi.org/10.1016/j.concog.2016.05.008>
- Teufel, C., Subramaniam, N., Dobler, V., Perez, J., Finnemann, J., Mehta, P. R., Goodyer, I. M., & Fletcher, P. C. (2015). Shift toward prior knowledge confers a perceptual advantage in early psychosis and psychosis-prone healthy individuals. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(43), 13401–13406.
<https://doi.org/10.1073/pnas.1503916112>
- The GIMP Development Team. (2019). *GIMP*.
- Torralba, A., & Oliva, A. (1999). Semantic organization of scenes using discriminant structural templates. *Proceedings of the IEEE International Conference on Computer Vision*, *2*, 1253–1258. <https://doi.org/10.1109/iccv.1999.790424>
- Torralba, A., & Oliva, A. (2003). Statistics of natural image categories. *Network: Computation in Neural Systems*, *14*, 391–412. https://doi.org/10.1088/0954-898X_14_3_302
- Trapp, S., & Bar, M. (2015). Prediction, context, and competition in visual recognition. *Annals of the New York Academy of Sciences*, *1339*(1), 190–198.

Bibliographie

<https://doi.org/10.1111/nyas.12680>

- Trouilloud, A., Ferry, E., Boucart, M., Kauffmann, L., Warniez, A., Rouland, J. F., & Peyrin, C. (2023). Impact of glaucoma on the spatial frequency processing of scenes in central vision. *Visual Neuroscience*, *40*, E001. <https://doi.org/10.1017/S0952523822000086>
- Trouilloud, A., Kauffmann, L., Roux-sibilon, A., & Rossel, P. (2020). Rapid scene categorization : From coarse peripheral vision to fine central vision. *Vision Research*, *170*, 60–72. <https://doi.org/10.1016/j.visres.2020.02.008>
- Van de Cruys, S., Evers, K., Van der Hallen, R., Van Eylen, L., Boets, B., De-Wit, L., & Wagemans, J. (2014). Precise minds in uncertain worlds : Predictive coding in autism. *Psychological Review*, *121*(4), 649–675. <https://doi.org/10.1037/a0037665>
- Van Essen, D. C., & Deyoe, E. A. (1995). Concurrent processing in the primate visual cortex. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 383–400). The MIT Press.
- Van Essen, D. C., & Maunsell, J. H. R. (1983). Hierarchical organization and functional streams in the visual cortex. *Trends in Neurosciences*, *6*, 370–375. [https://doi.org/https://doi.org/10.1016/0166-2236\(83\)90167-4](https://doi.org/https://doi.org/10.1016/0166-2236(83)90167-4)
- von Helmholtz, H. (1867). *Handbuch der physiologischen Optik*. Voss.
- Walsh, K. S., McGovern, D. P., Clark, A., & O’Connell, R. G. (2020). Evaluating the neurophysiological evidence for predictive processing as a model of perception. *Annals of the New York Academy of Sciences*, *1464*(1), 242–268. <https://doi.org/10.1111/nyas.14321>
- Watson, A. B., & Ahumada, A. J. (2011). Blur clarified : A review and synthesis of blur discrimination. *Journal of Vision*, *11*(5), 1–23. <https://doi.org/10.1167/11.5.10.Introduction>
- Wertheim, T. H. (1894). Über die Indirekte Sehschärfe. *Psychology Physiology Der Sinnesorgane*, *7*, 172–197.
- Williams, J. R., Markov, Y. A., Tiurina, N. A., & Störmer, V. S. (2022). What You See Is What You Hear : Sounds Alter the Contents of Visual Perception. *Psychological Science*, *33*(12), 2109–2122. <https://doi.org/10.1177/09567976221121348>

Bibliographie

- Witzel, C., Valkova, H., Hansen, T., & Gegenfurtner, K. R. (2011). Object Knowledge Modulates Colour Appearance. *I-Perception*, 2(1), 13–49. <https://doi.org/10.1068/i0396>
- Wyart, V., Nobre, A. C., & Summerfield, C. (2012). Dissociable prior influences of signal probability and relevance on visual contrast sensitivity. *Proceedings of the National Academy of Sciences of the United States of America*, 109(9), 3593–3598. <https://doi.org/10.1073/pnas.1204601109>
- Yan, C., de Lange, F. P., & Richter, D. (2023). Conceptual Associations Generate Sensory Predictions. *The Journal of Neuroscience*, 43(20), 3733–3742. <https://doi.org/10.1523/JNEUROSCI.1874-22.2023>
- Yeshurun, Y., & Carrasco, M. (1998). Attention improves or impairs visual performance by enhancing spatial resolution. *Nature*, 396(6706), 72–75. <https://doi.org/10.1038/23936>
- Yon, D., & Frith, C. D. (2021). Precision and the Bayesian brain. *Current Biology*, 31(17), R1026–R1032. <https://doi.org/10.1016/j.cub.2021.07.044>
- Yon, D., Gilbert, S. J., de Lange, F. P., & Press, C. (2018). Action sharpens sensory representations of expected outcomes. *Nature Communications*, 9(1), 1–8. <https://doi.org/10.1038/s41467-018-06752-7>
- Yon, D., & Press, C. (2017). Predicted action consequences are perceptually facilitated before cancellation. *Journal of Experimental Psychology: Human Perception and Performance*, 43(6), 1073–1083.