Multi-stability in visual perception and eye movements : does action control perception or vice-versa?

Kevin Parisot

Université Grenoble Alpes

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE GRENOBLE ALPES

Spécialité : **SIGNAL IMAGE PAROLE TELECOMS**

Arrêté ministériel : 25 mai 2016

Présentée par

Kevin PARISOT

Thèse dirigée par **Steeve ZOZOR, Directeur de recherche, CNRS**, et
codirigée par **Ronald PHLYPO, Maître de conférences, Grenoble-INP**,
et **Alan CHAUVIN, Maître de conférences, UGA**

préparée au sein du **Laboratoire Grenoble Images Parole Signal Automatique (GIPSA-lab)** et du **Laboratoire de Psychologie et NeuroCognition (LPNC)**
dans **l'École Doctorale Electronique, Electrotechnique, Automatique, Traitement du Signal (EEATS)**

La multi-stabilité dans la perception visuelle et les mouvements oculaires : l'action contrôle-t-elle la perception ou vice-versa ?

Multi-stability in visual perception and eye movements: does action control perception or vice-versa?

Thèse soutenue publiquement le **14 octobre 2020**,
devant le jury composé de :

Monsieur Jean-Luc SCHWARZ
Directeur de recherche, CNRS Alpes, Président du Jury
Madame Anna MONTAGNINI
Chargé de recherche HDR, CNRS Provence et Corse, Rapporteur
Monsieur Ralf Engbert
Professeur, Univeristé de Potsdam, Rapporteur
Monsieur Laurent MADELAIN
Professeur des Universités, Université de Lille, Examinateur
Monsieur Steeve ZOZOR
Directeur de recherche, CNRS Alpes, Directeur de thèse
Madame Anne GUERIN-DUGUE
Professeur des Universités, Université Grenoble Alpes, Invitée
Madame Aurélie CAMPAGNE
Maître de Conférences, Université Grenoble Alpes, Invitée

Multi-stability in visual perception and eye movements: does action control perception or vice-versa?

Kevin PARISOT

January 13, 2021
Version: final submission

Université Grenoble Alpes

Grenoble-Image-Parole-Signaux-Automatique (GIPSA-lab), Laboratoire de Psychology et NeuroCognition (LPNC)
Ecole Doctorale Electronique Electrotechnique Automatique et Traitement du Signal (EEATS)
Université Grenoble Alpes

Cognitive Sciences

Multi-stability in visual perception and eye movements: does action control perception or vice-versa?

Kevin PARISOT

1. Reviewer Anna MONTAGNINI
CNRS, Institut de Neurosciences de la timone (INT)
Aix Marseille Université

2. Reviewer Ralf ENGBERT
Experimental and Biological Psychology
Universtät Potsdam

Supervisors Alan CHAUVIN
Ronald PHLYPO
Steeve ZOZOR

January 13, 2021

Kevin PARISOT

Multi-stability in visual perception and eye movements: does action control perception or vice-versa?

Cognitive Sciences, January 13, 2021

Reviewers: Anna MONTAGNINI and Ralf ENGBERT

Supervisors: Alan CHAUVIN, Ronald PHLYPO and Steeve ZOZOR

Université Grenoble Alpes

Université Grenoble Alpes

Ecole Doctorale Electronique Electrotechnique Automatique et Traitement du Signal (EEATS)

Grenoble-Image-Parole-Signaux-Automatique (GIPSA-lab), Laboratoire de Psychology et NeuroCognition (LPNC)

Domaine Universitaire, Saint-Martin-d'Hères

38400 and Grenoble

Abstract

Multi-stable perception refers to the perceptual dynamics that emerge when the human's sensory system is confronted with a stationary, ambiguous visual stimulus. Though the stimulus is stationary, humans observe alternations in perception. Since its first observations by Necker (1832), multi-stable perception has been a tool to investigate the inference processes of the visual system when reconstructing a rich perceptual world from incomplete, and sometimes poor, sensory information.

In this thesis, the relationships and dependencies between the oculomotor action and perceptual systems are approached in the context of multi-stable perception. The main questions driving the investigation can be formulated as follows: can we infer percepts from oculomotor behaviour and thus, what are the (hierarchical) relations between the perceptual and oculomotor systems?

At first, we have studied ocular micro-movements in fixations that have been detected during exploration of a bi-stable visual stimulus with motion. We propose to classify them as micro-pursuit; a class of fixational eye movements, correlating with smooth, predictable, small-scale stimulus' target trajectories. We replicated these findings in an explicit pursuit task with a luminance change detection task, but only when the moving object was a target, and not when it was a distractor. Inter-experiment analysis suggests that the manipulation of task, stimulus target motion, and the level of ambiguity of the stimulus affect the generation of micro-pursuits: a result that may hint that bi-stable perception may play a role in the oculomotor decision to attend either the fixation cross, or the moving object.

We have modelled this behaviour with a predictive model based on an energy potential field in which gaze is represented by the dynamics of a unitary mass. We further extend this model to capture multi-stable perception. An exploration of the model's capacity to reproduce fixational eye movement—covering micro-saccades, micro-pursuits and stable fixations—is presented.

To further study perceptual multi-stability and oculomotor control, we used the moving plaid: a tri-stable stimulus, composed of two transparent gratings moving in different directions and visualised through an aperture, making perceived motion direction ambiguous. We investigate how the plaid's ambiguity can be manipulated at the individual subject level, using a probabilistic model and an experimental

protocol to estimate its parameters. As such, points of maximal ambiguity can be identified for an observer based on the manipulation of the gratings' transparencies. We further looked at oculomotor manipulation in the context of the moving plaid stimulus and provide a brief outlook at a no-report paradigm that aims to exploit eye movements to infer the perception dynamics of an observer.

This exploration aims to provide a road map for further investigation of perceptual and oculomotor coupling in multi-stable perception, and opens up to methods using neuro-imaging techniques to investigate multi-stability. The work presented here raises questions on the link between stability regimes and how bottom-up and top-down processes may play a role in modulating the brain into mono-, multi- or meta-stable dynamics.

Science popularisation abstract

Human vision enables us to acquire information about our environment, from a distance and with precision. However, the eye is a heterogeneous photon sensor with low resolution, when compared to cameras, but the brain has efficient hidden processes. It uses eye movements to create rich perceptions and thus, it is able to solve visual ambiguity problems. Here, we will show how multi-stability, a type of illusion where one's perception changes while the stimulus stays the same, can help us make a link between vision and action. We will present the micro-pursuit: a small eye movement used to track an object. We will use gravity to predict the relationships between action and perception. We will manipulate a moving plaid's ambiguity, a multi-stable stimulus, and will propose ideas to help us decipher perception in eye movements. At last, we ask, what does stability mean for perception?

Résumé

La multi-stabilité dans la perception visuelle et les mouvements oculaires : l'action contrôle-t-elle la perception ou vice-versa ?

La multi-stabilité perceptuelle désigne la dynamique perceptuelle, qui émerge lorsque le système sensoriel humain est confronté à une stimulation stationnaire et ambigüe. Dans ce cas, la perception de l'observateur alterne entre différents percepts, tandis que le stimulus reste constant. Depuis les premières observations de Necker (1832), la perception multi-stable est devenue un outil de recherche, utilisé dans les recherches visant à comprendre comment le système visuel se sert de mécanismes d'inférences, afin de reconstruire un monde perceptuel riche à partir d'informations sensorielles incomplètes.

Dans cette thèse, nous cherchons à étudier les relations d'inter-dépendence entre l'action oculomotrice et le système perceptuel, dans le contexte de la multi-stabilité perceptuelle. Les questions centrales de recherche sont les suivantes : est-il possible d'inférer les percepts à partir des mouvements oculaires, et si oui, comment se constituent les relations (hierarchiques) entre les systèmes perceptuel et oculomoteur ?

Tout d'abord, nous avons étudié les micro-mouvements oculaires fixationnels, que nous avons détecté dans une tâche d'exploration d'un stimulus bi-stable en mouvement. Nous proposons de classifier ceux-ci en tant que micro-poursuite, une classe de mouvements oculaires fixationnels, corrélant avec des trajectoires lisses, prévisibles et de faibles amplitudes du stimulus en mouvement. Nous avons reproduit ces résultats dans une tâche de poursuite explicite, accompagnée d'une tâche de détection de changement de la luminance, mais uniquement lorsque l'objet était la cible, et pas lorsqu'il était un distracteur. L'analyse inter-expérience suggère que la manipulation des tâches, du mouvement du stimulus, et du niveau d'ambigüité du stimulus, influe la génération de micro-poursuite: un résultat qui indique que la perception bi-stable pourrait jouer un rôle dans la décision oculomotrice de porter son attention sur la croix de fixation, ou l'objet en mouvement.

Nous avons modélisé ce comportement à l'aide d'un modèle prédictif, basé sur un champ d'énergie potentielle, dans lequel le centre du regard est représenté par

la dynamique d'une masse unitaire. Nous avons ensuite étendu ce modèle pour qu'il rende compte de la multi-stabilité perceptuelle. Une exploration des capacités du modèle à reproduire les mouvements oculaires fixationnels – en considérant micro-saccades, micro-poursuites et fixations stables – est présentée.

Afin d'étudier davantage les liens entre multi-stabilité perceptuelle et contrôle oculomoteur, nous nous sommes servis d'un stimulus : le plaid en mouvement, composé de deux grilles transparentes se déplaçant dans différentes directions et observées au travers d'un trou circulaire, ce qui engendre une tri-stabilité vis-à-vis de la direction du mouvement perçue. Nous examinons comment l'ambigüité du plaid peut être manipulée au niveau de chaque sujet, en utilisant un modèle probabiliste et un protocole pour en estimer ses paramètres. Ainsi, les points d'ambigüité maximale peuvent être identifiés pour chaque observateur.rice, en manipulant la transparence des grilles. Enfin, nous avons aussi regardé comment manipuler les mouvements oculaires avec le plaid en mouvement, et nous donnons un bref aperçu sur l'application d'un paradigme, sans rapport explicite de la perception, qui a pour but d'inférer cette dernière à partir de la dynamique oculomotrice. Cette exploration cherche avant tout à proposer une feuille de route, afin de poursuivre les questions de recherche autour du couplage entre les systèmes oculomoteur et perceptif, lorsqu'il y a multi-stabilité, et d'ouvrir vers l'utilisation de techniques de neuro-imagerie, appliquées à cette thématique.

Enfin, le travail présenté dans cette thèse pose des questions sur les liens entre les différents régimes de stabilité – tel que la mono-, multi-, ou méta-stabilité – et quel est le rôle des processus ascendants et descendants sur ceux-ci.

Résumé vulgarisé : La vision nous permet d'acquérir de l'information sur notre environnement, à distance, avec précision. L'oeil n'étant qu'un capteur de photons hétérogène de faible résolution, comparé aux appareils photo, le cerveau doit donc avoir des techniques efficientes cachées. Il utilise les mouvements des yeux pour créer des perceptions riches et ainsi, il est capable de résoudre des problèmes d'ambiguïté visuelle. Nous montrerons comment la multi-stabilité, un type d'illusion où la perception d'une personne change mais le stimulus reste le même, permet de mettre en lien vision et action. Nous présenterons la micro-poursuite : un petit mouvement de suivi d'un objet. Nous utiliserons la gravité pour prédire des liens entre action et perception. Nous manipulerons l'ambiguïté du plaid en mouvement, un stimulus multi-stable, pour évoquer les pistes devant permettre de décoder la perception dans les mouvements des yeux. Finalement, quel est le sens de la notion de stabilité dans la perception ?

Contents

Avant-propos / *Preface*

Dans cet avant-propos, je souhaite partager une touche d'humanité et de gratitude envers les personnes qui ont fait partie de ces années, de près ou de loin. Cette partie est écrite dans un mélange de Français et d'Anglais.

I first want to share a touch of gratitude to those who participated closely, or from far away, in my life during the years of the research presented in this manuscript. This part is written in a mixture of French and English.

Anne Guérin-Dugué, qui m'a invité à explorer le sujet fascinant de la multi-stabilité visuelle lors de mon stage de Master, et qui a su m'encourager à développer mes idées tout en prenant soin de les mettre à l'épreuve de son examen critique. Je suis aussi ravi que nous ayons, par ce projet de thèse, réussi à intéresser Steeve Zozor, sur un terrain qui n'est pas forcément le sien, et cette rencontre a permis de nombreux échanges passionnants, permettant ainsi de développer l'aspect modélisation de cette thèse. La particularité de ce travail se centre autour de son approche inter-disciplinaire, qui fut possible par la vivacité, la curiosité mais aussi la rigueur de Ronald Phlypo, qui malgré ses nombreux cours et obligations d'enseignements, a réussi à me transmettre un goût pour l'exploration de méthodes et de thématiques, alliant l'éclectisme des sujets d'intérêt à un scepticisme scientifique nécessaire. Enfin, je remercie Alan Chauvin d'avoir su pimenter cette expérience d'une légère chaleur et d'amusements, lui aussi malgré des obligations d'enseignements parfois lourdes, et cela, dans l'ouverture d'esprit, l'écoute et la rigueur de pensée essentielle dans un projet de recherche aussi varié que le nôtre ; ses mots furent par moments de réelles éclaircies dans la brume, parfois étouffante, d'un projet de recherche.

D'autres chercheur.se.s et collègues ont contribué aux idées travaillées et présentées dans ces pages. Dans le monde des francophones, je souhaite partager ma gratitude envers Jean-Michel Hupé, Nathalie Guyader, Simon Barthelmé, Aurélie Campagne, Jean Lorenceau, Julien Diard, Nathan Faivre, Jean-Luc Schwarz, Eric Castet, Laurent Madelain, Anna Montagnini, Sylvain Harquel, Laurent Ott, David Alleysson, Carole Peyrin, Antoine Coutrot, Hélène Devillez, Suzon Ajasse, David Méary, Pierre Comon, Gang Feng, Mauro Dalla Mura, Stéphane Rousset, Eric Guinet, Cédric Pichat et Marcela Perrone-Bertolotti.

D’autres ont aussi été présents et ont fait partie de ce voyage durant des échanges à la pause café, dans les couloirs, ou en dehors. Les collègues du GIPSA-lab : Camille, Emmanuelle, Quentin, Pedro, Pierre M, Raphaël, Jeanne, Imane, Cosme, Marc, Marion, Marielle, Julien, Bruce, Miguel, Lorenzo, Florent, Taia, Victor, Pierre N, Aziliz, Roza, Thibaut, Louis, Silvain, Gael, Tayeb, Malik, Allison, Christian, Bertrand, Marco, Olivier et Pierre-Olivier.

D’autres dans le laboratoire LPNC : Alexia, Emilie, Audrey M, Louise, Hélène M, Lisa, Ladislas, Chloé, Violette, Brice, Lea, Adeline, Méline, Audrey T, Elie, Olivier, Benjamin, Eve, Christopher, Laura, Lucrèce, Amélie, Richard, Martial et Hélène L.

I also want to thank the international researchers Rubén Moreno-Bote and Ralf Engbert for their support and their contributions.

Ces recherches ont aussi avancé grâce aux travaux de stagiaires enthousiastes et motivés, qui je l’espère, auront tiré une expérience de recherche intéressante dans leur collaboration. Je tiens donc à remercier Florian Millecamps, pour son travail sur les effets des consignes et de la volition sur la perception multi-stable, Maryiem Ahmida, qui nous a suivi sur deux stages autour du rôle de l’adaptation dans la perception multi-stable, Eva Aprile, qui a contribué largement sur les travaux d’expérience sans report explicite et subjectif, Emeline Lalisse, qui a repris les travaux sur l’adaptation, et enfin Juliette Lenouvel, qui a entamé l’expérience ’Gaze-EEG’ et qui poursuivra mes travaux dans une thèse que j’ai hâte de découvrir.

Évidemment, cette épreuve n’aurait pu être possible sans le soutien et l’amour que ma famille a su me donner, durant ces quatre années mais aussi les vingt-six qui ont précédés. Maman, Papa et Marc, je vous remercie donc d’une force éternelle. Merci aussi à la famille Parisot.

My gratitude also goes to Anand, Shankar, Sweata, Velumama, Mami & Aya.

Je remercie aussi les professeur.e.s qui m’ont aidé à grandir dans mon parcours académique, et qui ont permis cet aboutissement. Leurs rôles dans le parcours long qui permet de faire une thèse est non-négligeable.

Une thèse se vit dans un contexte de vie en dehors des laboratoires aussi. Le mien fut à Grenoble, une ville merveilleuse, vivante, et d’ouverture. Cette vie s’est manifestée surtout par de formidables rencontres amicales. Des amis qui ont su me faire rire, rêver, jouer, pleurer, changer, me stabiliser et me déstabiliser ; des amis qui m’ont fait vivre des années inoubliables. Je veux donc remercier les personnes qui étaient près de moi durant ces années. Celles et ceux au plus près, mes colocataires au fil des années : Émile, Erwan, Hugo, Angèle, Laura, Matthias, Éléonore, Anaïs,

Mika, Valentin, Camille, Marie, Solène et Louise. Je n'oublie pas nos chats Tatane, Mapouche et Babybelle, ainsi que les poules et le pioupiou.

Celles et ceux du théâtre : Fernand, Letizia, Margherita, Laurine, Mélanie, Julien, Lisa, Mary, Juliette T, Benjamin T, Benjamin V, Louise, Hélène, Thierry, Jean, Lucie, Aurélie, Alexis, Morgan, Joris, Philippe, Jérôme, Roberto, Virginie, Romano, Jenny, Germán, Hiba, Clara, Manu, Benjamin P, Félix, Marie G, Carole, Rachel, Nicolas, Elisa, Juliette L, Nastasia, Loona, Manon, Babeth, Vaïk, Hubert, John, et tant d'autres.

Aussi les amoureux et amoureuses de la montagne et des bars : Aurore, Tom, Ophélie, Mathieu, Ety, Charlotte, Kilian, et Clément. Et celles et ceux qui furent plus loins : Pierre-Alexandre, Touraj, Pierre, Dominique, Stephan, Adrien, Anne, Noémie, Jade, Alex, Luc, Mikey, Marie, Fiona, Elisabeth, Guillaume, Oulfa, Georges, et Lou.

And the international friends : Florencia, Eugenio, Emma, Eliad, Calleb, Tom, Virginia, Nina, Charlotte, Federico, and Alberto.

Enfin, il y a eu les indispensables présences furtives qui ont accompagné cette traversée. Alban, je te remercie d'être resté lié à moi par cette force de curiosité et d'enthousiasme que nous partageons. Nos pousses se sont nourries mutuellement durant ces années. Malgré la distance, nous avons pu partager des moments festifs aux quatre coins de l'Europe, des échanges constructifs sur nos thèses respectives, et des visions sur ce monde que nous cherchons tous les deux à comprendre. Cette thèse, je pense l'avoir écrite avec ta main bienveillante posée sur une de mes épaules. Antoine, j'imagine que ta main était posée sur l'autre épaule, et ton énergie, ta ressource, et ton ouverture, se sont écoulées tranquillement sur moi tout du long de cette épreuve. Je te remercie aussi pour tous ces échanges festifs, scientifiques, et philosophiques, malgré la distance, là aussi. Simon, tu as été à mes côtés dans le quotidien pendant ces années, et partager ces murs avec toi, s'ancrer ensemble dans la terre des taillées aura été source d'humour, de bringues, de délires et d'attention. Ta force et ta liberté sont contagieuses : merci. Gallia, ta générosité, ta force de vie, ton envie et ta gentillesse sont sans limites. Tu as été la brise chaude issue d'une rouge flamme qui m'a animé et émerveillé : je te remercie intensément. Enfin, Margaux, l'harmonie de nos souffles aura été l'atout de ma résilience dans ce voyage. Ta justesse, ta tendresse, ton écoute, et ton regard ont fourni les impulsions qui m'ont permis d'avancer, malgré toutes les épreuves, dans la douceur de la confiance partagée. Par nos songes partagés et l'élan de nos vitalités, je te remercie.

Puisqu'une thèse est aussi un cadre qui permet la réflexion intellectuelle, la prise de recul et l'ouverture, je souhaite enfin partager aux lecteurs et lectrices quelques

dernières pensées. Le contexte, dans lequel ce manuscrit a été écrit, peut se peindre tel un monde où les sciences modernes ont pu se développer grandement en quelques siècles, en parallèle avec des technologies qui ont, pour certaines, permis les travaux présentés. Ce monde technique domine de façon incontrôlée et dévastatrice la planète sur laquelle nous vivons tous, la Terre, au point que nous ayons nommé notre époque l'Anthropocène, une période où les activités humaines marquent majoritairement la géologie, où la biodiversité des écosystèmes s'effondrent, et où la composition atmosphérique est modifiée d'une ampleur sans précédent. Je souhaite simplement rappeler que la lutte contre la dérive de l'équilibre de ce monde est une épreuve collective débordant de sens. Un monde est complexe, et il pourrait bien, lui aussi, être multi-stable.

Publications & Talks

Articles

JOURNAL ARTICLE: **Parisot K.**, Chauvin A., Guérin-Dugué A., Phlypo R., & Zozor S. (*accepted*), Micro-pursuit: a class of fixational eye movements correlating with smooth, predictable, small-scale target trajectories, *Journal of Vision*.
CONFERENCE ARTICLE: **Parisot K.**, Chauvin A., Phlypo R., & Zozor S. (2019), Modélisation de l'ambigüité d'une multi-stabilité visuelle, *GRETSI 2019*.

Conferences

POSTER: Ahmida M., **Parisot K.**, Guérin-Dugué A., Phlypo R., Zozor S & Chauvin A. (2019), Does multi-stability involve adaptation? *GDR Vision 2019*.
POSTER: **Parisot K.**, Chauvin A., Phlypo R., & Zozor S. (2019), Modélisation de l'ambigüité d'une multi-stabilité visuelle, *GRETSI 2019*.
POSTER: **Parisot K.**, Chauvin A., Guérin-Dugué A., Phlypo R., & Zozor S. (2018), Quantifying pursuit movements through measures of spatio-temporal similarity, *GDR Vision 2018*.
POSTER: **Parisot K.**, Chauvin A., Phlypo R., & Zozor S. (2018), Multi-stability and fixationnal eye movements: an energy potential fields modeling approach, *Grenoble Workshop on Models and Analysis of Eye Movements*.
TALK: **Parisot K.**, Chauvin A., Phlypo R., & Zozor S. (2018), Modeling multi-stability and fixational eye movements, *Grenoble Workshop on Models and Analysis of Eye Movements*.
POSTER: **Parisot K.**, Chauvin A., Guérin-Dugué A., Phlypo R., & Zozor S. (2017), Predictable motion on a Necker cube leads to micro-pursuit-like eye movements and affects the dynamics of bistability. *19th European Conference on Eye Movements 2017*.
TALK: **Parisot K.**, Chauvin A., Guérin-Dugué A., Phlypo R., & Zozor S. (2017), A multi-stable gravitational potential approach to fixational eye movements., *European Conference on Visual Perception 2017*.
POSTER: **Parisot K.**, Guérin-Dugué A., & Phlypo R. (2016), Influence of visual

motion on reversal speeds in an ambiguous Necker cube experiment: an eye-tracking study. *GDR Vision 2016*.

Invited Talks in Laboratories & Seminars

TALK: **Parisot K.** (2019), Une cognition multi-stable ? Vision du cerveau en système complexe dynamique, *Rencontres Jeunes Chercheur·euses du LPNC*.
SEMINAR: **Parisot K.** (2018), Multi-stability & fixational eye movements: an energy potential approach, *SFB 1294 Data Assimilation - the seamless integration of models and data, Universität Potsdam*, invited by Prof. Ralf Engbert.
TALK: **Parisot K.**, Chauvin A., Guérin-Dugué A., Phlypo R., & Zozor S. (2017), Multi-stability & fixational eye movements: an energy potential approach, *Grenoble Fest @ LPP*, invited by Prof. Thérèse Collins.

Scientific Outreach

Parisot K., & Chauvin A. (2019), Quand voir... c'est faire! *Pint of Science Grenoble 2019*

Seminars Hosted

Moreno-Bote R. (2017), How does the brain solve causal inference? *GIPSA-lab, Images & Signal department seminar*

Before reading

Visual code for reading

This manuscript uses a visual code to provide the reader with different levels of depth and importance. Four levels are used :

> **Highlights and important content are displayed in bold quote format.**

Regular content is written in classical form such as displayed here.

Boxes.
Boxes are used to provide additional optional information which can be read independently on a topic. They are also used to provide reading instructions and publication information.

And footnotes are used for optional additional information in the flow of reading[1].

Manuscript architecture

The manuscript is composed an introductory chapter, a chapter with results for eye movement research, a chapter with exploratory works on theoretical approaches, a chapter with results for ambiguous visual perception research and a perspective chapter. The decision to follow such a narrative patterns was made to provide the building blocks towards the work and ideas presented in the final chapter.

[1] For instance, when a short additional comment might provide useful information to the reader but is not sufficiently developed to fit a box.

1 Ambiguity for the human visual system

> *The beauty of a living thing is not the atoms that go into it, but the way those atoms are put together. Information distilled over 4 billion years of biological evolution. Incidentally, all the organisms on the Earth are made essentially of that stuff. An eyedropper full of that liquid could be used to make a caterpillar or a petunia if only we knew how to put the components together.*
>
> ***— Carl Sagan***
> **"Cosmos: A Personal Voyage".**
> **Documentary (1990 Update).**
> **Episode 5: "Blues For a Red Planet", 1990.**

Visual experience has fascinated thinkers and scientists as far as written records exist. As one of humans' richest sensory modality, vision has driven many forms in arts, but it also has driven physicists to develop methods in optics to observe and study the stars, and now the cosmos. But what are the mechanisms proper to human vision? And how does our brain handle such amounts of information, and combines it with actions? A general review of what is known on perception, and more specifically visual perception, shall be given. More specifically, we will then delve into vision and how the multi-stable phenomenon arises in it, as well as the experimental methods used in the scientific literature. Finally, the motivations for studying multi-stability will be presented and the gaps in the literature will be identified.

Reading instructions.
This chapter is designed for vision researchers and dives into multi-stable perception. However, readers wishing further context, notes and reviews may be interested to follow the many pointers presented in boxes, for a more complete and introductory understanding of cognition and vision.

Contents

Theoretical context.
Readers who wish to read into the theoretical and philosophical context in cognitive sciences surrounding this work can do so in Appendix A.1.

1.1 Visual perception

Visual perception is the phenomenon that allows humans to sense their environment based on photons, in other words light, that is emitted or reflected. The environment can be composed of objects or events. Visual perception is, foremost, a set of processes used to acquire knowledge. These processes manipulate visual information that is captured by sensors: the **eyes**. Though there is a purely optical and physical aspect, visual perception also encompasses the cognitive activities that process the visual flow of information from the eyes. This is where it differs from cameras, for instance: they do not have knowledge of the scenes they record, whereas a human visually perceiving does. In this section we shall give the reader a brief overview of the stages and processes that visual information travels through in order to reach awareness or conscious experience.

From the eyes to the brain.
Readers who wish to understand the basis of low-level vision and how light is transformed into perception can read Appendix A.2 where an introduction to the field is given by following the information pathways: from the eyes to the brain.

1.1.1 A brief look at some basics of visual perception

Importantly, the visual system is complex and processes light information in such a way that it provides a rich reconstruction of the world.

Human vision has key features such as heterogeneous sensors with its retina coverage being unequal across the visual field it receives light from (Curcio and Allen, 1990). It also has multiple pathways to relay the sensory information to the occipital cortex. Visual information is projected in a retinotopic fashion in the cortex and neural signals encoding it, are processed through layers and fed forward to other areas of the brain (Wandell, 1995; S. Palmer, 1999). For more details, read Appendix A.2.

When we become visually aware of a perceptual object, its information flows in multiple networks ranging from the frontal to the parietal cortex. This process is not instantaneous—and though it may vary depending on the complexity of the

stimulus, observer habituation and more factors—, to become quasi-fully conscious[1] of a perceived object, the process may take at least 200 milliseconds (ms) (Kornmeier and Bach, 2012). Visual illusion, phenomena in which the visual system's expectation are tricked into building an incoherent representation of the stimulus, have been used in vision science to identify the processes as inferences, and more specifically as predictive coding (Friston and Kiebel, 2009). Moreover, illusions, such as multi-stable perception, are being used to uncover the neural correlates of consciousness and perception (Frässle et al., 2014).

The visual system is typically considered as a hierarchical structure of networks processing the information flowing in one global direction, that is upwards—namely from the retina to higher cognitive functions in the brain. Vision scientists consider the biological systems' inputs, the eyes, the lower level of this model, and the consciously experienced *qualias*—a proposed term to refer to a granular unit of conscious subjective experience, analogous to the *quantas* in physics (Dennett, 1993; Chalmers, 2007)—, the higher level of the system. The eyes and their movements are key properties of visual perception. Indeed, though our eyes could be considered as poor sensors, from an engineering perspective, we move them to capture information from different areas of the visual field, and to process it efficiently. Eye movements are typically broken down in three categories, when studied in the context of cognitive tasks: saccade, pursuit and fixation (Liversedge et al., 2011). However, they never remain still, and during fixations, micro-saccades, drifts and tremors can be observed (Martinez-Conde, Macknik, and D. Hubel, 2004).

> **These eye movements create changes of the visual input on the photoreceptor cells of the retina by shifting the retinal projection constantly. But the visual system likely evolved with this constraint, and is thought to exploit these noisy features to increase its capabilities (Hicheur et al., 2013; Rucci and Victor, 2015).**

A growing view in the field of cognitive sciences requires that the body and action be considered as part of cognition—in embodied or enactive cognition (Varela, 1996a). Linking oculomotor action to visual perception is becoming increasingly evident as eye movements show potential to be physiological markers of internal cognitive states (Kagan and Hafed, 2013; Spering and Carrasco, 2015; Shaikh and Zee, 2018); whether it be attention (Kuhn et al., 2009; Orquin and Loose, 2013; Denison et al., 2019), perception (Gold and Shadlen, 2003; Hafed and Krauzlis, 2006; Schütz, D. I. Braun, and Gegenfurtner, 2011; Boi et al., 2017; Kagan and Burr, 2017), learning

[1] Consciousness is not a quantifiable and measurable property of neural information, at the time of writing. And though we might be aware of some properties of a perceptual object, our conscious representation may neglect more fine details of the object.

and development (Eckstein et al., 2017), language processing (Engelmann et al., 2013), and reading (Kliegl et al., 2004). Though the lower parts of the visual system are starting to be well understood, the higher one goes along the visual hierarchy, the less clear network architectures and causal relationships become.

In fact, the visual system is full of asynchronous feedback mechanisms that make deciphering its workings a very complex task. For instance, parts of the visual signals are fed, as they go through the lateral geniculate nucleus (LGN), to the superior colliculus (SC) which has been correlated to oculomotor programming with other cortical areas such as the frontal eye field (FEF) and the lateral intra-parietal (LIP) cortex (Krauzlis, 2004; Hafed, Goffart, et al., 2009; Taouali et al., 2015; Peel et al., 2016; Krauzlis et al., 2017). But the two later areas are also tightly correlated to attention, a higher cognitive process than oculomotor programming (Astrand et al., 2015). In fact, these two functions may share *efference copies*, a set of copies of the information for motor programming enabling the system to have different levels of engagement with its action[2] (Jeannerod and Arbib, 2003). Note that other theories, such as *referent control of perception*, may also explain the link between motor action and perception (Feldman, 2016).

> **These points raise a series of questions. How is action and perception related? What tools can be used to simplify and understand such complex and intertwined interactions in between the motor and perceptual systems? How do these systems relate to conscious experience of the world?**

These questions linking the body's actions to its internal cognitive states may give insights and leads for some of the problems introduced in Appendix A.1 on the origins of consciousness, perceptual experience and its evolution. Our attempt to contribute relies on a trans-disciplinary approach, using a combination of methods from empirical sciences—namely psychophysics and neurosciences—and theoretical research—namely signal processing and computational modelling. In this work, we focused on a visual phenomenon in particular, multi-stable perception, as it allows to study changes of internal perceptual states while the stimulation remains stationary from a physical perspective, but is changed by the eyes' constant movements. Moreover, multi-stability occurs in different modalities (Schwartz et al., 2012) and relates to the coordination of sub-systems in complex systems dynamics (Kelso, 2012). Therefore, theoretical approaches to bi-stability might give key insights relating the questions asked here (Moreno-Bote, Knill, et al., 2011).

[2] i.e., one copy may contain the information for a desired action that is not carried out, while a latter one will only keep the neural information for action executed.

1.1.2 Eye movements

The eyes move; in other words they are dynamic and active. Their motion has one major consequence: it shifts the visual content of the retinal projection, and therefore the visual input flow changes. The eyes are located in spherical sockets that allow them to rotate on themselves. The movements are controlled by six strong and precise *extraocular muscles*. In fact, the **oculomotor system**, that controls the extraocular muscles enabling eye movements, can be very dynamic and shows various behaviours ranging from stationary **fixations** to ballistic and highly dynamic **saccades**.

> **The oculomotor system's main functions allows the visual system to fixate a point in space in order to accumulate information, or to track a target by keeping it in the foveal location on the retina as it moves across the visual field.**

Oculomotor dynamics can vary extensively depending on the tasks and actions of the observer as shown in Fig. 1.1 (Yarbus, 1967). The eye movements are captured in a bi-variate signal called **gaze** that situates the foveal position on the visual field over time. It is characterised by fixations during which the gaze is stable and the retinal image motion is small, and punctuated by saccades, a class of rapid and ballistic movements for static stimulation—i.e., a still image. When stimulation is dynamic—i.e., a video display—the oculomotor system produces fixations and saccades, but also in some cases **smooth pursuit** eye movements. These pursuits are used to track a target object moving across the visual field. Eye movements have been studied in close relation to visual perception and provide key information on the retinal image variations as well as insight on visual attention (Liversedge et al., 2011; Kowler, 2011). They can also be physiological marker of internal cognitive states, and more precisely of motion perception (Just and Carpenter, 1976; Spering, Pomplun, et al., 2011; Shaikh and Zee, 2018; Boccignone, 2019).

Eye & head movements.
Though eye movements are often coupled with head motion in ecological conditions—i.e., experimental conditions that are closer to everyday life, with less control and restrictions—in this work, we do not consider the latter and its interaction because it adds another set of degrees of liberty, thus increasing the complexity of scientific investigation. We focus our review on eye movements that are considered independently from head movements and with experimental setups where these head movements are restrained.

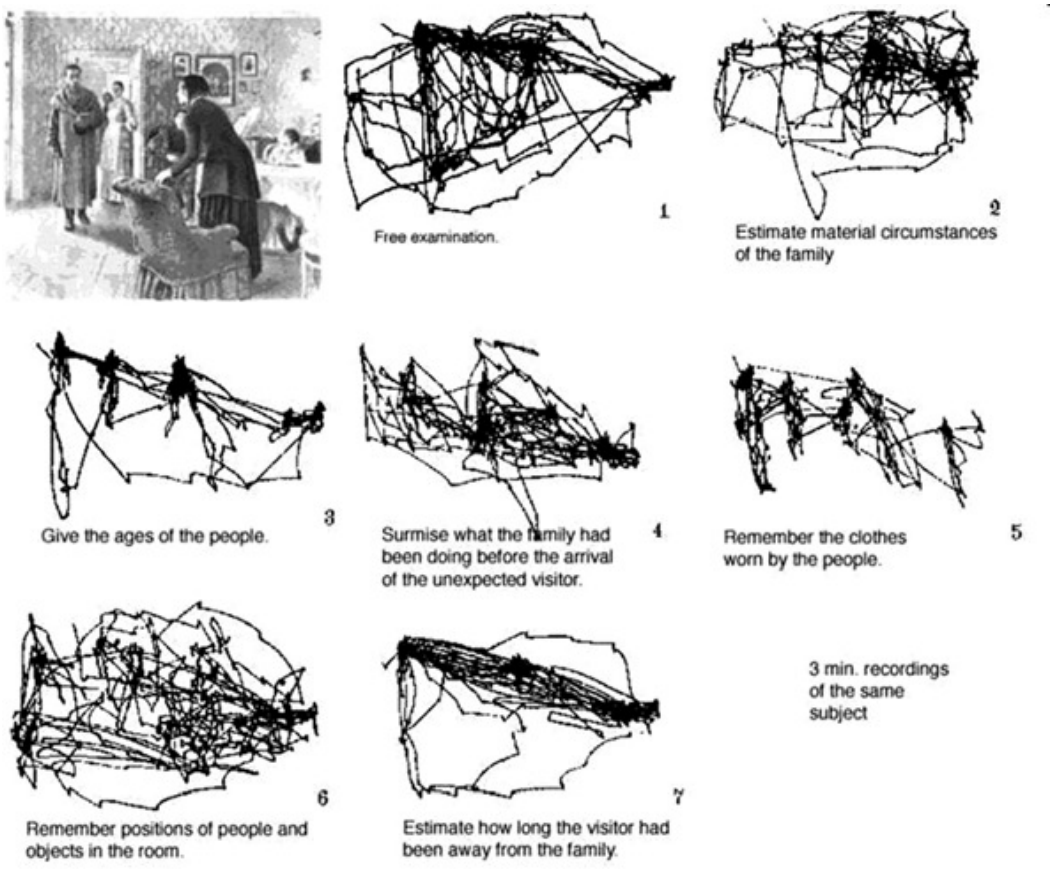


Figure 1.1. Eye movements & instructions. Figure from Yarbus (1967) showing the variation of the spatio-temporal dynamics of gaze for one stimulus. Different tasks were given and are reported bellow the 2D gaze traces.

Saccades: the rapid and ballistic movements

Saccades are fast, ballistic eye movements shifting rapidly the locus of the fovea on the visual field—diagram of a saccade shown on Fig. 1.2. They allow humans to explore, scan and search their environment by displacing the fovea, where the precise conic photoreceptor cells are located, to the area of interest. Saccades are also an energy efficient method to explore a scene (Liversedge et al., 2011) and are often used over other actions for humans, such as head or body movements. They allow the gaze to move from a spatial position to another in a scene. The oculomotor event lasts between 150 ms and 200 ms for planning and execution.

Saccades are characterised in terms of duration, amplitudes and velocities by the **main sequence**[3], a relationship that links the velocity and amplitude of a saccade to the time it takes to plan it (Bahill et al., 1975; Harris and Wolpert, 2006). The velocity of the gaze is very high relative to all other eye movements: within 30 ms, the eyes can reach a speed up to 900 visual degrees per second ($deg.s^{-1}$) (Goldberg et al., 1991). Saccades are typically defined by displacement, velocity and acceleration thresholds—above 0.15 deg, above 30 $deg.s^{-1}$ and above 9500 $deg.s^{-2}$, respectively. However other algorithms for detection exist based on adaptive methods and *glissade*[4] detection (Nyström and Holmqvist, 2010; Behrens et al., 2010) or Bayesian classification in which an algorithm learns and adapts probability functions related to motion properties of the gaze for saccade detection (Tafaj et al., 2013; Mihali et al., 2017). As you are reading this text, you are in fact doing a series of saccades, moving across words and sentences.

Although we know we can move the eyes, interestingly, our conscious visual flow seems unaffected by the movements. By moving the eye, and thus the retinal image, saccades should generate blurry moments in the visual experience. However, it is not the case as the brain applies mechanisms that guarantee visual constancy (S. Palmer, 1999). This is referred to as *saccadic suppression* and though it is highly effective in ecological conditions, some experiments have shown that transsaccadic perception can occur (Burr and Ross, 1982; Castet and G. S. Masson, 2000). Though saccades are generally direct movements, there are experimental paradigms and associated phenomena[5] that show that it is not always so.

Smooth eye pursuits: the target tracking movements

Smooth pursuits are slower eye movements that have been studied in the context of visual tracking of an object. Therefore, the function of pursuit is to maintain the tracked target on the fovea by matching the spatio-temporal properties of the target's displacement with the eyes. As a consequence, pursuits are defined as an oculomotor phenomenon with two phases: (1) a catch-up saccade followed by (2) a target pursuit or maintenance phase (Lisberger et al., 1987). Unlike saccades, pursuits are considered as *smooth* as they do not show high acceleration and jerky movements in the maintenance phase. If the tracked target has erratic motion,

[3] A notion taken from astronomy in which, when plotting colour versus brightness, all stars follow a path in that space. In eye movement research, the field hypothesises that evolution has led humans to maintain an advantageous trade off between accuracy and duration of the saccade movement.
[4] Glissades are defined as movements that follow saccades with slight overshoot and when the gaze needs to be readjusted.
[5] i.e., Anti-saccades or saccade deviation.

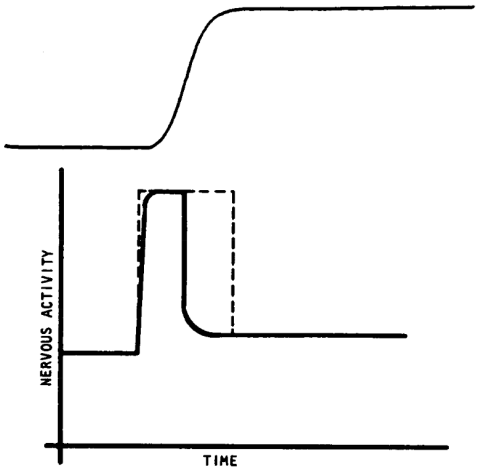


Figure 1.2. Saccade spatial shift. Diagram taken from Bahill et al. (1975) showing at the *top* the spatial shift due a 10 deg saccade and a proposed nervous signal controller associated to program it, *below*.

the oculomotor system will not track it as it becomes unpredictable and saccadic. Therefore, pursuits are slow oculomotor behaviours, comparatively, with velocities being restrained to a range of 20 to 90 deg.s^{-1} (Komogortsev and Karpov, 2013; Krauzlis, 2004) and latencies dependent on the catch-up saccade properties.

Moreover, smooth pursuit movements are dependent on visual stimulation as they attempt to fixate a moving target on the fovea by moving the eyes (Rashbass, 1961; Do Robinson, 1965; Liversedge et al., 2011). Thus, they also require constant visual feedback so that gaze can be adjusted and its position or velocity updated. Though pursuits are mostly studied with a clear and explicit target, research has shown that the phenomenon can be applied to more stimuli: random-dot kinematogram (RDK)[6] (Heinen and Watamaniuk, 1998), illusory perceptual motion (Madelain and Krauzlis, 2003) or even motion after-effect (MAE) motion (D. Braun et al., 2006). Since smooth pursuits have mostly been studied in explicit dot tracking experiments, this has constrained the development of explicit measurement and detection of the oculomotor event.

> **The functional role of the pursuit as an oculomotor process is to maintain a target of interest on the high acuity foveal region of the retina (Spering and Montagnini, 2011).**

Interestingly, its properties are linked to its definition's two phases of initiation and maintenance. For instance, detection of pursuits is based on the measurement

[6] A stimulus made of a cloud of points in which numerous points have Brownian motion while an arbitrary number of them have a coherent motion within the cloud.

of particular properties for the initiation phase (catch-up saccade); by looking at latencies between 80 to 120 ms (Krauzlis, 2004; Carl and Gellman, 1987) and retinal positioning at the centre of the fovea. Therefore, this phase has a temporal constraint that depends on saccade properties; a ballistic motion of gaze with high velocity—linked to amplitude by the *main sequence* relationship from (Bahill et al., 1975)—and the retinal position's change of location for the region of interest in the stimulus.

For the maintenance phase, measures of gaze and retinal errors and retinal slip[7] are used to verify that position, and velocity, of the gaze and target are matched, respectively (*more details on pursuit measurement in the box below*). Human observers typically track targets up to a speed of 100 deg.s^{-1} (Spering and Montagnini, 2011), though pursuits are mostly considered to be smooth and precise at speeds inferior to 30 deg.s^{-1}. It is noteworthy that the upper range leads to corruption of the pursuit epochs with catch-up saccades when velocity of target is high (De Brouwer et al., 2002). The maintenance phase, in which retinal image is stabilised, is interpreted to rely on a feedback loop where the oculomotor system must estimate and correct a velocity matching error between gaze and target.

Pursuit measurement.
Measuring the quality of tracking has been done by computing gain as a result of modelling the smooth pursuit system as closed-loop system (St-Cyr and Fender, 1969). This measure is effective in the experimental protocols in which a target appears on screen and participants are tasked to follow its motion. Pursuit is mostly studied for tracking a single point on a uniform background, however, other stimuli in motion can lead to pursuit movements (Heywood and Churcher, 1971; Heywood and Churcher, 1972). These other stimuli can lead to pursuit phenomenon in conditions—i.e., RDK (Heinen and Watamaniuk, 1998), line figures (G. Masson and Leland Stone, 2002), illusory perceptual motion (Madelain and Krauzlis, 2003) or MAE (D. Braun et al., 2006)—that are less coherent with the two phases structure described in the previous paragraphs, making it harder to detect them with these measures. The measure of gain and the models associated have been questioned for tasks where a percept is pursued, rather than a dot (Leland S Stone et al., 2000).

Fixations: the stationary state of visual accumulation

When the eyes are not moving—in between saccades and pursuits—they are stabilised in fixations. A period or epoch of the gaze signal is classified as a fixation if it cannot be classified in a type of movement and when the amplitude of displacement

[7] Retinal errors: $\boldsymbol{q}_R = \boldsymbol{q}_G - \boldsymbol{q}_S$, where $[\boldsymbol{q}_R, \boldsymbol{q}_G, \boldsymbol{q}_S]$ are bi-variate signals of the $[x, y]$ coordinates on the screen plane converted into visual degrees (deg), for the retinal image, the gaze and the stimulus, respectively. Retinal slip: $\dot{\boldsymbol{q}}_R = \dot{\boldsymbol{q}}_G - \dot{\boldsymbol{q}}_S$, which is same computation, but on the target and the eyes' velocities.

is smaller than 1 deg (Martinez-Conde, Macknik, and D. Hubel, 2004). However, the eyes never stay still. The study of fixations and fixational eye movements (FEM) have grown in the recent decades as increasingly affordable measurement equipment have facilitated this growth (Rolfs, 2009). During a fixation, information is accumulated for the visual system as the region of observation is treated by the highly sensitive and precise foveal region of the retina. Hence, during a scene exploration or search task, human observers tend to scan the visual field with saccade-fixation combinations also known as **scanpaths** (Noton and Stark, 1971a; Noton and Stark, 1971b). They are visible in Fig. 1.1 and characterise the spatio-temporal properties of an observer's oculomotor behaviour when facing a given task. In most of these tasks, the fixations tend to last on average 300 ms, though they may be much longer in other tasks.

FEM have different dynamics and are classified as **micro-saccades**, **drifts** or **tremors** (Martinez-Conde, Macknik, and D. Hubel, 2004). The dichotomy separating FEM from larger, macro, eye movements can possibly be explained by methodological constraints related to task choice, measurement equipment (Appendix A.3), analyses and classification. A possible explanation is that the reported small amplitude eye movements are miniature versions of the more studied smooth pursuits and saccades, and thus, they may have the same functional role to cognition and vision. In fact, theories that link FEM to active vision have been developed, in which the visual system uses the noisy properties of the FEM to enhance its capabilities and enable the detection of subtle orientation changes (Hicheur et al., 2013) or reach *hyper-acuity* (Poletti, Listorti, et al., 2013; Rucci, Iovin, et al., 2007)—i.e., human vision shows capacities to detect changes at smaller resolutions than their cone mosaic should allow (Appendix A.2), if no signal processing was carried out by the brain in higher parts of the visual system. However, the identification and classification of FEM, and more largely eye movements, are still debated and unsettled (Rolfs, 2009; Hessels et al., 2018).

Micro-saccades & small amplitude saccades

Micro-saccades have varying definitions and the algorithms used to detect them have changed over the years. Given that the majority of algorithms are based on thresholds, either on speed or acceleration, only threshold based algorithms will be discussed in this paper—one can refer to Hoppe and Bulling (2016) for alternative approaches. Thresholds used in these algorithms are not absolute (as used for saccade detection and definition): velocity thresholds are defined with respect to the median velocity for every trial (Poletti and Rucci, 2016; Krauzlis et al., 2017),

or by absolute deviation of the velocity distribution within the fixation combined with a binocularity criterion (Engbert and Kliegl, 2003), or even a Bayesian classifier with priors on velocity and magnitude (Mihali et al., 2017).

Once micro-saccades are detected and classified as events, it is possible to extract oculomotor drift as the complementary epochs in the signal. It is also worth noting that, though the function of micro-saccades has been, in the past, the subject of controversy, the literature now agrees that they are a small amplitude eye movement strategy used for visual exploration and acuity (Martinez-Conde, Macknik, and D. Hubel, 2004; Engbert and Kliegl, 2003; Rolfs, 2009; Kowler, 2011; Hicheur et al., 2013; Poletti, Listorti, et al., 2013). Therefore, they can be considered to have a dependency on the visual signals. And if they can be used as an exploration strategy, they might have some level of volition involved in the process. The use of volition as a criterion for oculomotor event definition and classification is unreliable, as it has been shown that dissociating voluntary oculomotor control and awareness is not straightforward (Poletti and Rucci, 2016). For instance, saccades are often produced without explicit awareness though they serve a voluntary task to find visual information.

Recent research results and reviews tend to minimise the debated multiple roles of micro-saccades, and interpret them as small amplitude saccades (Poletti and Rucci, 2016; Sinn and Engbert, 2016). Therefore, in this consideration of eye movements, micro-saccades help readjust the preferred foveal area against the stimulus' area of interest, hence having a similar functional role as saccades.

Oculomotor drifts & tremors

Once micro-saccades are detected and classified as events, it is possible to extract oculomotor drift as the complementary epochs in the signal (Cornsweet, 1956). Drifts are defined as the low frequency and large amplitude[8] component of FEM activity that lead to long term (relative to the fixation time) exploration of the area of fixation, also referred to as persistence (Engbert and Kliegl, 2004; Engbert, Mergenthaler, et al., 2011), which is a term used to describe a signal that has high correlation between its observation and its lag. Ocular drift is thought to be due to the viscosity of the medium in which the eyes rest in the socket; when fixation starts after a saccade, the eyes still have momentum in the same or opposite direction to the saccade.

[8]Relative to other FEM components; all these movements are within 1 deg of amplitude.

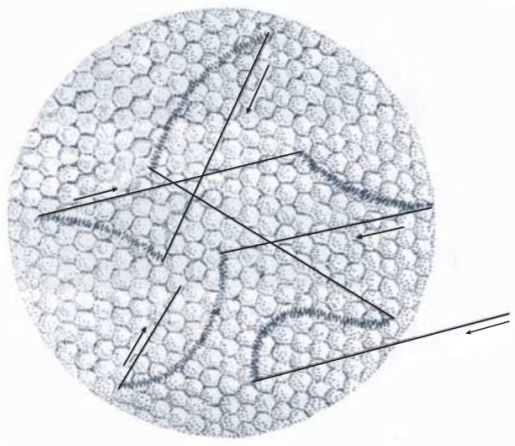


Figure 1.3. Fixational eye movements. Representation of FEM dynamics with ballistic movements corresponding to micro-saccades, low frequency oscillations for drifts and high frequency oscillations for tremor. Figure taken from Martinez-Conde, Macknik, and D. Hubel (2004).

Tremors are also notably not well studied as they remain difficult to measure and distinguish from measurement noise (Martinez-Conde, Macknik, and D. Hubel, 2004; Krauzlis et al., 2017). They represent the small amplitude oscillations in the signal (see Fig. 1.3), which can be confused with measurement noise. However, tremors are thought to originate in the noisy components of the extra-ocular muscles nervous control system. They can be characterised as a mechanism that reduces the gaze motion due to drift and keeps the locus of fovea persistently in the same area; this is often analogous to Brownian motion and random walks (Engbert, Mergenthaler, et al., 2011).

Historically, FEM are considered as noise in the oculomotor system, although this view is gradually being contested with recent evidence that they may help to relocate a preferential locus of the fovea (Putnam et al., 2005) on the scene in order to give hyper-acuity to humans for instance (Rucci and Casile, 2005; Rucci, Iovin, et al., 2007; Zozor et al., 2009; Rucci and Victor, 2015).

Other movements: vergence, vestibular & optokinetic.
More types of eye movements exist though they will not be covered in details in this work. Vergence movements are used when a tracked target approaches the observer by having the eyes move in a disconjugate manner—in smooth pursuit movements, when the target moves along a plane, like a computer screen, the eyes move in a conjugate fashion. Vestibular movements, i.e., the vestibulo-ocular reflex (VOR), correspond to compensatory eye movements, when the head moves, in order to stabilise a target on the fovea. Finally, the oculo-following reflex (OFR) occurs when a large portion of the visual field has a uniform motion across the retina (Michalski et al., 1977; Quaia et al., 2012)—e.g., when looking outside through the window while being inside a moving train. The opto-kinetic nystagmus (OKN) is a composite gaze pattern in which an object is followed by smooth pursuit until the object leaves the visual field. The eyes will maintain an object on the fovea in the slow phase until it is not possible, and find a new object to stabilise in a fast phase.

Eye movements classification

The classification of all these eye movements is a key methodological topic of the research field because other fields, such as vision science, use oculomotor events as signals giving information on the observer's task at hand and the stimulus. The definition of each category or class plays an important role on the output. Though eye movements have been studied for many decades, the variety of their dynamics and the quality of measured signals make it difficult for the research field to agree on definitions (Hessels et al., 2018). Indeed, eye movement signal classification is approached with different methods (Coutrot et al., 2018), with consequences on their interpretation (see Fig. 1.4 for an overview of methods).

The traditional approach has been to look at the *parameters* of eye movements which gives interpretability but often reduces spatial and mostly temporal information. For instance, one can look at (i) fixations' durations, numbers, dispersion or clusters, (ii) saccades' amplitude, velocity, direction, duration or latency, (iii) blinks' frequency or duration, or (iv) pupil size. Another approach is to use the *spatial distribution* of the oculomotor signal in order to derive bottom-up distribution-based metrics such as Kullback-Leibler divergence (KLD), correlation coefficient (CC), similarity (SIM) or earth moving distance (EMD) (Rajashekar et al., 2004; Le Meur, Le Callet, et al., 2006; Toet, 2011; Judd et al., 2012). Alternatively one can use location-based metrics such as the area under the curve, normalised scanpath saliency, percentage of fixation or information gain (Riche et al., 2013; Bylinskii et al., 2018). An approach, driven by web-based experimental work, where the stimulus is divided into sections, has focused on *string-based*, *common scanpath* and *geometric* comparisons of gaze signals (Le Meur and Baccino, 2013; N. Anderson et al., 2015). Finally,

the *probabilistic* approaches use transitions matrices, graphs, Markov processes, Fisher vectors, Gaussian mixture model or spatial point processes (Galdi et al., 2016; Boccignone, 2019).

The diversity of methodologies presented here (Fig. 1.4) can be explained by the multitude of contexts in which eye movements are studied and by the multi-variate aspect of the signals extracted in eye movement studies.

Eye tracking.
For further information on eye tracking signals and apparatus, Appendix A.3 provides a review.

Now that we have covered the foundations of eye movements' scientific literature, one can go back towards the visual information processing in the brain, while considering that the sensory input is subject to variation with eye movements.

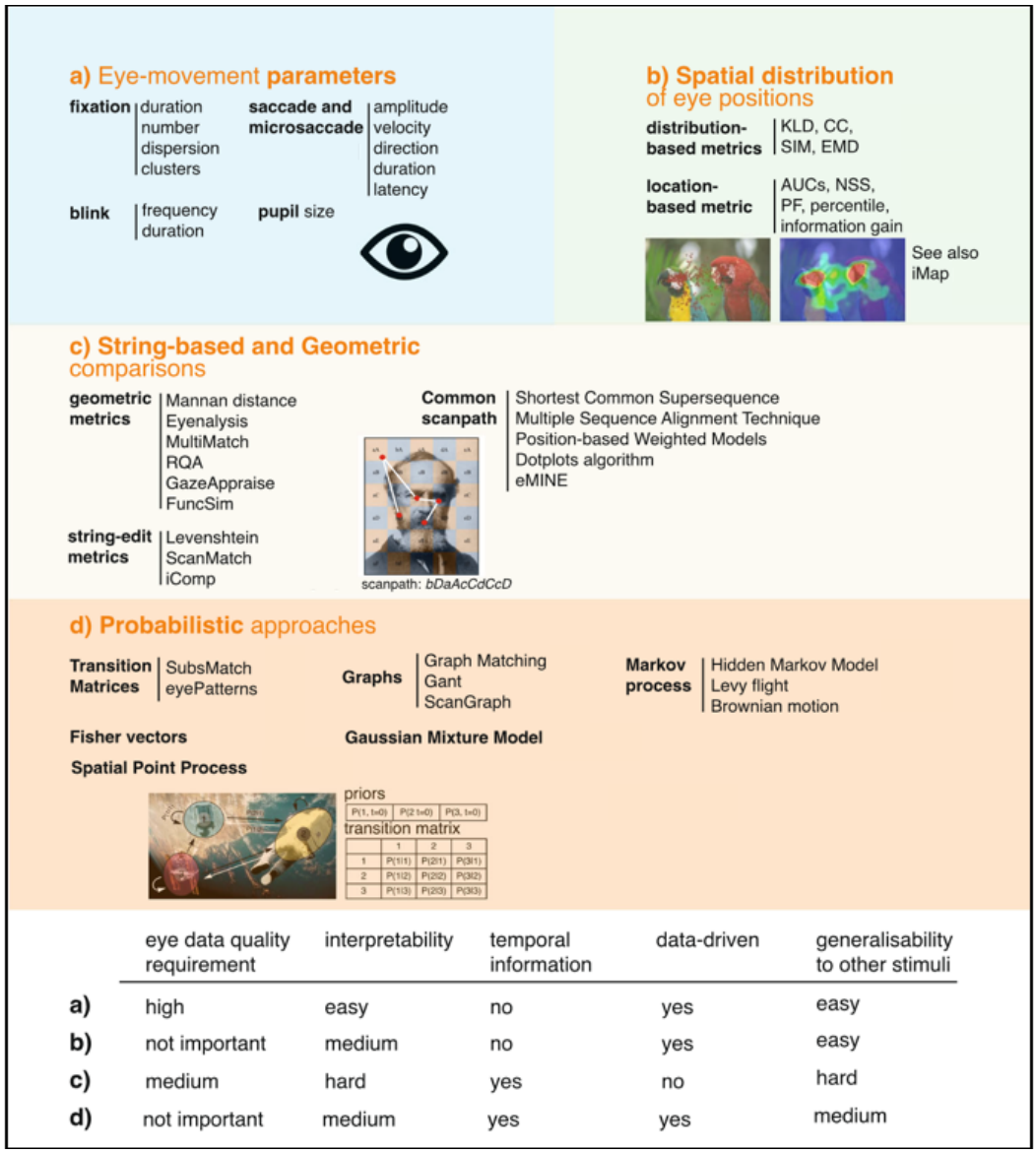
1.1.3 Visual perception as an inference mechanism

In the next paragraphs, we look at how vision is an active inference process, where sensory information is mixed with prior information, to reconstruct a rich perceptual world, from an incomplete, and sometimes poor, sensory world.

Visual inferences

Vision is often considered as a *feed-forward process* that progressively extract features based on the statistics of the retinal image. Also, the system keeps a map of the visual space across the different stages of the visual system—V1, V2, V3, etc. Neural network models, for instance, are inspired by our knowledge of the human visual system, capable of achieving very high performances in detection, recognition and classification tasks (Castelluccio et al., 2015; Gide, Karam, et al., 2017). However, this hierarchical, feed-forward view is still too simple to reflect natural cognition and perception[9]. Visual perception involves more functions and capacities than passive observation and classification of scenes or objects. Perception in humans and other mammals is an active process and interacts with many other systems and may have evolved in only a few hundred thousand years from simple light detector to a complex system capable of scene decomposition, driving actions, three dimensional

[9] Though there are parallels with cortical anatomy and physiology, on top of behaviour performances, the *deep learning* models use implausible mechanisms, for instance, back-propagation, to learn the features.



	eye data quality requirement	interpretability	temporal information	data-driven	generalisability to other stimuli
a)	high	easy	no	yes	easy
b)	not important	medium	no	yes	easy
c)	medium	hard	yes	no	hard
d)	not important	medium	yes	yes	medium

Figure 1.4. Eye movement signal processing.
Overview of eye movement modelling and comparison methods taken and adapted from Coutrot et al. (2018). **a)** is a list of eye movement events parameters, **b)** is a list of spatial distribution methods for gaze analysis—i.e., Kullback-Leibler divergence (KLD), correlation coefficient (CC), Similarity (SIM), earth moving distance (EMD)—, **c)** is a list of string-based and geometric approaches for signal comparisons and **d)** regroups probabilistic approaches to eye movement analysis. The table at the bottom gives a qualitative appreciation and requirements for each cluster of method listed above.
For more details on all these methods, please refer to Coutrot et al. (2018).

environment mental representations, etc (Nilsson and Pelger, 1994; Lewicki et al., 2014).

> **Action is an important component to consider for vision (Aloimonos et al., 1988); so much so that the sensors, the eyes, never stop being active in information retrieval (Rucci and Casile, 2005; Kagan and Hafed, 2013), and action is driven by goals, behavioural states and memory.**

In classical feed-forward models, the information collapses towards a decision, but here one can consider an intermediate level in which goals can intervene in an **inferential process** (Knill and Richards, 1996; Kersten and Yuille, 2003; Kersten, Mamassian, et al., 2004; Shams and Beierholm, 2010; Moreno-Bote, Knill, et al., 2011; Moreno-Bote and Drugowitsch, 2015). The properties of the environment are not captured by the sensors; shape, motion, texture, colour, identity, and the many other features that humans systematically use to describe and interact with objects are *inferred*. These properties are entangled in spatio-temporal visual patterns.

Inference as a core mechanism of perception is key to understand how the visual experience, for instance, suppress the visual interference of the blood vessels in the eye in front of the retina, the hole made by the optical nerve in the retina, or even the fact that internal representations are three dimensional while the retinal image is two dimensional. In fact, most known visual illusion emerge from exploiting the inferential system such that it has to solve a problem in a *metameric*[10] fashion or to give an experience that is incoherent with the physical world—i.e., such as in multi-stable perception—see Fig. 1.5 for a schematic representation of the illusory phenomena (Mamassian, 2006).

Visual illusions as stimuli for vision science

Visual illusions have been a source of fascination for thinkers all the way back to Aristotle (Eagleman, 2001). In psychology, the rise of *Gestalt* theory brought illusory stimuli to the research community as phenomena that could be used to study the mechanisms of the brain (Köhler, 1929; Wertheimer, 1938; Wuerger et al., 1996; Bach and Poloschek, 2006). Since then, visual illusions have been used as stimuli to study and gain empirical knowledge on contour (Anstis, 2013), surface filling-in (Pinna and Grossberg, 2005), adaptation (Anstis et al., 1998), motion

[10] Perceptual metameres are phenomena in which two distinct physical objects or properties are merged into one perceptual object. This is studied, for instance, in colour perception (Hardin, 1988; Wandell, 1995) or cross modal fusion (Hillis et al., 2002).

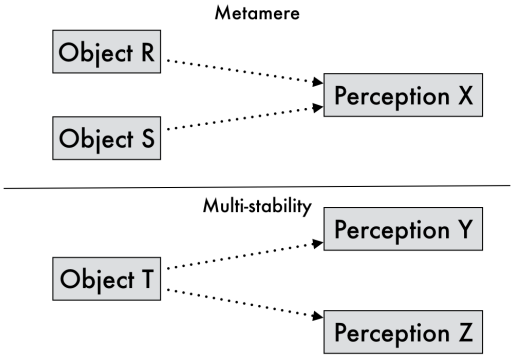


Figure 1.5. Metamere & Multi-stability. Diagram showing the difference between multi-stable perception and perception metamere. Reproduced from original diagram in Mamassian (2006).

perception (Anstis, 1980) and perceptual decision (D. Leopold and Logothetis, 1999), for instance[11]. Visual illusions support the theory that perception is an inferential process. The brain reconstructs and represents the environment internally so that actions and decision can be made.

Inferential problems are linked to incompleteness of information and past knowledge. In more formal terms, such as in the Bayesian framework, we refer to these as the probability of observing the data and prior information. In fact, a growing trend in cognitive sciences, using visual illusory phenomena, is oriented on **predictive coding**, an approach where vision is considered as an inference on visual information, other modalities and memory (Friston and Kiebel, 2009; Shams and Beierholm, 2010). To summarise, the brain is seen as a system that attempts to predict efficiently an interpretation of the neural sensory information. Visual illusions then correspond to situations, or conditions, in which the perceptual system mistakenly ignores differences in the physical world, for metameres. Or on the other hand, they may lead to a curious experience; the seemingly spontaneous change of a perceptual object into another one, while the physical stimulus remains unchanged, in the case of multi-stability.

Illusions like multi-stable perception or binocular rivalry have recently been used to gain insight on the neural correlates of visual awareness (Eagleman, 2001; Frässle et al., 2014). They are of interest as they feature a change of state of the perceptual system when the stimulation remains constant—see Fig. 1.5 for a schematic representation of the problem. Hence it is possible, with sufficient temporal resolution neuro-imaging methods—e.g., electro-encephalography (EEG),

[11] For more on visual illusion, the reader can delve into `https://michaelbach.de/ot/` or Martinez-Conde and Macknik (2017).

magneto-encephalography (MEG)—, to decode how perceptual change can occur in stable and controlled presentation condition (Parkkonen et al., 2008; Kornmeier and Bach, 2014). Alternatively, research has also focused on the location of percept coding in the brain by identifying the neural correlates during the duration of a quasi-stable percept. Using methods with high spatial resolution—e.g., functional magnetic resonance imaging (fMRI)—, researchers have identified multiple cortical areas, ranging from cortex areas V1 to middle temporal (MT) and depending on the stimulus used, in which the illusion changes seem to occur (Sterzer et al., 2009).

Figure 1.6. **Circle Limit IV (Heaven and Hell)**, by M.C. Escher. July 1960, Woodcut, printed from two blocks.

1.2 Vision & ambiguity: how does the brain handle it?

In this section, we delve more deeply in the behaviour of our visual system when facing **ambiguity**. Ambiguity is a universal term, used in a variety of context, and relates to the uncertainty associated with the interpretation of given information. This means that multiple interpretations are probable, and none of them clearly dominate the other ones. It differs from vagueness as the latter is evoked when having any interpretation is difficult given the information presented. Ambiguity can be used in reference to character's personality or motives in arts, such as the Caterpillar in Lewis Caroll's *Alice's adventures in Wonderland*[12] or Hamlet in Shakespeare's *Hamlet*[13]. Artists like Escher have exploited ambiguity in its visual manifestation to create fascinating visual works (for an example, see Fig. 1.6). In vision sciences, the classical example for ambiguity is the Necker cube (Necker, 1832), shown in Fig. 1.7. Even when the figure is hand drawn on paper, this simple stimulus clearly shows to an observer that its orientation's perception will change over time. This phenomenon has been defined as bi-stability, and it is the type of visual illusion of interest, in this thesis, from empirical and theoretical perspectives.

[12]The Caterpillar is ambiguous in his structure, as he is described to have a head that can be viewed as a human male's face or being a caterpillar's end head with legs.

[13]Hamlet is classic example of a protagonist showing moral ambiguity in literature. He has dual objectives as he tries to protect his mother and avenge his father. While Ophelia's death shows he has a human side with emotions, he carries on his vengeance leading to the death of many innocent characters.

One key phenomenon, within the domain of visual illusions, is **multi-stable perception**: when perception changes over the time of observation but the physical stimulus remains constant (D. Leopold and Logothetis, 1999). Multi-stable perception refers to the emergence of **multi-stability**, here, in visual perception. Multi-stability occurs when complex systems, with multiple sub-systems within, such as a brain, create multiple preferred states rather than one[14] (Kelso, 2012). This is not exclusive to the brain and can be observed in other domains such as physics (Hamedi et al., 2013), biology (Gonze et al., 2017), computational networks (Mao, 2012), climate science (Mitra et al., 2015), and more. However, the phenomenon has been investigated for over two centuries in vision sciences, since Necker (1832), a chemist, first reported the illusion formally (Fig. 1.7). In fact, there are different paradigms in experimental psychology to generate such illusions: **binocular rivalry**, in which two separate images are presented independently to each eye of the observer through an experimental setup, **ambiguous figures**, in which the stimulus is a static image (presented to both eyes simultaneously) too ambiguous for a single interpretation, or percept (Wernery, 2013), or **ambiguous videos**, where image motion will be ambiguous and generate alternation of percepts.

Multi-stable perception has been characterised by the following properties, common to many ambiguous stimuli:

- Irrepressible—an observer cannot avoid perceptual change over prolonged observation of a multi-stable stimulus.
- Mutual exclusivity—the percepts, i.e., the interpretations an observer will have of the stimulus, can only be experienced once at a time.
- Unpredictability—perceptual changes cannot be predicted (Lehky, 1995), at the time of writing, and seem to be stochastic, or at least are modelled as such.
- Percept duration distribution—tailed distribution such as Gamma and Log-Normal distribution are used to model reported percept durations (Levelt, 1967).
- First percept—the longest percept in duration, it has an idiosyncratic bias, and it may provide information on the continuous viewing empirical probabilities of perception (Hupé and Rubin, 2003; Mamassian and Goutcher, 2005).

Multi-stable perception properties.
The properties listed above are further described and expanded upon in Appendix A.4.

[14] In which case, the system would be said to be mono-stable.

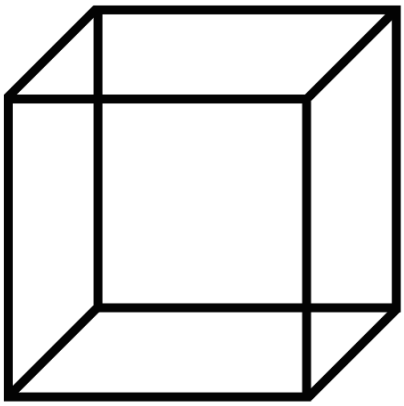


Figure 1.7. **Necker cube:** one of the simplest bi-stable stimuli, in which the geometric orientation of the cube is made ambiguous by the lack of perspective, i.e., vanishing lines, or the absence of occlusion of the back side of the cube. The ambiguity of the cube was first documented by Necker (1832).

In most studies, the simplest form of multi-stability is studied, namely bi-stability, in which there are variation between two quasi-stable states, here percepts.

Bi-stability: stimuli & phenomenon description

Visual bi-stability occurs in binocular rivalry and for most ambiguous figures—such as the classics, e.g., the Necker cube (Necker, 1832), Rubin's vase/face illusion (Parkkonen et al., 2008) or the rabbit/duck illusion (McManus et al., 2010), presented in Fig. 1.9. And bi-stability appears also in some artists' work, like for instance, Salvador Dali (Fig. 1.8) or Mauritis Cornelis Escher (Fig. 1.6), who exploited the visual phenomenon in various forms. In bi-stability, two states of perception, or percepts, are alternatively experienced by human observers. In binocular rivalry, where two different images are presented to each eye independently, Levelt (1966) presented four propositions to describe the dynamics:

(i) Stimulus dominance—the ratio of time duration of a percept over the other—depends on the strength of the stimulus; the strength can be modulated by controlling luminance and contrast.

(ii) Increasing the strength of the stimulus in one eye reduces the phase time of the stimulus in the other eye.

(iii) Reversal speed increases as a consequence of proposition 2.

Figure 1.8. Slave Market with the Disappearing Bust of Voltaire, by Salvador Dali, 1940. This art piece plays on spatial frequencies to hide two percepts and interpretations of the scene: one being a slave market scene, and with the right distance, one can see a bust of Voltaire appear and fade in the centre.

(iv) Increasing contrast of the images in both eyes reduces the time phases, and thus increases the reversal speed, however stimulus dominance remains unaffected. This was already observed by Breese (1899) and suggests independence between reversal and percept suppression in binocular rivalry.

These principles have been extended to some ambiguous figures in some studies (Mamassian and Goutcher, 2005; Chopin, 2012) however the second proposition is not always valid when contrast values go beyond a certain range (Jan W Brascamp et al., 2006). The hypothesis of independence between reversal and suppression suggests that two different mechanisms are at work which reflect such behaviour; one that maintains while another reverses.

1.2.1 Experimental paradigms of multi-stability

As mentioned above, multi-stable perception can be achieved, experimentally, in a variety of ways, and with various stimuli. We will now look at the main experimental paradigms in which the phenomenon has been studied, and their associated results.

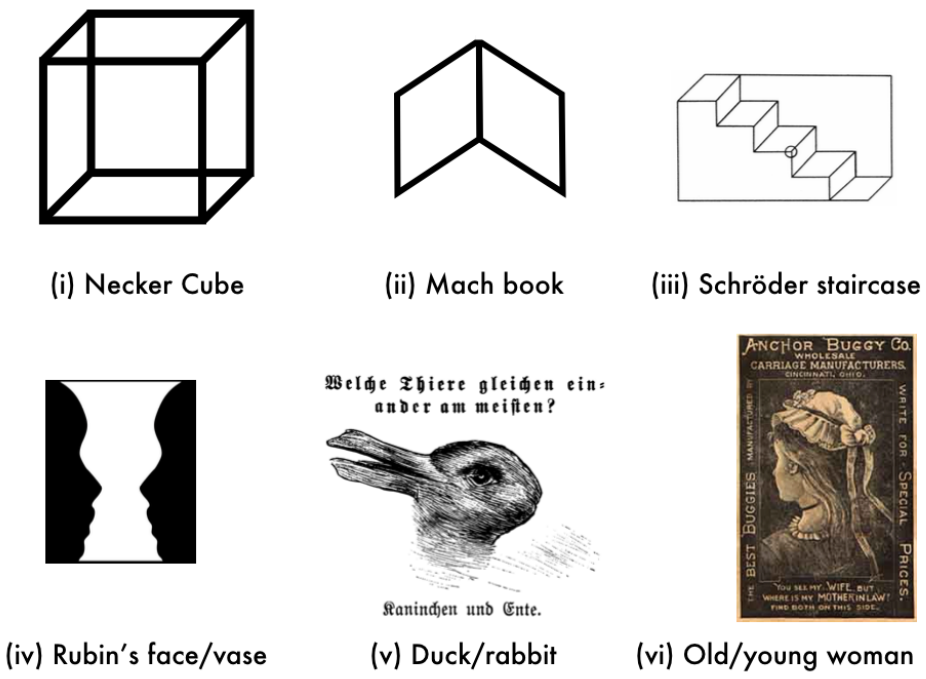


Figure 1.9. **Ambiguous figures.**
Top: different ambiguous figures of the categories perspective reversal: *(i)* the Necker Cube, *(ii)* the Mach book, and *(iii)* Schröder's staircase (taken from Wernery (2013)).
Below: an example of figure-ground bi-stability with *(iv)* Rubin's face/vase, and semantic rivalry with *(v)* the duck/rabbit illusion and the *(vi)* the old/young woman illusion (taken from Wernery (2013)).

Binocular rivalry

Researchers have developed sophisticated paradigms and experimental setups to study bi-stability in human perception. As such, binocular rivalry is an experimental paradigm that has been extensively used to investigate bi-stability.

Binocular rivalry consists in showing two different images or stimulus to each eye independently at the same retinal location—for a review we refer to Blake and Tong (2008). Fig. 1.10 shows different experimental setups used for binocular rivalry. The independence of stimulus presentation for each eye can be achieved, for instance, using a set of mirrors or polarised glasses. As a consequence of the setup, visual perception of the observers will alternate between the two images presented. First reports of binocular rivalry date back to as far as the 16th century (Wade, 1998). Though the setup might suggest that bi-stability is driven by specialised inter-ocular inhibitory processes rather than by competition of higher stimulus representations, there is compelling evidence that it is not the case and that binocular rivalry bi-stability might occur just in the visual cortex (D. Leopold and Logothetis, 1999). Note that the angular size of the stimulus can lead to different behavioural patterns. Indeed, the size of the stimulus will have an impact on what receptive fields and their associated neuronal networks code for the percept in the retinotopic visual cortex. Evidence suggests that the size of the minimal unit of such neuronal populations do not exceed 0.1 deg (Blake, O'Shea, et al., 1992).

Monocular rivalry

Monocular rivalry has a similar experimental setup to binocular rivalry but the same image is shown to both eyes; however, the stimulus can be a mixture, i.e., a superposition of two images, or a grating. In this instance, bi-stability will occur though after a longer fused initial interpretation. When comparing binocular and monocular rivalry, many similarities were reported, such as gamma distributed percept durations. Monocular rivalry is also called pattern rivalry and is considered closer to ambiguous figures in the range of bi-stable phenomena (Wernery, 2013).

Ambiguous figures

Ambiguous figures (Fig. 1.9) are images that do not contain sufficient information in order for the brain to stabilise itself on a single interpretation and which leads it to the experience of multi-stable alternations. Some of them have gained popularity

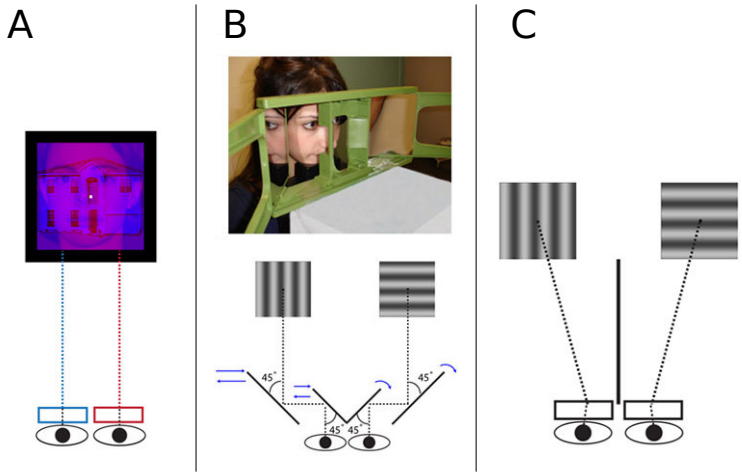


Figure 1.10. Binocular rivalry setups.
A: diagram of a setup using red-blue googles and with competing stimuli being composed only of red or blue colours.
B: picture and diagram of a setup using a miror stereoscope to project independent images on each eye.
C: Diagram of a setup using prism goggles, where each prism bends light creating an effective barrier between each eye.
Diagrams are taken from Carmel et al. (2010).

as they produce the phenomenon with no need for sophisticated viewing setup, and as most observers share and experience the illusion. Most ambiguous figures lead to bi-stability with two interpretations possible, though it is usually possible to consider an additional one: the meaningless drawing, a hidden third percept of bi-stability. For instance, the Necker cube can be perceived with two three-dimensional interpretations, however the third percept, in this case, refers to the "*flat*" two dimensional drawing of the cube. One should note that most studies do not record or take into account this third hidden percept, as its observation occurs after long exposure to the stimulus (Wernery, 2013).

Ambiguous figures can be distinguished in different categories due to the nature of the competition that operates and due to the computational properties involved in decoding the traits that determine the interpretations. However, in all cases, the physical stimulus remains unchanged while the observer's subjective experience alternates. Perspective reversing figures refers to images where a two dimensional drawing provides insufficient information for its three dimensional interpretation to be unique. Usual properties involved in this type of figure are symmetry and low semantic content, with both interpretations being very similar.

The most famous one is the Necker cube (Fig. 1.7) (Necker, 1832) and can also be extended to a lattice of Necker cubes (see Fig. 1.11, from Kornmeier and Bach

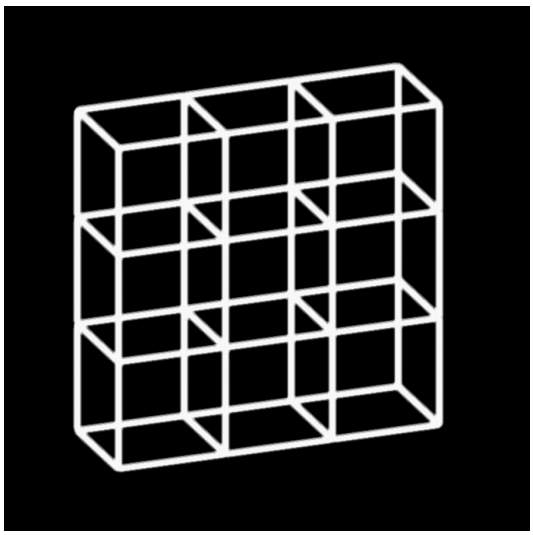


Figure 1.11. Necker lattice. Lattice of ambiguous bi-stable Necker cubes drawn, provided by J. Kornmeier.

(2012)). The Schröder staircase—Fig. 1.9-iii (Schröder, 1858)—is also an insightful example. Finally the Mach book (Fig. 1.9-ii) is yet another simple drawing that leads to perspective reversing bi-stability (Mach, 1901). Figure-ground reversing stimuli are related to Gestalt psychology, with their interpretations alternating between a foreground figure and a background shape standing out in an observer's consciousness, while the other is suppressed. The Rubin vase/faces (Fig. 1.9-iv) is also a popular example. Content reversal stimuli are characterised by the switch being due to the nature of the content observed in the subjective experience, not its perspective or figure-ground contrast. Examples such as the duck/rabbit and the old/young woman figures (Fig. 1.9-v,vi) were first published in non-scientific domains and were later adapted as stimuli for the study of perception.

Videos: dynamic bi-stable stimuli

Furthermore, bi-stable illusions can emerge in video stimuli; they are called structure-from-motion (Fang and He, 2004; Brouwer and Ee, 2006). They are related to visual kinetic depth effects (Wernery, 2013). The literature reports many results on the rotating sphere[15] and rotating cylinder stimuli—see Fig. 1.12. Apparent motion quartets[16] are also used to induce bi-stability, where the interpretations vary in the orientation of an inferred motion from dots blinks. Moving plaids can be used to produce a rivalry of direction too (Pressnitzer and Hupé, 2006).

[15] For an example, go to http://www.michaelbach.de/ot/col-equilu/index.html.
[16] For an example, go to http://www.michaelbach.de/ot/mot-sam/index.html.

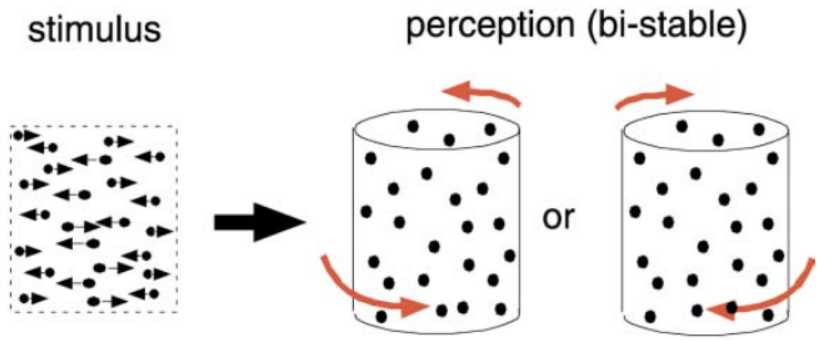


Figure 1.12. Structure from motion. Diagram showing the bi-stable illusion of structure from motion, that can be generated when randomly spaced dots within, a cylinder for instance, move according to, a horizontal sinus wave for example, and create the illusion of a rotating cylinder. Perception of the direction of rotation spontaneously changes for observers, creating the bi-stable perception phenomenon. Diagram taken from Fang and He (2004).

Finally, motion-induced blindness[17] has been shown to have temporal dynamics similar to ambiguous figure perception (Bonneh and Donner, 2011), though its link to bi-stability remains less obvious. Ambiguous figures also allow studying the interaction of specialised and distant neural networks known to code or operate decisions on certain types of information such as movement, face recognition, colour, perspective, and more. Thus, it is important to keep in mind that some ambiguous figures can lead to switches in perception within a category of cognitive function, i.e., intra-categorical competition, and across different cognitive functions, i.e., cross-categorical competition (Ishizu and Zeki, 2014).

Paradigm comparisons

Binocular rivalry and ambiguous figures are the two dominant paradigms used to study bi-stable perception; researchers have found many similarities and some differences in terms of observed dynamics and data fitting to models. One of the key aspects of bi-stability, the Gamma distribution of percept durations, has been reported, with similar results to binocular rivalry, for the Necker cube (Jan Brascamp et al., 2005), orientation rivalry (L. v. Dam and Ee, 2005), auditory bi-stability (Pressnitzer and Hupé, 2005; Pressnitzer and Hupé, 2006), monocular rivalry, motion-induced blindness (Wernery, 2013), or rivalry between moving gratings known as the moving plaid stimulus (Hupé and Rubin, 2003; Pressnitzer and Hupé, 2006; Moreno-Bote, Shpiro, et al., 2008; Moreno-Bote, Shpiro, et al.,

[17] For an example, go to http://www.michaelbach.de/ot/mot-mib/index.html.

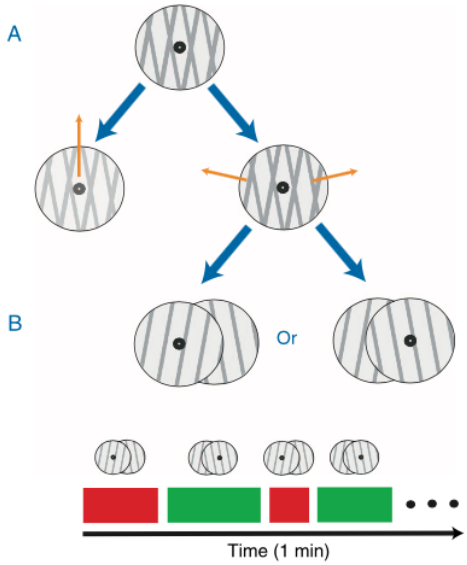


Figure 1.13. Moving plaid.
A: Diagram of the moving plaid stimulus and its two motion percepts with orange arrows.
B: A third percept can be introduced if observers are asked to differentiate the depth order of the two gratings for the transparency percept.
Below, the timeline of perceptual dynamics are shown, with perceptual changes occurring over time of observation of the stimulus.
Diagrams taken from Moreno-Bote, Shpiro, et al. (2008).

2010)—see Fig. 1.13 for a schematic explanation. The moving plaid stimulus is described in details, and a literature review of results related to it will be provided and discussed later, in Chapter 4. Another key property, the independence of phase durations and there unpredictability, has been reported for plaids (Rubin, Hupé, et al., 2005), auditory bi-stability (Pressnitzer and Hupé, 2006) and many ambiguous figures.

Moreover, differences in mean percept duration can be found for rivalry, orientation rivalry and moving plaids, where percept durations are found to be much longer than in binocular rivalry (Ee, Van Dam, et al., 2005; Wernery, 2013). Studies have found that voluntary control and observer strategy are more effective on the Necker cube as opposed to rivalry (Ee, Van Dam, et al., 2005; Meng and Tong, 2004). Finally, mutual exclusivity of interpretations is a property that relates to ambiguous figures more accurately than rivalry as observers report having fused percepts in the latter. It is worth noting that trans-magnetic stimulation (TMS) of the visual cortex has been shown to affect binocular rivalry but not ambiguous figure bi-stability, strongly suggesting that the conflicts in perception do not occur for the

same neuronal populations in these paradigms, though the phenomena are closely linked (Pearson, Tadin, et al., 2007).

1.2.2 State of the art for subjective reports

Before covering the topic, it is important to consider that, in an experimental setup where perception changes, the reversal can be of two nature: (i) *endogenous* reversals refer to changes of perception caused by internal processes in the visual system, and the brain, while (ii) *exogenous* reversals refer to changes of perception caused by external modification of the stimulation, in the physical display of the experimental setup.

The simplest form of experimental paradigm[18] is based on presenting the stimulus and asking for the observer to explicitly report perception. This is usually done by using assigned keys on a keyboard that correspond to a percept. Participants can be asked to either report changes in perception by a brief key press, or to keep the key pressed as long as a percept is perceived. For both cases, motor programming has to be considered as part of the response, as it introduces variable latencies. This time may vary greatly from one participant to another, and from one trial to another, from one perceptual change to another. In fact this issue is amplified when high temporal resolution neuro-imaging techniques are applied to study multi-stability, as it becomes difficult to estimate the precise moment of endogenous reversal and have an onset for perception change (Kornmeier, Ehm, et al., 2007; Kornmeier and Bach, 2012). Overall, one can assume that the motor response may take between 200 ms and 600 ms, but as shown in Fig. 1.14, multiple internal motor processes are involved before measure acquisition, and each can add variability on the time stamp. As most experiment in the literature use key press reporting, these caveats apply to most results (Ballanger and Boulinguez, 2009).

Experimental paradigms review

Two approaches have been explored by researchers: (i) *continuous* viewing during which the stimulus remains stationary and participants report their perceptual dynamics as it changes, and (ii) *discontinuous* viewing during which the experiment provides temporal windows during which the observers may report their perception.

[18] We refer to experimental paradigms as families of protocols and approaches in methodologies that have a common structure and constraints, beyond parameter changes.

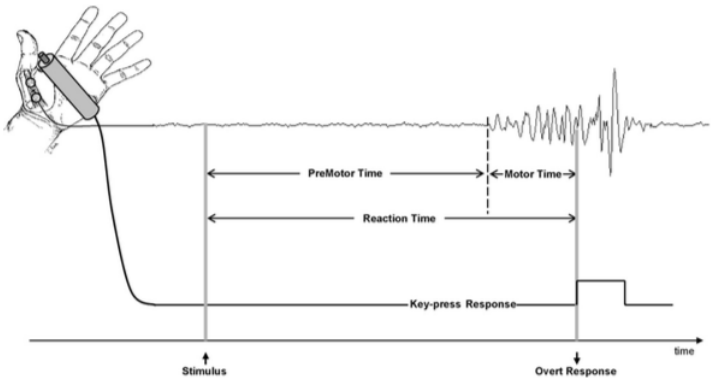


Figure 1.14. Key press. Diagram of the motor programming process for reaction time key press events, estimated from empirical data in electro-myography (EMG) in Ballanger and Boulinguez (2009).

Continuous viewing. Researchers have used the **continuous viewing paradigm** for empirical research in multi-stable perception because what sets the phenomenon apart from other visual illusions, is the mid-term[19] dynamics of perception. Indeed, most perceptual tasks in vision science use short trials to study the visual system, during which a stimulus will be presented for a duration smaller than 5 seconds, typically. This allows experimenters to control presentation condition and accumulate a large sample of trials per condition. However, multi-stable stimuli have varying mean percept durations with most being over 3 seconds: in fact this is why the percept is considered to be *stable* for a period of time. Note that static stimuli such as the Necker cube tend to have shorter percept durations (Zhou et al., 2004; Wernery, 2013) than dynamic stimuli like the moving plaid (Hupé and Rubin, 2003; Hupé and Rubin, 2004; Moreno-Bote, Shpiro, et al., 2008; Moreno-Bote, Shpiro, et al., 2010).

As a consequence, trial durations tend to be larger, though they vary greatly from one experiment to another, ranging between 20 seconds and 3 minutes. Trial duration quickly becomes a pragmatic trade-off to consider when designing experiments as observers will experience fatigue over prolonged stationary stimulation and reactivity to perceptual changes may vary more. For neurosciences, continuous viewing protocol have been used successfully in fMRI studies with results showing percept coding in the visual cortex and some other cortical areas (Sterzer et al., 2009), but also present difficulties to estimate the precise time at which the perceptual reversal occurs and to synchronise signals for event relate potential (ERP) and time-frequency (TF) analyses in MEG and EEG studies (Parkkonen et al., 2008; Kornmeier and Bach, 2012).

[19] By mid-term we mean effects that are in the order of magnitude of a tens of seconds.

Discrete viewing. To solve synchronisation issues raised in the previous paragraphs, some researchers have tried to exploit a known aspect of bi-stability: when visual presentation is interrupted by a mask also referred to as the inter-stimulus-interval (ISI), percept reversal rates[20] can be manipulated such that perception is quasi-stabilised for one percept (D. Leopold, Wilke, et al., 2002; Kornmeier, Ehm, et al., 2007).

We refer to this approach as the **discrete viewing paradigm**, also known as discontinuous presentation protocols. Defenders of this approach argue that given the flow inputted to the visual system is ecologically interrupted by eye blinks, the system is used to treat discontinuities and that multi-stable perception is fundamentally discontinuous (Kornmeier and Bach, 2012). Furthermore, ERP and TF analyses are greatly improved as temporal noise, i.e., phase shift, is reduced since the protocol forces the observer to respond at a given time, after the stimulus and the ISI were displayed (see Fig. 1.15) and an onset can be estimated (Parkkonen et al., 2008; Kornmeier and Bach, 2012). However, the issue cannot be as simply solved. Visual awareness is experienced as continuous but input feed interruption such as blinks are endogenously generated and do not influence the system with the same power as an exogenous ISI as shown by the variation of mean phase duration depending on ISI (D. Leopold, Wilke, et al., 2002). Indeed the impact of eye blinks on bi-stable perception has been shown to be minimal and indirect; in fact, L. v. Dam and Ee (2005) discuss that blinking rates are impacted and decrease when key press motor programming occurs. Such issues reorient questions on the impact of key press on perception. Their results support the idea that given blinks are endogenous events, the visual system is given the information of their occurrence and adapts to the task events.

In fact, Brascamp et al. (2009) showed that perceptual changes occur on larger time scales when using a discrete presentation protocol with ISI. They further argued that perceptual memory may play a role in the dynamics of multi-stable perception, by having similar dynamics at larger temporal scales. Furthermore, when looking at the dynamics in a probabilistic way, such as the one proposed by Mamassian and Goutcher (2005), the impact of blanks on survival probabilities[21] is evident, as the temporal dynamics suggest that subjects have high perceptual biases on their first percept, with the system having mechanisms to stabilise near equi-probability.

[20] Reversal rates or perceptual change speed is a measure used to estimate based on data in a trial how quickly perceptual changes occur. It is typically expressed as reversals per seconds and gives an interpretation and value on the dynamics of perception. One can compute it simply with: $r = n_X/t_T$ with r the reversal rate, n_X the number of perceptual changes and t_T the trial duration in seconds.

[21] Survival probability refers to a probability value given to the dominant percept that indicates the chance that it will remain dominant in the next iteration.

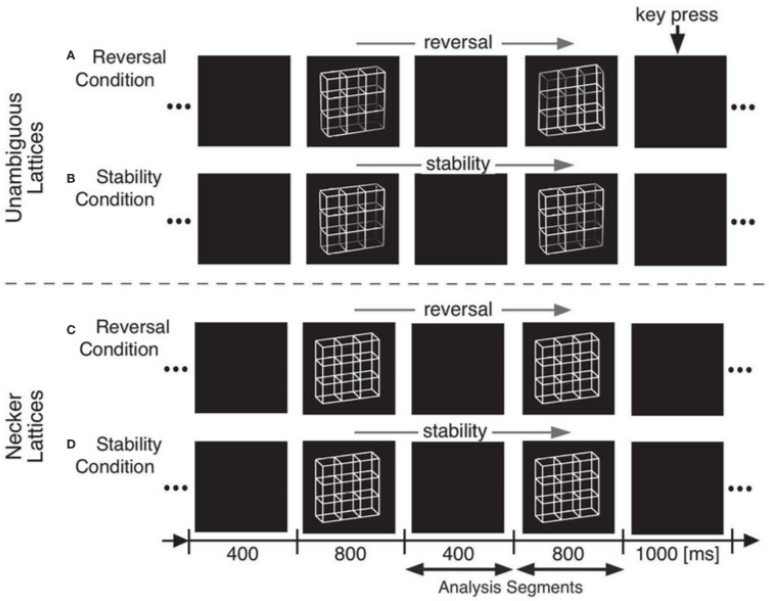


Figure 1.15. Discrete viewing paradigm. Experimental protocol from Kornmeier and Bach (2012) showing unambiguous (A & B) versus ambiguous (C & D) conditions contrast. Each condition shows examples where a perceptual reversal (A & C) is reported in the ISI, and perceptual stability (B & D) is reported during the ISI.

Known interference effects linked to key press reports

Motor response is used in subjective report experiments in the form of mostly key press on a computer keyboard, but also mouse click or even oral report. For instance, it is established that motor response performances are dependent on task difficulty or workload (Veltman and Gaillard, 1998). Furthermore, motor responses are not equivalent; saccades can be generated with much faster latencies, in just 100 milliseconds for natural scene tasks, compared to around 300 milliseconds when using button press (Crouzet et al., 2010; Guyader et al., 2017). The motor response implies attentional shifts and thus can affect the visual decoding and spatial frequency appearances of the stimulus for the observer (Yeshurun and Carrasco, 1998; Barbot and Carrasco, 2017; Barbot, S. Liu, et al., 2018). In fact, the effects of **attention** on multi-stable perception are not negligible.

In binocular rivalry, changes in attention can explain the dynamics of bi-stability (Dieter et al., 2016; Li et al., 2017), but this is not as clear for ambiguous figures. However, a review of motor control and learning with regards to attention exposed the conclusion that as participants get accustomed to the task, their attention is not distributed in the same manner as at the end of the experience (Song, 2019). This in turn affects perceptual performance for the visual system. Furthermore,

motor response is known to introduce unwanted neural activity in neuro-imaging experiment as programming the response activates neural networks dedicated and not involved in visual perception as per say (Kornmeier and Bach, 2012).

> **Overall, little is known on the direct contribution of key press responses on the dynamics of multi-stable perception besides its impact on attention and eye movements (L. v. Dam and Ee, 2005). To further study the contribution of indirectly related motor action, researchers need to develop, before hand, methods that can be enable perception decoding without relying on key press, so that the conditions can be contrasted.**

1.2.3 State of the art for objective reports

Objective report experimental protocols aim to use other methods to measure perceptual changes, without being entirely dependent on the observers' explicit responses.

Difficulties related to subjective report paradigms have been presented above, but other approaches are being developed. They rely on finding physiological markers or signatures of multi-stable perception that can be measured on participants during the task. These markers can be estimated by building up on subjective report protocols, during which data can be labelled by observers, considering them as *ground truth*—note that given some of the arguments mentioned above on motor and attentional interference, this notion should not be considered as absolute. As these signals may have some level of variability, this type of approach heavily relies on signal processing and modelling. In fact, this type of problems can be approached by machine learning methods, in which the signatures can be learned by an algorithm, to detect the markers of a percept or of perceptual changes. A model of signatures needs to be established either by theoretical works or by data-driven techniques, so that it can be applied to scan data and detect matches. The data can be of many kinds; for instance, one can use physiological data such as muscle activity measurement with EMG, or neural activity with EEG, MEG, electro-physiology, or blood activity with fMRI, or even eye movements using eye trackers. In the following paragraphs, we review attempts across the literature to identify perceptual dynamics in visual multi-stability, with an emphasis on eye movement studies.

Eye movements

Studying the dynamics of eye behaviour provides a mean to understand how forces external to the visual system may influence perceptual reversals. Or in other words, this approach can provide a characterisation of system noise coming from eye movements, fixations and blinks that may lead to switching. Necker himself had already described the influence of fixating certain aspects of the cube (Necker, 1832). Thus, the question of whether the percept, at a given time in bi-stable perception, is due to eye movement or whether the movement is a consequence of perceptual reversal has been an important source of investigation (L. C. v. Dam and Ee, 2006a). Eye tracking devices allow researchers to have quantitative and precise measurements, providing information on the impact of different events that are featured in human vision: fixations, saccades, smooth pursuit, blinks, pupil dilatation and micro-movements.

Observers of the multi-stable stimuli, and more specifically with the Necker cube, often report finding strategies to control their perception by fixating different features of the cube. Indeed, up to a certain extent, subjects can control the rate of perceptual reversals, though they can never fully stop the reversals (Ee, Van Dam, et al., 2005). Some studies have aimed to pinpoint the role of fixations and gaze position in ambiguous vision; for instance, it has been shown that eye position and percept dynamics of the Necker cube in a free viewing paradigm are closely linked (Einhäuser, Martin, et al., 2004). However this finding has been contested (Long and Toppino, 2004; L. v. Dam and Ee, 2005). As most scientists are interested in the markers that correlate or are causally linked to the reversal of perception, a series of studies have looked at eye movements and their implication in multi-stability.

Saccades and micro-saccades. Studying the impact of saccades, and other eye movements, on bi-stable perception is crucial in understanding the impact of shifting the visual input, along different neural populations, on the subsequent perceptual dynamics. In an attempt to quantitatively characterise eye movement related strategies operated by subjects in free viewing perceptual rivalry, no positive correlation was reported between saccades and perceptual changes (L. C. v. Dam and Ee, 2006a). Thus, it seems unlikely that participants use eye movements to stabilise or accelerate reversal rhythms. Furthermore, some interesting differences were highlighted: for instance, for ambiguous figures, no or weak correlations were observed between saccadic movements and reversals, though strong correlations were found in binocular rivalry (L. C. v. Dam and Ee, 2006a). And correlation between micro-saccadic eye movements and the following percept was shown. Hence, researchers have

attempted to differentiate the impact of eye movement from retinal image shifts on perception.

Retinal image shifts. In this perspective, retinal—and more specifically foveal—image change has been shown to be the factor linked to perception reversals in binocular rivalry by investigating the impact of stimulus motion on a saccade-less observation (Blake, Sobel, et al., 2003; L. C. v. Dam and Ee, 2006a). Using stimulus motion to excite changing populations of neurons, based on retinotopic mapping, and coding of the percept, in V1, it was expected that, in theory, the effects of neural adaptation[22], would be minimised. Blake and colleagues showed that stimulus motion speed, in subjects' visual field, had an effect on the average reversal speed for binocular rivalry. However, their data were collected on very few subjects, making generalisation questionable given the high inter-individual variability observed in multi-stability.

Pursuits & opto-kinetic nystagmus There have been attempts to use OKN on binocular rivalry to infer perceptual dynamics without depending on subjective reports (Frässle et al., 2014). The authors used a typical OKN generating stimulus and presented it separately to the two eyes and asked participants to report using key presses in one condition, and **no-report** in the other. In the latter, they analysed the gaze signal dynamics to infer the perceptual dynamics. This allowed them to then contrast neural correlates of multi-stable perception using fMRI. But their method was in fact improved by another team, using the same psychophysical setup, without the fMRI, but focusing on gaze signal processing and data interpretation to solve signal interruptions and displacement due to blinks and saccades (Aleshin et al., 2019). They proposed a method to estimate the **cumulative smooth pursuit** by identifying smooth pursuit epochs in the signal and interpolating the signal over unwanted other epochs. These are, to the author's knowledge, the first convincing attempts at applying objective measures of perceptual dynamics using eye movements, featuring experiments with **blind** no-report conditions. However, no such results have been reproduced yet on ambiguous figures.

Eye blinks. There has been little investigation of the impact of blinks on multi-stable dynamics though their role could theoretically be linked to studies of intermittent presentation of ambiguous stimuli. Indeed, blinks should act as blank periods where

[22] more details on adaptation will be given in the review of models, however, for now, adaptation refers to a process that drives a baseline oscillatory behaviour in bi-stability

the stimulus is not being fed to the visual system. Thus, results observed when manipulating and controlling discontinuous presentation times should offer insights on how blinks affect bi-stability. Researchers have shown that by setting the blank period, i.e., the ISI, to longer durations (above mean phase time), it is possible to stabilise the perception of an ambiguous stimulus, while short blanks will lead to high probabilities of reversals (D. Leopold, Wilke, et al., 2002). However, when examining data around blinks, in free viewing conditions, no positive correlation was found between blinks and reversal dynamics (L. v. Dam and Ee, 2005). This suggests that the blank created by a blink could be profoundly different from a stimulus removal, with perceptual memory[23] playing a part in the dynamics of multi-stability. Indeed, scientists have shown that using intermittent presentation can be misleading and lead to perceptual alternation cycle with specific characteristics and dependencies (Brascamp et al., 2009). When looking at the dynamics in a probabilistic way, such as the one proposed by Mamassian and Goutcher (2005), the impact of blanks on percept survival probabilities tends to create a bias towards the first percept.

Pupilometry. Pupil dynamics can be recorded and analysed through eye tracking data and can be used in vision sciences. Attempts to link pupilometry and perceptual changes have been proposed (Einhäuser, Stout, et al., 2008), however, the proposed correlations have been shown to reflect the planning of the motor response over the endogenous experience of reversal of perception (Hupé, Lamirel, et al., 2009). Moreover, pupilometry has been shown to be a less effective means to measure higher cognitive functions such as attention or work memory (Meghanathan et al., 2015).

Neuro-imaging results review.
A review of neuroscience results on multi-stable perception, which also provide objective report methods, can be read at Appendix A.5. It is not essential to understand the results presented in Chapters 2 to 4 as the planned EEG experiment could not be carried out.

To synthesise, finding oculomotor markers of perceptual changes, when the human visual system faces an ambiguous stimulus, is not trivial, as the many studies covered in the paragraphs above show. Although some empirical results show that oculomotor events—e.g., micro-saccades, OKN or retinal image shifts—can provide information on the perceptual dynamics, eye movements remain difficult to control, since they also depend on attentional and

[23] Perceptual memory refers to short term memory, present in the lower visual cortex, and used in the processes that reconstruct the visual experience (Magnussen and Greenlee, 1999).

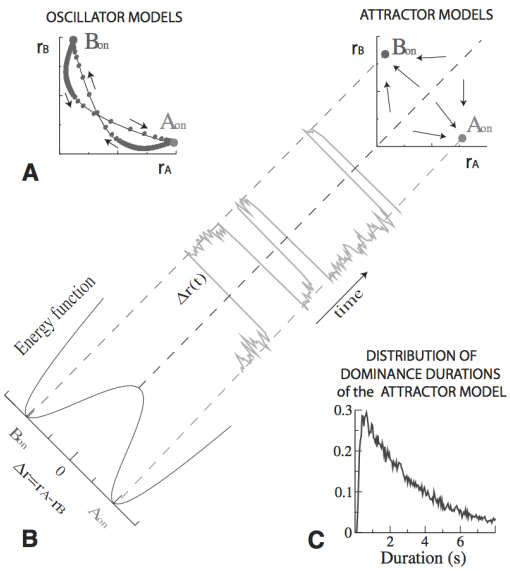
intentional factors. Finally, objective report methods require advanced signal processing to decode systematically signatures of percepts, or perceptual change, in oculomotor, physiological or neuroimaging data. Such constraints require researchers to consider modelling the phenomenon in order to predict the investigated and searched markers.

1.2.4 Explanatory models of multi-stable perception

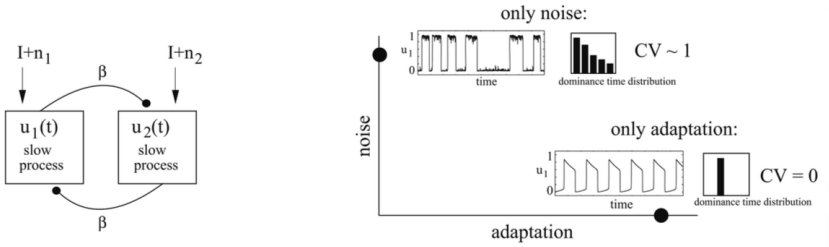
Historically, the scientific community modelling multi-stable perception was divided by two opposing approaches. On one hand, a top-down view, supported by the observations of voluntary control, knowledge of reversibility, priming and cognitive load (see Tab. A.1 in Appendix A.4), was developed. On the other hand, a bottom-up view, driven by the observations of initial adaptation, local adaptation, multiple-figure presentation, reverse-bias, (dis-)continuity of presentation and viewing parameters (see Tab. A.1 in Appendix A.4), emerged. This division was maintained as both sides found evidence backing their type of models (Wernery, 2013). Furthermore, bi-modal experimentation has shown that a centralised hypothesis seems less likely than a distributed competition in unconscious perceptual decision making (Hupé, Joffo, et al., 2008). Quantitative models have been proposed, by different research groups, to provide a theoretical and computational tools to understand multi-stable perception. They can be classified with two main approaches: oscillators and attractors.

Oscillators

Oscillator models are based on noisy oscillator circuits with adaptation being the main driving force dictating reversal temporal dynamics. A comparison of four oscillator models based on cross-inhibition for binocular rivalry showed that, with different parameters and gain functions, dynamics followed Levelt's *Proposition IV* of monotonic decrease of phase durations with the increase of the stimulus' strength (Shpiro, Curtu, et al., 2007). Oscillators have been used to account for **adaptation**, an internal theoretical force thought to explain the choice of percepts, their duration and Levelt's Proposition IV (Moreno-Bote, Rinzel, et al., 2007). It applies a slow negative feedback that gives models an oscillatory characteristic; as the system enters percept *A* (P_A), adaptation slowly reduces the probability of maintaining P_A over its competitor, percept B (P_B). As shown in Fig. 1.16aA, the



(a) Diagrams comparing oscillator and attractor models.



(b) Bi-stable models.

Figure 1.16. Bi-stability.
(a) Diagrams comparing oscillator and attractor models.
A (*left*): oscillator models have deterministic trajectories between the percepts A and B in the firing rate space (r_A, r_B) for a neural population coding the bi-stable perception. This leads to periodic oscillations between A and B.
A (*Right*): an attractor models have two locations in the (r_A, r_b) space and when the system is initialised, it falls to the nearest lower energy point or attractor. No change will occur unless noise is added.
B: the energy function for percepts A and B in the attractor models, and an example of perceptual time series by computing the difference of firing rates $\delta r(t) = r_A - r_B$.
C: the distribution of dominance durations for attractor models with noise.
Figure taken from Moreno-Bote, Rinzel, et al. (2007).
(b) Bi-stable models.
Left: a diagram of the architecture of a bi-stable model with slow negative feedback β and noisy inputs $I + n$, with I the input and n the noise. The model has slow negative feedback in the form of adaptation $(u_1(t), u_2(t))$.
Right: the consequence on alternation dynamics when the strength of noise and adaptation vary, with two extreme example points where the alternative force is absent and the models is either noise-driven or adaptation-driven. This diagram shows how noise provides the characteristic tailed distribution of dominance durations, while adaptation impacts the mean of that distribution.
Taken from Shpiro, Moreno-Bote, et al. (2009).

dynamics of the system act as a deterministic[24] oscillator. The period in each percept becomes regular and the system's state changes has periodicity that can be estimated and defined (Fig. 1.16b).

Adaptation is interpreted in physiological and neuronal terms as due to synaptic depression and spike rate or frequency adaptation (Shpiro, Curtu, et al., 2007). When neurons are pushed in a response regime over a prolonged period[25], the neurons gradually decrease in excitation for a constant input and adapt. This is in fact understood and modelled at the level of membrane current and action potential propagation dynamics (Dayan et al., 2001). It is observed and pervasive in many neural systems of human cognition and relates to the change of the system's response over time although the input remains the same. For instance, other visual illusions, such as MAE, are linked to neuronal adaptation (Anstis et al., 1998).

In the context of perceptual multi-stability, it is thought that adaptation occurs over populations or networks of neurons coding for one percept over its alternative (Shpiro, Moreno-Bote, et al., 2009; Moreno-Bote, Knill, et al., 2011; Huguet et al., 2014). Hence, it acts as a suppression or inhibitive mechanism that provides oscillations in the perceptual time discourse (Hupé, Signorelli, et al., 2019). However, the dynamics of perception have been shown to be unpredictable, and therefore, adaptation and oscillation are not sufficient to explain the empirical data (Lehky, 1995; Shpiro, Curtu, et al., 2007), and some form of stochastic process should be considered.

Attractors

Attractor models (Fig. 1.16A) propose that noise acts as the main component that directs perceptual reversal in time and adaptation only modulates the process. With such an approach, the system is modelled as a particle in landscape following a random walk (Einstein, 1956) and gradient descent (Kelley, 1999). An attractor for each percept (P_A & P_B) is set in an energy landscape. Noise, which remains poorly defined and characterised, is the driving force that helps overcome energy barriers between attractors and lead to percept reversals (Moreno-Bote, Rinzel, et al., 2007). It may refer to a variety of *negligible* interactions with other systems, thought not

[24] Determinism refers to the absence of noise on the system in the parameter space described.
[25] Depending on the neurons, it may vary from a few seconds to a few minutes to adapt.

to be of crucial importance for perception, or it may also refer to neuronal noise as modelled in leaky integrate-and-fire neurons (Moreno-Bote and Parga, 2005).

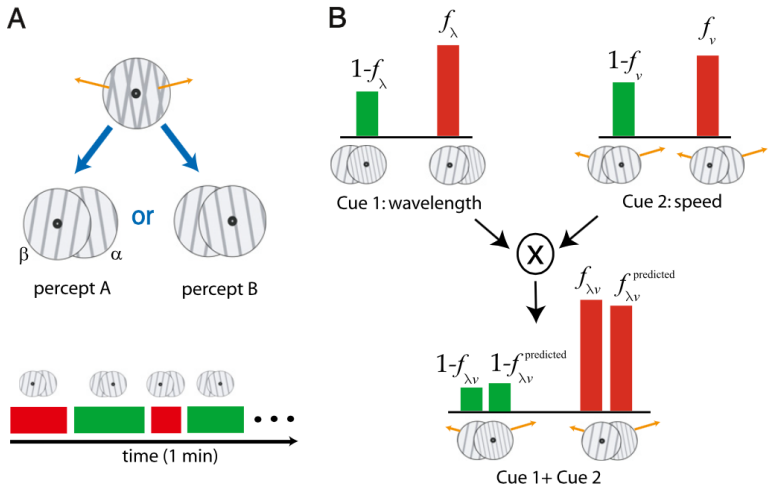
Consequently, models based on adaptation will showcase stable periodicity if the noise is removed, whereas noise-driven models will have no reversals of perceptual states (Shpiro, Moreno-Bote, et al., 2009). Moreover, if adaptation is removed from attractor models, the distribution of phase times would become Exponential, not Log-Normal or Gamma. Hence, a combination of these structures seems to be necessary as an exclusively attractor-based or oscillator-based model is not realistic given the arguments cited and the defining features of bi-stability (for an illustration, see Fig. 1.16 & Fig. 1.16b). For instance, an attractor model, with weak adaptation, has been implemented, and studied, with firing rate mean-field and in spiking cell-based neural networks (Moreno-Bote, Rinzel, et al., 2007). In Moreno-Bote, Knill, et al. (2011), the authors showed that when viewing ambiguous gratings motion with a moving plaid bi-stable stimulus, the fractions of dominance[26] of each percept depending on a cue manipulation (speed or wavelength) follow a multiplicative rule (see Fig. 1.17a). In fact, they showed that this is a key aspect of Bayesian sampling, thus suggesting that the visual system may act as such a sampler. They further showed that an attractor neural network can sample probability distributions in a Bayesian way (Fig. 1.17b), hence reinforcing the proposed idea.

Forces

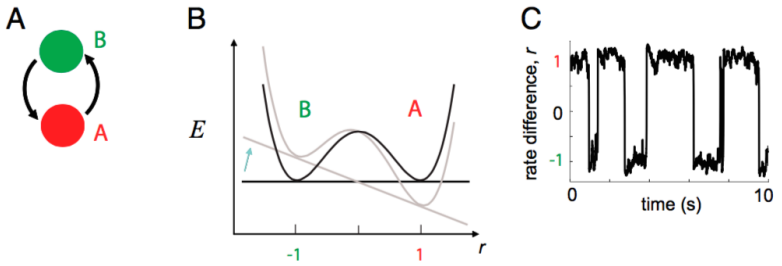
The models described above operate through the dynamics of a particle in an energy landscape, corresponding to its parametric space. Different forces are applied to the system and result in dynamical changes of its position in its parametric space. These forces are interpreted to be related to the input strength, adaptation and noise.

Input strength. Because multi-stability focuses on having the input kept constant and observing the states of the system vary, input strength is rarely manipulated, and often omitted from models. However, some models can be expressed with an input variable such as the one shown in Fig. 1.16b. Input strength relates to Levelt's propositions (Levelt, 1966; Levelt, 1967) for binocular rivalry, where the variable can be easily controlled by increasing luminance or contrast in the display of one eye over the other, for instance. In most multi-stable perception stimuli, such as ambiguous figures, we consider the input to remain constant throughout the

[26] The fraction of dominance here corresponds to the empirical probability of a percept P_A over total observation time.



(a) Bi-stable moving plaid.



(b) Bi-stable model.

Figure 1.17. Bi-stability and the moving plaid stimulus.
(a) Bi-stable moving plaid.
A: a diagram of the bi-stable moving plaid stimulus with transparency, or depth ordering, percept competition and a schematic timeline of perceptual dynamics.
B: a diagram showing how the multiplicative rule is applied, schematically, to fractions of dominance with different cues to build the probability distribution of each percept given the cues.
(b) Bi-stable model.
A: a diagram of competing neural networks coding for a percept each.
B: the model's energy function with attractors for each percept as a function of the firing rate difference of the two networks ($r = r_A - r_B$) and an adaptation force (the arrow).
C: the rate difference over time, with noisy oscillations, showing residence in each percept.
Figures taken from Moreno-Bote, Knill, et al. (2011)

observation time. Input strength can be seen to directly affect the probability of observing a percept, and in attractor models, it corresponds to the minimum level of energy for an attractor.

Adaptation. Adaptation represents a force that affects the relative depth of the attractors by providing a slow negative feedback. When the system particle is located in an attractor, its depth is reduced, thus increasing the probability of the competing attractor—illustrated by the grey arrow in Fig. 1.17bB, Moreno-Bote, Knill, et al. (2011). If adaptation is the only force applied to the system, one obtains a periodical dynamic of perception as shown by Fig. 1.16bB (Shpiro, Moreno-Bote, et al., 2009). Adaptation has a variety of interpretations in the literature (Huguet et al., 2014), but it corresponds, in theory, to a slow negative feedback loop. It is sometimes explained by referring to neural adaptation[27], a phenomenon that occurs when neurons are constantly stimulated over a period, their firing rate threshold adapts and shifts. Adaptation can lead, for instance, to visual illusions such as motion after effect (Anstis et al., 1998). However, adaptation can be also explained by mutual inhibition mechanisms, in which the activation of a population of neurons coding for a percept leads to its competitor gradually inhibiting the current percept (Hupé, Signorelli, et al., 2019). This leads to the same periodic observations, if only adaptation is the driving force in the system. Whatever the interpretation and biological plausibility, adaptation, in models, refers to a deterministic force or mechanism that provides oscillatory behaviours. It is then mixed with a stochastic component to obtain dynamics similar to bi-stable perception.

Noise. Neural noise is present in biophysical systems and corresponds to the intrinsic electrical fluctuations in the neural signals that do not code the information processed by a neural network (Huguet et al., 2014). However, more generally, noise can refer to stochastic processes that impact a system. In multi-stability modelling, it is an essential component to explain unpredictable percept durations, however, experimental observations or characterisations are scarce. In the attractor model family, noise has been shown to be necessary to reproduce the percept duration distributions observed in multi-stable perception (Shpiro, Curtu, et al., 2007; Shpiro, Moreno-Bote, et al., 2009). In fact, it is necessary to have both adaptation and noise's strength balanced to reproduce empirical observations, as both forces seem to be involved in alternations such that the system must operate near the boundary between being driven by adaptation or noise.

[27] Spike-frequency adaptation or synaptic depression.

Bi-stable models thus use adaptation to change the likelihood of a percept being chosen, by increasing the probability of the competing alternative percept. However, this is not sufficient: noise will provide the necessary energy to pass over the energy barrier remaining. Hence noise drives the moment of alternation, and adaptation which state is chosen.

If the noise enables the exploration of the multiple states, **stochastic resonance** is said to occur (Gammaitoni et al., 1998). Stochastic resonance is a phenomenon known to occur in biological systems in which a system takes advantage of its internal noise to enhance its performances (McDonnell and Abbott, 2009). Some works have shown how stochastic resonance might take place in the visual system (Kim et al., 2006; Funke et al., 2007).

1.3 Why do we study multi-stable perception?

This section provides some arguments motivating the study of multi-stable perception stimuli. Indeed, as reviewed in previous sections of this chapter, multi-stability is an illusion that occurs when visual inference cannot find a single stable interpretation to perceive. Cognitive systems, functions and research themes linked to this phenomenon include attention, decision making and complex system dynamics. Relevant results from these research fields are reported for a contextual understanding of this thesis' work.

1.3.1 Premotor theory of attention

Attention is a cognitive process that allows an individual to select an object or a feature and to focus on its processing over the rest of the stimulation. For instance, it allows to enhance our perceptual capacities on a part of the visual stimulation (Yeshurun and Carrasco, 1998). As James et al. (1890) originally wrote:

> "*[Attention] is the taking possession by the mind, in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought. Focalisation, concentration, of consciousness are of its essence.*"

In vision sciences, **spatial attention** has been extensively studied, and numerous experimental paradigms[28] have been developed and investigated, giving insights and observations (Rizzolatti, Riggio, Dascola, et al., 1987; Posner and Petersen, 1990; Posner and Dehaene, 1994; Rizzolatti, Craighero, et al., 1998; Petersen and Posner, 2012). In fact, vision and attention have been studied closely together, and attention has become a necessary consideration for theoretical work (Gide, Karam, et al., 2017), in a wide range of model families:

- **bottom-up** approaches including feature integration (Itti et al., 1998), spectral residual analysis (Hou and L. Zhang, 2007), superpixel segmentation (Z. Liu et al., 2013) or proto-objects (Yanulevskaya et al., 2013), and more[29],
- **deep learning** approaches including deep neural networks, sparse deep learning and Boltzmann machines (Gide, Karam, et al., 2017),

[28] e.g., Posner, Stroop, etc.
[29] For a complete and recent review reading Gide, Karam, et al. (2017) is recommended.

- and **top-down** approaches using the known features of higher visual processing mechanisms such as facial detection (Cerf et al., 2008; Gide, Karam, et al., 2017).

Classical theory of attention relies on a *supramodal* control mechanism that orients the information processing of the visual scene in the brain to enhance performance (Posner and Dehaene, 1994). This suggests that a dedicated neural network, anatomically and functionally, orients attention; this cognitivist conceptualisation has often linked attention and consciousness studies (Wyart and Tallon-Baudry, 2008). But connexionist approaches have shown that there is no need for a dedicated system and that it may be a distributed phenomenon (Rizzolatti, Riggio, Dascola, et al., 1987). The latter is known as the **premotor theory of attention** and explains the control mechanisms of attention as being dependant on weaker activation in a series of fronto-parietal networks (Desimone and Duncan, 1995; Rizzolatti, Craighero, et al., 1998). The study of attention has shown links to oculomotor dynamics as these networks are highly correlated and dependant to both attention and eye movement programming (Posner, 1980; Hoffman and Subramaniam, 1995; Kuhn et al., 2009; Engbert, Trukenbrod, et al., 2015; Kalogeropoulou and Rolfs, 2017; Meyberg et al., 2017).

Evidence from psychology studies

Using the Posner experimental paradigm, Rizzolatti, Riggio, Boris M Sheliga, et al. (1994) showed the *meridian effect*: a delay in response time when participants have to respond in the contra-lateral side with respect to the hemifield where the cue is. Thus attention was located. The meridian effect is a strong argument in favour of premotor theory of attention as it can be explained by how eye movements are executed: when a cue indicates the position of the expected stimulus, the observer prepares a saccade towards this expected position. If the expected target does not appear in the cued location, the brain has to reprogram a saccade, which adds a delay in the reaction time. If the target appears at the location cued, the saccade is carried out more efficiently, as the premotor programming corresponds to an attentional boost in performance. This can be observed, especially in eye movements, with saccade deviations—see Fig. 1.18 (B. Sheliga et al., 1995). In fact, this has been causally demonstrated by restraining eye movements, attention was affected such that detection performances dropped hence showing that oculomotor programming and attention are intrinsically linked (Craighero, Carta, et al., 2001; Craighero, Nascimben, et al., 2004).

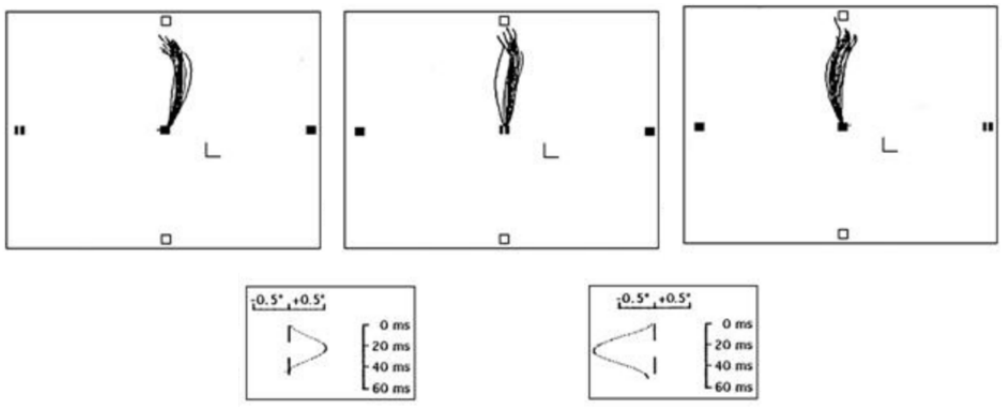


Figure 1.18. Spatial attention. Gaze traces from the experiment presented in B. Sheliga et al. (1995), in which the meridian effect in spatial attention is exposed by saccade deviation generated by cues in the contra-lateral side of target location.

Evidence from neuro-imaging & neuro-physiology studies

Neuro-imaging studies further showed that the cortical neuronal networks used for eye movement and visuo-spatial attention are shared in parietal, frontal and temporal lobes (Corbetta et al., 1998; Nobre et al., 2000). The networks reported in these studies are analogous, for humans, to the monkeys' FEF and LIP areas, both known for voluntary control of eye movements (Astrand et al., 2015). Moreover, event-related fMRI data on blind individuals reinforce the idea that attention and eye movements share the same neural circuitry in FEF (Garg et al., 2007).

Invasive neuro-physiology data also show activity of neurons recorded in monkey's SC, a core neural network for oculomotor programming (Kustov and David Robinson, 1996). Neurons in SC increased their excitability when attention was paid to the location it needed to make eye movements to, for the task. In fact, some researchers have shown that it is possible to enhance spatial perception by changing the oculomotor signals inside the brain (Moore and Fallah, 2001). Two monkeys had to make a manual response when detecting a transient dimming of a peripheral visual target while experimenters micro-stimulated the FEF cortical area. The authors reported that sub-threshold stimulation of a specific area of FEF led to a decrease for the psychophysical detection threshold of the stimulus, hence improving performance, when the target was positioned in the motor field corresponding to the stimulated neurons. Such evidence argues for a causal relationship between eye movement control and allocation of spatial attention. These results have since been replicated and investigated further, showing that when stimulating sub-regions of the FEF, visual and oculomotor performances can be improved (Moore and Armstrong,

2003; Ekstrom et al., 2008). Furthermore, these results have been causally replicated on human subjects using TMS-fMRI by showing stimulating human FEF leads to systematic effects on fMRI signals in the early visual cortex V1 (Ruff et al., 2006).

Attention & intention for action & perception

Premotor theory of attention makes attention intrinsic to motor commands and thus, action. Indeed, more specifically, eye movements, attention and visual perception are highly inter-linked phenomena (Posner, 1980; Hoffman and Subramaniam, 1995; Kuhn et al., 2009; Engbert, Trukenbrod, et al., 2015; Kalogeropoulou and Rolfs, 2017; Meyberg et al., 2017) as the first changes the input, the second orients the former and the latter processes the information. In fact, attention and intention are closely linked during visual tasks (Kohler et al., 2008), notably longer ones during which attention may vary and fluctuate over time (Esterman and Rothlein, 2019). On the other hand, intention can be controlled with relatively more constancy by means of task design and instructions (Firestone and Scholl, 2016). While attention often refers to bottom-up driven changes in behaviour (e.g., eye movements) and thus in perception, intention is often conceived as a top-down signal driving motor commands and affecting sensory inputs. Premotor theory offers a distributed mechanism in neural networks whereby the attention is not driven by an external supra-modal function, but by preactivation or sub-threshold levels of neural excitation in the action or perception related networks. Therefore, in this theoretical approach, attention is placed as an intrinsic component modulating both action and perception, but also provide a commonality and is shared across both systems.

This consequence is a key aspect that shall be exploited in this thesis' theoretical work, and provides a pathway to bridge our understanding of eye movements and multi-stable visual perception (Rolfs, 2015; Li et al., 2017; Mirza et al., 2019; Song, 2019)—but see Parr and Friston (2019) for a critic of the premotor theoretical approach.

1.3.2 Perceptual decision over time

When the visual system infers a perceptual representation of the sensory inputs, it takes many unconscious decisions, referred to as *perceptual decisions*.

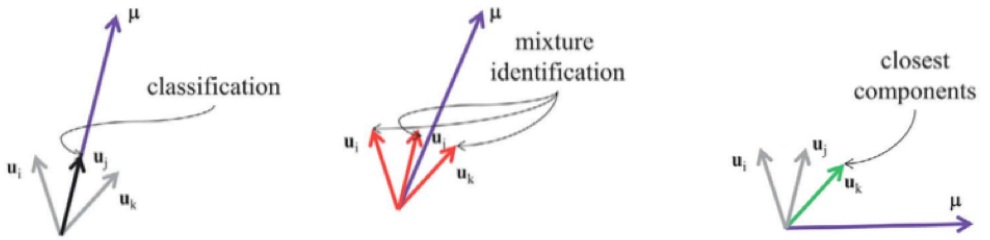


Figure 1.19. Decision models. Figure taken from Moreno-Bote and Drugowitsch (2015) showing the results of simulations for three types of decision problems: classification, mixture identification and closest component.

Percepts as options for perceptual decisions

Perception reduces the sensory information into interpreted perceptual objects in a mental space, as presented in Section 1.1.3. The process uses inferential mechanisms to generate percepts, and this requires the brain to make perceptual decisions. Decision making is a phenomenon that occurs in many cognitive functions and can be linked to neural network architecture reducing the input into a restrained set of outputs (Hérault and Jutten, 1994). In fact, in a recent computational study, Moreno-Bote and Drugowitsch (2015) showed that spiking network models can solve high dimensional causal inference problems. Using spiking network models in numerical simulations, they looked at the network's behaviour for hard discrimination classical problems, complex mixture identification problems and closest component problems (Fig. 1.19). This type of operation follows the accumulation of evidence from sensory input which is associated to a cost function (Drugowitsch et al., 2012). In the latter study, the authors used a RDK on a diffusion model and observers and obtains similar reaction time distributions and could identify a contribution they interpreted as the urgency signal which was independent of stimulus strength (i.e., difficulty). These studies suggest that perceptual decision operates at fast time scales with neural populations accumulating evidence towards an inference (i.e., a percept), and are able to take a hard decision to classify, but also to identify the proportions of the input leading to the outcome. When no class exist for a stimulation, the network can converge towards the closest state. Empirical studies using continuous tracking methods such as track pads over decision making provide data showing similar drift to those engaged by the models (Zeljko et al., 2019).

Inference mechanisms

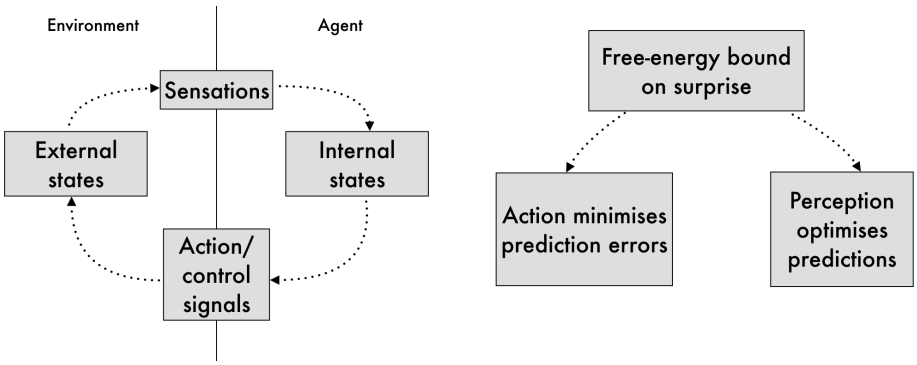
The states of decision making show parallelism to the attractors in multi-stable models and the Bayesian framework is used in both cases to explain the system's dynamics (Moreno-Bote, Knill, et al., 2011). Moreover, some researchers have shown that these attractors, in the network's parametric space, are not static and may evolve over time, with neuronal plasticity being a key feature to enable that (Malhotra et al., 2017). By combining an ideal observer model and a noisy judgement task in which observers had to trade off accuracy (i.e., the accumulation of evidence) and speed (i.e., more rewards), the authors showed that participants can modulate their internal decision boundary, though they did so in a sub-optimal way. Therefore, key theoretical bridges exist between findings in perceptual decision making and multi-stable perception. In the latter, the processes for accumulation of evidence continues over time and the suppressed percept ends up re-emerging, thus suggesting that the visual system is constantly inferring perceptual objects.

1.3.3 Multi-stability as a regime of stability in complex systems dynamics

The theory of **predictive coding** (Fig. 1.20a) with the **free energy principle** (Fig. 1.20b) proposes that the brain computes sets of prediction errors based on established and learned priors (Friston and Kiebel, 2009; Friston, 2010). The theory is based on the assumption that the brain is Bayesian (Chater et al., 2006), namely its computations attempt to solve the following generic equation, corresponding to Bayes' theorem:

$$p(\boldsymbol{\Theta}|\boldsymbol{X}) = \frac{p(\boldsymbol{X}|\boldsymbol{\Theta})p(\boldsymbol{\Theta})}{p(\boldsymbol{X})} \tag{1.1}$$

where $\boldsymbol{X}$ is a set of data (i.e., sensory information) and $\boldsymbol{\Theta}$ is a set of parameters. $p(\boldsymbol{\Theta})$ is the *prior* which corresponds to the probability that the brain have such a state, independent of the sensory information. $p(\boldsymbol{X}|\boldsymbol{\Theta})$ is the conditional probability of observing the sensory inputs given the current state of the system, also referred to as the *likelihood*. $p(\boldsymbol{X})$ is the marginal probability that normalises the $p(\boldsymbol{\Theta}|\boldsymbol{X})$ *posterior* distribution. The latter corresponds to the probability of the brain being in a state, defined by $\boldsymbol{\Theta}$, given the sensory input. The posterior is the probability law that is investigated in cognitive studies as it reflects behaviour based on the mixing of prior and likelihood distributions. The inference computation in the context of active vision can be interpreted using the schematic models described in Fig. 1.20. A consequence of the Bayesian foundations is that multiple systems can



(a) Predictive coding framework. **(b)** Free energy principle.

Figure 1.20. Predictive coding & free energy.
(a) Diagram depicting the loop for perceptual inference, in which an agent interacts with its environment by capturing an estimation of the environment's state through sensations, that modify its internal states, causing the agent to act, which in turn, affects the external states.
(b) The free energy principle postulates that the system is driven by surprise, i.e., when expectations are not met, and can act on its actions to minimise prediction errors or on its perceptual computations to optimise predictions and inferences.
Diagrams are reproduced from Friston (2010).

be integrated in hierarchical models with forward propagation of inference errors being a bottom-up coupling of systems, and backwards propagation being top-down influences.

The theory also considers that as complex systems, the neural networks learn to form attractors which encode *what* is being perceived, while neural activity itself encodes *where* the attractors are located in the network's parametric space. When considering such types of organisations for complex systems, multi-stability can be seen as a regime of stability, in which a complex system may settle for a time (Kelso, 2012). However, other regimes such as mono-stability and meta-stability exist—for an illustration see Fig. 1.21—and a system's regime depends on its history (i.e., what is has learned). Visual multi-stability occurs on highly trained visual systems—for instance, there is evidence that uninformed children are not capable of experiencing perceptual reversal on ambiguous figure (Mitroff et al., 2006)—and one can assume that priors are well established for adult participants and attractors are unlikely to vary. Hence, ambiguous stimuli reveal how the visual system, a mature and complex one, may enter a regime of multi-stability where the brain re-evaluates its predictive errors on sensory information, and operates perceptual decisions based on accumulated evidence and noise. Furthermore, similarities between the error back-propagation algorithm used for predictive coding learning and synapse's Hebbian

plasticity have been shown (Whittington and Bogacz, 2017), suggesting it is a good model to study how the brain may learn preferred states (i.e., attractors) for a certain type of problems and their boundaries may evolve.

Studying the brain as a complex system, composed of a multitude of inter-connected networks, requires to look at how these components interact: i.e., what is the level of integration or segregation of sub-systems? Though the task is not trivial as billions of neurons are involved with a much larger number of connections, researchers have tried to propose metrics to quantify and estimate such complexity, based on the theory of information, by looking at mutual information between sub-systems (Tononi et al., 1994), for instance. This type of measure is difficult to apply to neuro-physiological data as they are in essence, largely incomplete, but it is useful when working with complex computational models. Indeed, this provides tools to investigate the relative coupling and synergies between sub-systems over computational visual processing for instance.

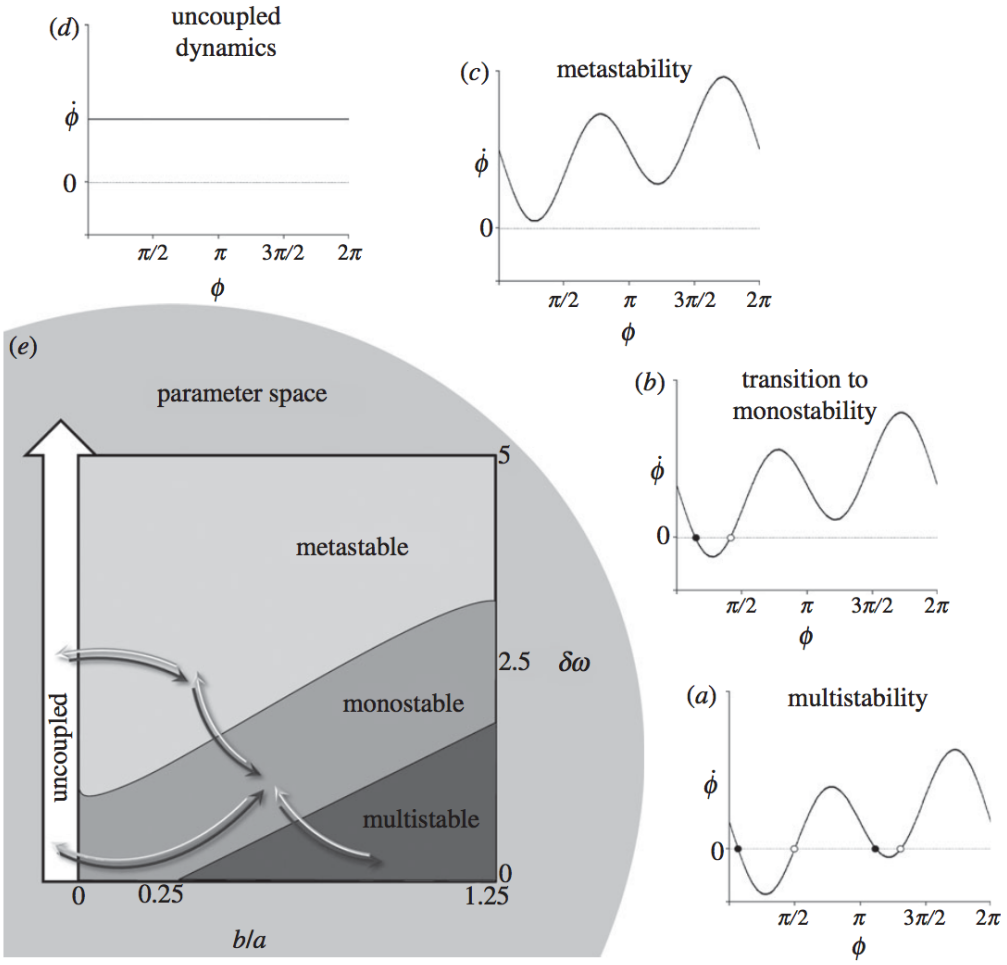


Figure 1.21. Mono-, multi-, meta-stability. Diagram taken from Kelso (2012) showing the changes of regimes as a dependency of a system's sub-systems' inner coupling. Coordination dynamics have four described regimes.
(a) Multi-stability where, here, two stable *attractors* are represented by the filled circles on the landscape of coordination variable coupling the systems oscillations together, the relative phase parameter ϕ, and two unstable *repellors* as unfilled circles.
(b) Mono-stability corresponds to a shift of the coordination variable ϕ such that only one attractor and one repellor remain at $\dot{\phi} = 0$.
(c) Meta-stability in which no attractor or repellor exist, with a system effectively using sub-system coordination, though no stable state can be found.
(d) Uncoupled dynamics where no sub-system coordination are present.
(e) The system's parameter space marked with areas corresponding to the described regimes and arrows showing possible transitions as the system's parameters and its coupling vary—b/a is a coupling parameter, internal to the system, while $\delta\omega$ represent the observed coordinated dynamics.

1.4 State of the art synthesis

In this section, a synthesis of the Chapter 1 is given by providing the gaps identified in the literature for visual multi-stability with eye movements and EEG studies. The content is split such that these gaps are specified for theoretical and computational models on one hand, and empirical studies on the other. Following this summary, the aspects that are addressed in this thesis are specified, thus providing the motivation for the rest of this manuscript.

1.4.1 Identified gaps

Theoretical and computational models

In theoretical works, the key gaps identified are (1) the lack of generalisation from bi-stable to multi-stable models, (2) the lack of consensus on the interpretation of the adaptation force in models, (3) the lack of studies on noise characterisation, (4) not many models implementing active vision processes for multi-stable phenomena and (5) the perspective of addressing the problem as dynamical complex systems coupling for multi-stable perception.

Bi-stable models, but little on multi-stable models. As reviewed in Section 1.2.4, research on computational models of multi-stable perception has focused on bi-stable phenomena. Attractor models composed of a deterministic component, adaptation, and a stochastic one, noise, are able to generate perceptual dynamics similar to observations measured on human participants in experiments. This is particularly true for replication of percept durations distribution. However, the focus has been primarily on the simpler case of a system with two attractors, or states. Few studies have approached problems with three or more states, though some attempts have been reported (Huguet et al., 2014).

> **In fact, generalisation to tri-stable perception provides new problems and may lead to further insights on the mechanistic properties of perceptual inference: for instance, understanding whether percepts interact hierarchically can be valuable to identify the temporal and serial aspects of inferential processes in the brain.**

Furthermore, bi-stability models rely on adaptation and noise, but does tri-stability require new components? Or does it clarify the controversial interpretation for adaptation?

Adaptation's interpretation is still debated. Adaptation is a force in theoretical models of bi-stable perception that has generated disagreements within the literature. It is defined as a deterministic slow negative feedback that allows oscillatory changes of perception if all other forces are removed (Shpiro, Moreno-Bote, et al., 2009). However, its interpretation, namely its physiological basis, is not clear. Some authors refer to neural adaptation (Pastukhov and J. Braun, 2013), often experienced in visual phenomena—e.g., in motion after effect (Anstis et al., 1998), or in colour perception—but is often explained by low-level bottom-up processes, where neurons adapt to a saturated state in order to provide sensitivity again after at least 20 seconds. This mechanism can be implemented by synaptic depression or spike-frequency adaption. An alternative view is that the slow negative feedback occurs in the brain by means of mutual inhibition connections between networks coding for the competing percepts (Hupé, Signorelli, et al., 2019). Further investigation on the physiological nature of this deterministic force in bi-stable models is thus still needed.

Noise's interpretation is elusive. Multi-stable perception dynamics have been shown to be replicated best when adding a stochastic process, in theoretical works (Shpiro, Moreno-Bote, et al., 2009). In fact, there is some evidence that bi-stable visual perception for binocular rivalry is not chaotic, and that given one perceptual duration, it is not possible to predict systematically the length of the next duration (Lehky, 1995). Furthermore, noise is an elusive notion that can be referred to for all processes that are not explained. Hence, identifying what might act as noise in the inference of an ambiguous stimulus for the visual system is a challenging and less addressed issue. But the role noise plays is not necessarily detrimental; for instance, stochastic resonance, in which the noise brings the energy for a change of the system's state (Gammaitoni et al., 1998), might provide an interesting explanation for multi-stable perception (Kim et al., 2006). Indeed, this could bring evidence towards showing that the brain has evolved to have optimal and preferred states to which it converges to, when doing perceptual inference, but also has a stochastic mechanism to allow it to explore new possibilities, thus improving its capacities to evolve and adapt (McDonnell and Abbott, 2009).

Addressing multi-stable perception as an active vision process. Some research has been carried out on trying to identify correlations between eye movements and visual multi-stability. However, from a modelling perspective, models have been focused on perception exclusively, and no model, to the author's knowledge, have provided a mechanistic interaction between the oculomotor system and the perceptual one (Fig. 1.22). In fact, investigating such coupling in the paradigm of active vision may provide key results on the identification of stochastic processes in perceptual inference. Results on hyper-acuity have shown for instance that fixational eye movements provide changes to the visual signal that are exploited by the brain to enhance its performances (Rucci and Victor, 2015). Thus, the noise, when decoding perceptual ambiguous stimuli, may be linked to variations operated in the motor system driving the changes in the visual flow: the oculomotor system.

> **The field may benefit from propositions giving computational models that can take into account gaze dynamics, that can be objectively measured, and perceptual subjective dynamics.**

Such propositions may be done in the form of a theoretical framework in which models can be developed in both spaces with their interactions; they would allow, if generative, the possibility of making quantified predictions, through computational simulations and the generation of synthetic data, that can be tested in empirical studies. Such perspectives would lead the community in theoretical studies of multi-stable perception to investigate the phenomenon by studying how these systems are coupled and their synergy.

Developing the study of multi-stable perception as complex system coupling problem. Visual multi-stability is linked to the inferential process that occurs in decoding visual information. As reviewed in Section 1.1 and Appendix A.2, many layers of processing are involved. Information is gradually interpreted into perceptual objects through parallel operations. This implies that many sub-systems are involved and that the signals are combined and re-combined asynchronously, though perceptual experience is known to flow continuously. A recent article has proposed and sketched out a framework for considering how all these sub-systems involved might be treated as a complex larger system and may be studied as such (Kelso, 2012). It focuses in considering the coupling of these sub-systems by observing whether they are in a synergistic regime or whether they tend towards segregation. Moreover, these considerations should be done with dynamical processes.

> **In the framework proposed by Kelso (2012), multi-stability becomes a regime of stability for complex dynamic systems, and there-**

fore, a larger context is provided to understand how multi-stable perception might be a marker of an evolving visual system's coupling states.

This might help formulate hypotheses on the longer temporal dynamics of visual multi-stability, and may help understand why it seems to be absent in children of young age and some patients with neuropathologies.

Eye movements & multi-stable perception

For empirical studies, key identified gaps related to this thesis are as follows; (1) some oculomotor markers of multi-stable perception have been identified, (2) high temporal resolution neuro-imaging techniques combined with eye-tracking could provide further knowledge on the processes occurring at the moment of reversal, (3) many issues remain with continuous viewing paradigms, more so with neuro-imaging techniques and (4) further development of no-report protocols are needed to remove subjective and attentional shift biases.

Oculomotor markers of multi-stable perception. Though studies have been carried out in attempts to find eye movement correlates of bi-stable perception, these have often been done with multiple and varied stimuli, and more importantly, with an empirical and exploratory approach. In other words, most studies investigated whether saccadic dynamics or fixation location might be correlated with perceptual events, but clear evidence of such relationships are sparse.

An alternative approach can be formulated as follows: given a chosen stimulus and associated models for eye movements and perception, one can establish specific hypotheses and predict gaze markers linked to perceptual changes and the stimulus dynamics.

This type of approach have been attempted successfully in recent works using binocular rivalry and OKN eye movements (Frässle et al., 2014). The markers allowed the authors to investigate neural correlates of perceptual changes and with recent methodological improvements, this type of research can be further investigated (Aleshin et al., 2019).

Joint gaze and neuro-imaging investigations. Eye movement and neuro-imaging research on visual multi-stable perception have produced many results, independently. Moreover, methods combining both recordings have sprung in the past two decades, enabling researchers to look at interaction between physiological and neural dynamics. In addition to oculomotor markers of perceptual change in multi-stable perception, reviewed in this chapter, neural dynamics can be used to decode internal events (see Appendix A.5). This can be done by looking in TF analysis for Gamma band energy periods at percept reversals and Alpha band energy during percept duration (Kornmeier and Bach, 2012). MEG and EEG, especially, should provide adequate data to investigate the processes that occur around perceptual reversal.

Improve continuous viewing paradigms. Multi-stable perception is characterised by changes of perception over a continued period of observation. This differs from most cognitive experiments for which trials tend to be short and thus, tasks and attention tend be relatively controlled. A key challenge here, on the other hand, is that participants will observe the stimulus and report spontaneous perceptual changes over periods typically beyond 20 seconds, and will do so in a repetitive manner. This tends to introduce variability in participants' behaviours. Some researchers have reported methods that can provide a relative compensation for these fall outs, using discontinued stimulus presentation with blank intervals and the report task being deported to the end of the trial (Kornmeier, Ehm, et al., 2007), or by sampling the perception at periods over long observation times by giving an auditory cue to report (Mamassian and Goutcher, 2005).

> **Beyond attentional control, key press report requires motor commands and impacts attention and perception.**

Thus, no-report experimental protocols are a step towards improving continuous viewing protocols.

Extend no report methods for neural correlates of consciousness research. No-report protocols rely on inferring the content of conscious perception based on implicit physiological markers. Such markers may be eye movements or neural dynamics for instance. This approach provides a bridge with the emerging literature on neural correlates of consciousness (NCC). The results from NCC can provide insights on identifying neural markers for perceptual changes in multi-stability. The contrary is true as well, the development of methods on multi-stable stimuli can help drive research on NCC and provide a methodological basis to develop the field.

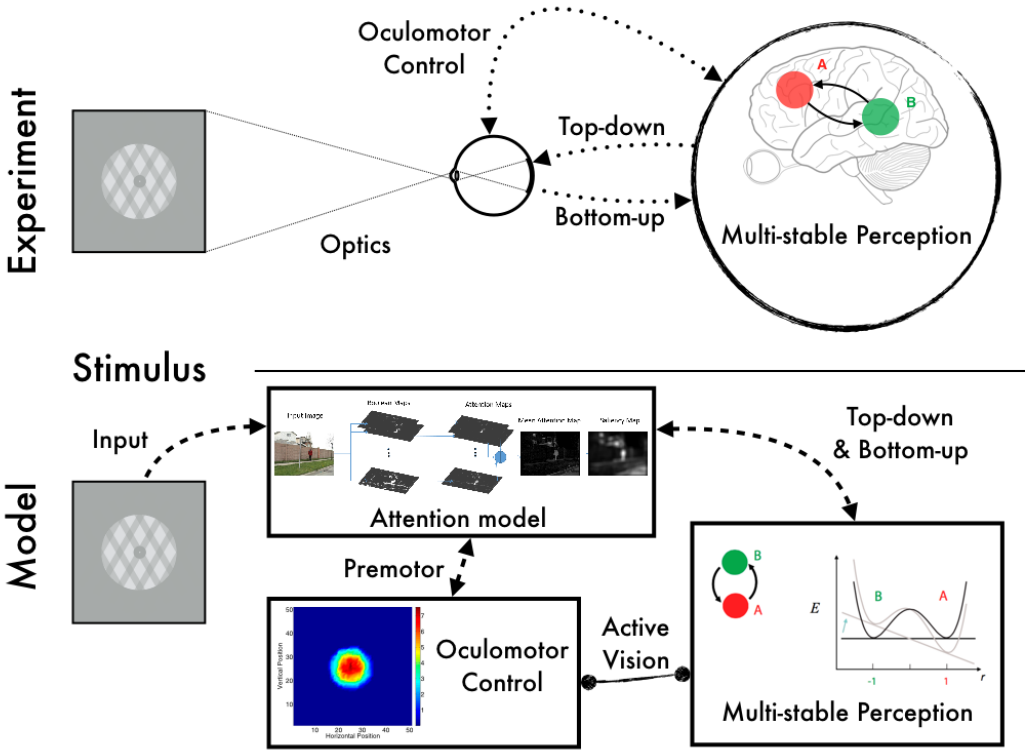


Figure 1.22. Synthesis & simplified overview of the literature review. This diagram provides a simplified overview of the state of the art for experimental and modelling research on the processes involved in visual multi-stability. In this thesis, we approach multi-stability with models and experiments. We also focus on the relationship with relationship between oculomotor control and perception, defined as *active vision*, i.e., vision being coupled to the body's action to operate. The stimulus chosen for most of the study is the *moving plaid*, featuring tri-stable ambiguity on perceived motion direction and depth ordering. Diagram composed of figures from the literature (Engbert, Mergenthaler, et al., 2011; Moreno-Bote, Knill, et al., 2011; J. Zhang and Sclaroff, 2013; Huguet et al., 2014).

1.4.2 Gaps addressed in this work

In this work, we do not address all the gaps but we focus on some by following an approached based on theoretical models developed. To address the problem more simply, first, linking the field of eye movements and multi-stable perception is necessary. In fact, investigating oculomotor dynamics by looking at eye movement classification, and providing a model, that can join both phenomena, is key to predict observation that relate perceptual multi-stability and oculomotricity.

In Chapter 2, we provide a definition and evidence of *micro-pursuits*, fixational eye movement dynamics showing similarity with a stimulus' motion. We look at how micro-pursuit may relate to other FEM, like micro-saccades, and propose metrics to measure them, based on spatio-temporal similarity of bi-variate signals.

In Chapter 3, a set of models based on *gravitational energy fields* provides a framework for a theoretical understanding of active vision in the context of multi-stable perception. It also explains all eye movement dynamics using the motion of attractors, which can be linked to perceptual attractors, modulated by attentional and intentional forces. Experimental work of this manuscript focuses on a particular tri-stable stimulus: the *moving plaid*.

In Chapter 4, contributions on the manipulation of its ambiguity using the gratings' transparency parameters, and the perceptual inference from gaze data based on its motion ambiguity. The latter is a test to apply theoretical and empirical understanding of the moving plaid and eye movements in order to investigate the role of eye movements in perceptual decision when stimulation is ambiguous.

In Chapter 5, a synthesis of the body of work presented in the thesis are presented, as well as some preliminary works on oculomotor control in the moving plaid and no-report paradigm, and perspectives are discussed.

2 Micro-pursuits: a class of fixational eye movements

> *Consciousness is only possible through change; change is only possible through movement.*
>
> — ***Aldous Huxley***
> **"The Art of Seeing".**

Human vision and eye movements are intrinsically linked as the latter change the visual input projected on the retina. Though our visual representation is stable, the eyes never truly stay still and generate small amplitude FEM that can be interesting markers of cognitive states. Research in the field of FEM has been extensive on micro-saccades, but less is known about drift and slow movements. Drift and slow movements tend to be considered as independent from visual stimulation, since larger eye movements are typically used to explore the visual field. However, we have detected small amplitude (fixational), slow movements when the task comprised a visual target with a highly predictable trajectory. In addition, the gaze showed high similarity with the target trajectory, measured through maximally projected correlation. Individual and group analyses gave significant results both in an implicit (Necker) and an explicit (Cross) pursuit task experiment, but not in a secondary implicit (Square) pursuit task experiment. The inter-experiment analysis results suggest that the manipulation of task, stimulus target motion, and the complexity of the stimulus may play a role in the generation of micro-pursuits.

Publication.
The work presented in this chapter had been accepted for publication to the *Journal of Vision*, as an article under the title *Micro-pursuit: a class of fixational eye movements correlating with smooth, predictable, small-scale target trajectories*, Kevin Parisot, Steeve Zozor, Anne Guérin-Dugué, Ronald Phlypo, & Alan Chauvin, and has been revised since this thesis' first review.
CAUTION: the micro-saccade analysis was corrected following the defence's discussion, in which an error in the analysis was pointed out by the examiners Laurent Madelain. The discussion was significantly modified as no secondary main sequence were detected in the corrected analysis. The *Journal of Vision* article has been accepted for publication and the reader should refer to the version in Appendix C rather than the following content, which was left as a trace of the defence's discussion.

Contents

2.1 Introduction

The main function of eye movements is to orient the gaze towards parts of a visual scene (Yarbus, 1967; S. Palmer, 1999; Liversedge et al., 2011). To accomplish this goal, the human oculomotor system has the capacity to generate a wide variety of movements that can be categorised based on their spatio-temporal dynamics: amplitude, velocity, and acceleration.

Rapid and ballistic eye movements (saccades): classified based on displacement, speed, and acceleration thresholds, e.g., displacement above 0.15 degrees (deg), velocity above 30 deg.s^{-1}, acceleration above 9500 deg.s^{-2}, though other detection criteria exist (Nyström and Holmqvist, 2010; Behrens et al., 2010; Mihali et al., 2017). These criteria have become their definition. But, absolute threshold criteria have been criticised for their lack of functional, physiological or formal justifications. For example: the clear dichotomy between fixations and saccades has been loosened (Ko, Poletti, et al., 2010).

Slow eye movements (smooth eye pursuits, slow oculomotor control): classified based on a simple velocity criterion, e.g. smooth pursuit ranges from 20 to 90 or 20 to 100 deg.s^{-1} (Krauzlis, 2004; Komogortsev and Karpov, 2013; Spering and Montagnini, 2011), though pursuits are considered smooth and precise only at speeds up to 30 deg.s^{-1}. If target velocity is too high for the pursuit system, catch-up saccades can compensate for the accumulated position error created by the difference between target and gaze velocities, also known as the retinal slip (De Brouwer et al., 2002).

Eye fixations: usually defined as any eye movement with an amplitude below 1 deg. They specifically include FEM which form a generic class of small-amplitude eye movements (ocular drift, tremor and micro-saccades) sharing dynamic characteristics with regular (macro) eye-movements at smaller scale (Otero-Millan et al., 2013; Krauzlis et al., 2017).

The aim of this chapter is to focus on FEM and more specifically the subclass of slow FEM, which we term micro-pursuit eye movements. We provide evidence of micro-pursuit eye movements, providing an adapted metric that reveals their existence in three different experiments. Thus, we will describe the different class of slow eye movements, with their functions and metrics.

Notation used.
Throughout the chapter, subscripts R, G, and S will respectively refer to the retinal image, the gaze, and the stimulus. We will use $\boldsymbol{q}_G$ and $\dot{\boldsymbol{q}}_G$ to define the gaze position (in deg) and velocity (in deg.s^{-1}), respectively (analogously for $\boldsymbol{q}_R$ and $\dot{\boldsymbol{q}}_R$ or $\boldsymbol{q}_S$ and $\dot{\boldsymbol{q}}_S$). The bold notation indicates that we deal with a 2D column vector of coordinate (and a 2×2 matrix when a capital letter is used), the over-lined notation $\bar{\cdot}$ refers to the mean over a set of trials and let the tilde notation $\tilde{\cdot}$ refers to the median over a set of trials, for all metrics. The $\pm$ sign precedes standard deviation values associated with mean values, while for median values we report mean absolute deviation (mad).

2.1.1 Slow eye movements: different kinds of motion

The functional role of (smooth) pursuit is to maintain a—usually moving—target of interest on the high acuity foveal region of the retina (Spering and Montagnini, 2011). Tracking is believed to be controlled by retinal errors, the difference between gaze and target positions, or retinal slip, i.e. $\boldsymbol{q}_R \doteq \boldsymbol{q}_G - \boldsymbol{q}_S$, the difference between gaze and target velocities or speed vectors of the gaze and of the target stimulus, i.e., $\dot{\boldsymbol{q}}_R \doteq \dot{\boldsymbol{q}}_G - \dot{\boldsymbol{q}}_S$. According to Orban de Xivry and colleagues (Orban de Xivry and Lefevre, 2007), pursuit relies mostly on reducing retinal slip and is modulated, in a smaller way, by position and acceleration errors.

In order to detect and measure the quality of slow eye movements, metrics have been defined that associate gaze with the target stimulus position. For smooth pursuit, tracking quality is measured through *gain* (see Section 2.2 section for more details). This measure has shown its effectiveness in experimental protocols where a target appears on screen and participants are tasked to follow its motion. Pursuit is mostly studied for tracking a single point on a uniform background, although other stimuli in motion also lead to pursuit movements, for instance, random-dot kinematograms (Heinen and Watamaniuk, 1998), line figures (G. Masson and Leland Stone, 2002), illusory perceptual motion (Madelain and Krauzlis, 2003), or after-effect motion (D. Braun et al., 2006). In tasks where a percept is pursued, rather than a stimulus, the measure of gain and the associated models have been questioned (Leland S Stone et al., 2000).

Among the slow eye movements, we also find reflexive movements such as the VOR, the OFR, or the OKN. The VOR is a reflex eye movement that compensates head motion in order to maintain a stable retinal image. Though the VOR expression may be similar to pursuit, it is only generated when the head is free to move. The OFR is a reflexive eye movement in response to a sudden change of a wide-field image (Michalski et al., 1977; Quaia et al., 2012). The reflex is mainly attributed

to the tracking of motion in peripheral vision (Ilg, 1997). The OKN is a composite gaze pattern in which an object is followed by smooth pursuit until the object leaves the visual field. At this point, the gaze returns to the object's initial position (fast saccadic response) at the starting position of the pursuit. VOR, OFR and OKN are eye movements solicited in specific visual stimulation and experimental contexts, which require the manipulation of a large part of the visual field, not a smaller perceptual target, as with pursuit.

To summarise, pursuits have been studied as large-scale eye movements with amplitudes exceeding 1 deg (60 min-arc) in which a percept with motion is tracked by the gaze, such that the retinal slip is minimised. The metric used to measure pursuit has been velocity gain since it has been shown that the oculomotor system uses motion information for movement control and closed-loop feedback models have been proposed to explain observed data.

2.1.2 Do small amplitude pursuits exist?

Fixational eye movements

We have just described the three principal classes of eye movements, where saccades and pursuits are distinguished from fixations based on the amplitudes and velocities involved. In fact, it is well known that during the fixation the eye never stands still (Ditchburn and Ginsborg, 1953) and continuously produces FEM subdividing *fixations* into the following sub-classes (Kowler, 2011): **micro-saccades**, composed of ballistic small amplitude and fast gaze shifts (Rolfs, 2009; Poletti and Rucci, 2016); **slow drifts**, small velocity (< 0.5 deg.s^{-1}) displacements of the gaze (Nachmias, 1961; Yarbus, 1967); and **tremors (or physiological nystagmus)**, aperiodic high-frequency oscillations of the eye (30-80 Hz and amplitudes of up to 50 sec of arc) (Nachmias, 1961; Martinez-Conde, Macknik, and D. Hubel, 2004). Some of these phenomena, like micro-saccades, have been studied extensively over the past decades—see Fig. 2 in (Rolfs, 2009)—and consensus has emerged on the functional and neurological similarities between large-scale saccades and micro-saccades (Ko, Poletti, et al., 2010; Sinn and Engbert, 2016). Research has also been conducted on tremor, but due to their small amplitude and high frequency it is impossible to distinguish them from noise using video-based eye-trackers (Ko, Snodderly, et al., 2016). Therefore, tremors will not be considered in our study. The class of slow drifts, and more particular small-amplitude pursuits, seems less covered in

the literature, which can be explained by the technical difficulties associated with eye-tracker precision, especially video-based ones, at such small scales (Wyatt, 2010; Choe et al., 2016). As we want to focus on the latter, we will give a detailed review of literature on slow drifts small-amplitude movements.

Ocular drift: a simple random process or stimulus-dependent?

These slow and small movements are the consequence of a slow control system of eye position (Cunitz, 1970) described in literature as a mere drift of the eye (Dodge, 1907), OFR (C.-Y. Chen and Hafed, 2013), or—more recently—as small amplitude pursuits (Skinner et al., 2018).

In early studies of FEM, when subjects had to fixate a static dot, eyes drifted slowly with an upper velocity limit at 0.5 deg.s^{-1} and mean velocity of 5 min-arc.s^{-1} (Yarbus, 1967). Their trajectories were considered as random and involuntary processes since they showed dynamics similar to Brownian random walks (Ratliff and Riggs, 1950; Engbert and Kliegl, 2004) as well as independence between the two eyes (Cornsweet, 1956). However, Ditchburn and Ginsborg's work (Ditchburn and Ginsborg, 1953) provided evidence that direction of eye movement is not completely random during drift; it is idiosyncratic. Nachmias (Nachmias, 1961) replicated this finding in an experiment where a fixation target was switched on and off during 3 seconds cycles. He found that each of the 2 subjects have preferred drifting direction but this preferred direction can be modified by changing the visual environment. The author interpreted the idiosyncratic direction preference as specific to muscular response and reasserted that nonrandom ocular drifts occur in fixations while providing evidence that drift direction can be modulated by the visual environment. More recently, a variety of experiments have shown that drift can take properties and characteristics close to other known oculomotor phenomena (Poletti, Listorti, et al., 2010; C.-Y. Chen and Hafed, 2013; Skinner et al., 2018; M. Watanabe et al., 2019).

As mentioned, drift can be viewed as part of a slow control system, enabling gaze to capture a target, whether static or dynamic. Here, we will discuss two studies that show evidence of slow eye movements correlating with the target stimulus, and as such related to our proposition of adding a subclass to the FEM: that of micro-pursuits.

Chen and Hafed (C.-Y. Chen and Hafed, 2013) studied the impact of micro-saccades on visual perception and investigated the relationship between micro-saccades and drift. Their experiment contained two major tasks. The first task required two

monkeys to stare at a fixation dot where a change in luminance of the dot or a peripheral white flash was introduced to induce a higher probability of micro-saccade generation. Drift velocity was analysed before and after the micro-saccades using either direct velocity measurements or spatial dispersion (by spatial binning and box counts). Both measures showed an increase in drift velocity post-micro-saccadic movements with respect to pre-micro-saccadic movements or baseline movements. They also showed that eye drift mainly occurs in the direction opposite to the micro-saccade, which is interpreted as corrective slow control of the gaze position. The second task consisted of a sinusoidal grating that started moving at predefined delays after the onset of a micro-saccade (or after 500 ms if no micro-saccade was detected). The authors analysed the speed and direction of early drift of the eye, namely the OFR, according to the direction of the grating and the time of grating onset based on micro-saccade detection. Indeed, they reported that (i) the drift directions were in the opposite directions of the micro-saccades and (ii) the eye velocity was reduced when the grating's motion was initiated during micro-saccade and was enhanced when the motion was initiated after micro-saccade. The OFR being an indicator of "*the sensitivity of early motion processing to retinal-image slip after a micro-saccade*", the OFR, and thus motion perception, are suppressed during the saccade and enhanced after. Their overall findings suggest that there is a single slow gaze control system that control both fixation and eye movement position in the presence of a fixed target or a slow moving background linked to the motion perception system. Conclusions suggesting a subtle coupling between micro-saccades and drifts are also reinforce by previous reports (Engbert and Mergenthaler, 2006).

Part of this idea had already put forward by Murphy and colleagues (Murphy et al., 1975). In their experiment, they asked participants to maintain their gaze on a present or absent fixation dot while a grating in the background moved horizontally at velocity ranging from 0.08 deg.s^{-1} to 8 deg.s^{-1}. In a second condition, the participants had to follow the grating. Eye movement velocities were analysed for trials without saccades. The study shows that when participants have to stare at the fixation dot (i) they have an ability to keep gaze fixed when the fixation dot was present, and (ii) an OFR -— a smooth displacement of the eye in the direction of the grating's movement but with smaller velocities—is detected when the fixation dot was absent. In contrast, when the task was to follow the grating, participants showed clear smooth, slow movement in the direction of motion with velocity as low as 0.08 deg.s^{-1}.

Both these studies confirm the existence of a slow movement within a fixation that track a slow velocity target or counteract the displacement of a micro-saccade. These

slow movement of pursuit or fixation stabilisation are thought to be under a same slow control system, although the tracking mechanism seems not to be triggered when the movement is initiated during a micro-saccade.

Ocular drift and slow motor control

Drift has been linked to slow control of the eyes during fixation in the context of investigating links between visual stimulation and drift motion.

In a series of experiments, Kowler and Steinman (Kowler and Steinman, 1979a; Kowler and Steinman, 1979b) have investigated how expectation, over a stimulus and task, can induce anticipatory smooth and slow eye movements. The authors implemented a task in which participants had to track a dot moving by steps (with three frequencies: 0.25, 0.375 or 0.5 Hz) along a horizontal segment of 3.3 deg amplitude. They showed that eye movements' direction and latency depend on predictability of target displacement. Furthermore, they showed this effect to remain even when the level of predictability was manipulated and when a distracting secondary task was imposed (Kowler and Steinman, 1981). In fact, they provided evidence that anticipatory eye movements—which they also named involuntary drifts in the direction of future target motion—depended on the history of prior target motions (Kowler, Martins, et al., 1984). To understand whether the slow control of ocular drift is driven by position or velocity signals, they carried out an experiment in which they manipulated drift by changing the configuration of reference points, thus varying the difficulty of fixation of a central point (Epelboim and Kowler, 1993). Their analyses used gaze position data and bi-variate contour ellipse area (BCEA) computation for quantification of gaze dispersion. As such, they provided evidence that the oculomotor system does not rely on visual position signals, but rather on retinal image slip, in order to implement slow motor control. This creates a parallel with the known models for smooth eye pursuit described above.

In addition, in a recent paper, Watanabe and colleagues (M. Watanabe et al., 2019) reported a study that links ocular drift, micro-saccades, and pupil area on voluntary eye movements preparation. They observed anticipatory drifts prior to stimulus appearance and they argue that these anticipatory eye movement may reflect volitional action preparation. Interestingly, the authors provide a replication of previous results on anticipatory drift with a video-based eye tracker while applying correction to their gaze signals for pupil deformation.

Overall, these studies show that slow eye movements are present during fixation. These movements can control for a fixation position, can track large target and depend on expectation. Authors have postulated that all these behaviours are under control of a unique system.

Small amplitude pursuits

As mentioned higher, smooth pursuits are large-scale eye movements with amplitudes exceeding 1 deg (60 min-arc). A small set of studies found eye movements within a fixation that share characteristics with smooth pursuits, except for their amplitude. Though there are references to smooth pursuits of small amplitude as far as in Yarbus' book (Yarbus, 1967), most papers in the literature have reported the phenomenon in an indirect manner.

In a study on drift in the absence of visual stimulation or with afterimages, horizontal smooth drifts were reported (Heywood and Churcher, 1971). Although their description corresponds to pursuit dynamics, they did not define the observed movements as such. The authors published a follow-up paper showing that, depending on the eccentricity of the afterimage, oculomotor dynamics are more or less smooth and show low velocities, hence could be interpreted as pursuits (Heywood and Churcher, 1972). Further, while attempting to study oculomotor control capacities when presenting a moving grating background with a fixation point, Murphy and colleagues (Murphy et al., 1975) reported eye movements that correspond to small amplitude pursuits. When investigating the lack of compensation of the VOR when the head was free, Martins, Kowler and Palmer (Martins et al., 1985) studied whether a smooth pursuit system might interact with the VOR. Their data provided a qualitative description that small amplitude pursuits are related to the velocity of target motion. The following finding was reported: foremost, the effectiveness of smooth pursuits varied with target velocities. At the lowest average velocities of a tracked point[1] (0.0025-0.125 deg.s^{-1}), smooth pursuit was the most effective, i.e., retinal-image speed during smooth pursuit was about the same as retinal-image speed during low target velocities. At higher target velocities (0.25-1 deg.s^{-1}), smooth pursuit was less effective for retinal image stabilisation and at the highest velocities (1.5-2.5 deg.s^{-1}), smooth pursuit was totally ineffective.

[1] Here, we present the velocities rather than frequencies to provide comparable measures across reviewed articles. However, in most cases, the target signal corresponds to a sinusoidal movement, thus velocity is not constant over a period.

More recently, small amplitude pursuits have been reported again, in very different contexts. In a study of eye drift and its relationship to retinal image motion—investigating whether the latter drives the former through retinal or extra-retinal information—Poletti and colleagues (Poletti, Listorti, et al., 2010) declared the following observation: "*small pursuit-like eye movement with amplitudes comparable to those of fixational drifts are under precise control of the oculomotor system*". Finally, a precise characterisation of rhesus macaque oculomotor control for rectilinear sinusoidal motion of a target with amplitudes inferior to 0.5 deg and velocities below 2.5 deg.s^{-1} was recently reported (Skinner et al., 2018). The amplitude and frequency of the sinusoidal motion was modulated and gaze signals were analysed using gain and compared to filter responses; filters are, here, used as models to show how the oculomotor system could display different behaviours based on input frequencies—on gaze position and velocity. Furthermore, they showed that the gaze signals had eye velocity spectrum with peaks at target frequency and that pursuit gain was highest at 1 deg.s^{-1}.

> **Overall, *pursuits* have been observed for a range of velocities (0.05-2 deg.s^{-1}) and amplitudes (1.9-30 min-of-arc) which qualifies them as FEM. Given the classification in the FEM research field—in which only micro-saccades, drifts, and tremors are considered—these observations raise questions on the nature and potential definition of micro-pursuits or fixational pursuits.**

This chapter focuses on the presentation of micro-pursuits in three contexts: (i) presentation of metrics that fit the theoretical requirements to detect micro-pursuit, (ii) detection of the oculomotor phenomenon in (a) a dual task experiment (Necker) in which its elicitation was not explicitly made to participants, and (b) an explicit tracking experiment (Cross) and an implicit distractor setup (Square). Our hypothesis was that if the perceptual system has to detect a change in a moving stimulus with a predictable trajectory, the oculomotor system is likely to follow the target even if the participant is instructed not to do so (fixation task). But, since the fixation task inhibits large deviations, only small amplitude pursuit eye movements are generated. Furthermore, a computational model of pursuit eye movements based on gravitational energy fields is presented in Chapter 3 that accounts for the two contrasting objectives (fixation vs. pursuit). In our data analyses, we made use of a measure of inertia for gaze dispersion and maximally projected correlation (MPC) for similarity, since they are simple methods that showcase clear advantages in our context. The latter also offers a metric that can be physically interpreted as it is able to capture similarity between two trajectories of different scales and spatial offsets.

2.2 Micro-pursuits

2.2.1 Quantifying pursuit movements (metrics)

To propose a definition of micro-pursuit movements, existing metrics for ocular movements will be discussed, since they will orient our choices for proposing metrics and hence our working definition.

Classical smooth pursuit is measured by velocity—or retinal slip—gain (gain $=$ $\|\dot{\boldsymbol{q}}_G\|/\|\dot{\boldsymbol{q}}_S\|$ with $\dot{\boldsymbol{q}}_G$ the gaze velocity and $\dot{\boldsymbol{q}}_S$ the stimulus velocity), which is consistent with its closed-loop modelling (Liversedge et al., 2011). Position gain is also used—although to a lesser extent,—for instance, when dealing with catch-up saccades (Orban de Xivry and Lefevre, 2007). For the various drift phenomena described in the previous section, a variety of metrics have been used to study FEM dynamics (e.g. gaze position, velocity, acceleration, gain, and BCEA). For instance, gain measurement was used for analysis in the case of the small amplitude pursuits of monkeys on uni-variate sinusoidal motion (Skinner et al., 2018). But the authors went further and provided a spectral analysis using Fourier transform on eye signals to identify the fundamental frequency and harmonics with the expected target frequencies. However, gain is a uni-variate metric which does not extend to multi-variate problems. Thus, it can be used adequately only for pursuit of a target moving on a line, rather than a plane, like the visual field. Fourier analysis shares the same issue as it looks for a frequency in a uni-variate movement, typically horizontal.

In studies of ocular drift (Epelboim and Kowler, 1993), BCEA[2] was used to quantify the spatial variance—inertia, or spread—of the gaze. The authors obtain orientation preferences through the inferred relative anisotropy of the ellipse. Though this metric is clearly conceived for bi-variate signals, it does not provide spatio-temporal correlation between gaze and a target signal in the way gain does. Meanwhile, the box-count method used in more recent studies permits to compute dispersion of the gaze data over time, though it may suffer, like gain, from measurement noise, especially with video-based eye tracker (Engbert and Mergenthaler, 2006; C.-Y. Chen and Hafed, 2013). To summarise, (i) some metrics, e.g. BCEA, box count, inertia, can be used as quantifiers for the spread of a bi-variate gaze signal during an epoch,

[2]The surface area of the ellipse such that the data belong to this area with a probability of 68% when a two dimensional Gaussian fits the data; roughly speaking, up to a factor, it is the determinant of the empirical covariance matrix.

and these metrics are useful descriptors for drift and slow movements, and (ii) other metrics, e.g. gain, Fourier analysis, correlation, can be used to quantify similarity between two bi-variate signals, to quantify the quality of a pursuit between gaze and a stimulus in motion. Each metric presents a trade-off that should be considered based on a theoretical definition and prediction.

2.2.2 Micro-pursuits: a working definition

Given the reported observations of small amplitude pursuits, the following constraints need to be considered to define a **micro-pursuit**.

Amplitude as indicated by the prefix of its name, and as an analogy to micro-saccades, the micro-pursuit must be of small amplitude, within the range of fixational eye movements; typically below 1 deg.

Velocity ; micro-pursuit should consist of slow zye movements, similarly to drift, or smooth pursuit but at a smaller scale, with velocities below 2 deg.s^{-1}.

Tracking ; micro-pursuits occur when a percept with motion across the observer's visual field is tracked. But, as pursuit involves matching the motion of a target by that of an observer in real time, micro-pursuit measurement of tracking should reflect the spatio-temporal interaction between the dynamics of two bi-variate signals. Hence, *similarity* between gaze dynamics should be evaluated. Because the eye movement amplitude is within the fovea's size, deformation may occur in the tracking of predictable bi-variate signals. Therefore, any similarity metric should exhibit both scale and translation invariances—spatial offset invariance may also be beneficial for measures from eye-trackers with lower precision and accuracy.

Duration ; the phenomenon of tracking a moving target requires by definition that it is done over a sufficiently long epoch. Thus, micro-pursuit should not occur over brief epochs such as saccades and micro-saccades.

We propose that gaze signal epochs satisfying the above description be considered as **micro-pursuits**. As this is a proposed working definition, micro-pursuits may correspond to entire eye fixation periods, making it possible for micro-pursuit to be punctuated by other FEM. Once its properties are defined more precisely than above and detection algorithms can be developed, it will be possible discriminating micro-pursuits from other FEM, like micro-saccades.

2.2.3 Descriptive statistics for the classification of micro-pursuits

Choosing an adequate metric for analysis was key, given the constraints presented in the previous section and our experimental setup. Two metrics, **inertia** and **maximally projected correlation** (MPC), are used in this work; they provide complementary information about the data. The first is a measure of the spatial dispersion of the gaze within a fixation to investigate the marginal dynamics of the gaze during FEM. The second metric gives a quantification of similarity—and hence interaction—between the gaze and a target. Compared to works in the literature with similar observations (Martins et al., 1985; Skinner et al., 2018), an essential aspect was to have a metric that could reflect similarity with noise robustness, as well as scale and translation invariance. Moreover, this was needed in the context of movements in the plane, rather than rectilinear ones for which uni-variate measures are sufficient. A benefit from such considerations is to propose a generalised metric for micro-pursuit that could be applied to track perceived motion in the two-dimensional visual field projected on the retina. MPC offers a method to quantify spatio-temporal similarity between two bi-variate signals. Furthermore, inertia and MPC can both be applied on the gaze signals in fixation epochs detected by video-based eye-tracker algorithm. Their mathematical relationship is detailed more in-depth in Appendix B.1.

2.2.4 Measuring gaze dispersion with inertia

The dispersion of gaze within a fixation was computed using a measure of inertia, a metric used to quantify the spread of a cloud of data points with respect to a fixed point, usually its empirical mean. Here, we used a similar, but generalised formula based on the mean quadratic distance from an arbitrary reference point. As such, in the case of stimulus motion, we can compute inertia with respect to the stimulus' centre of gravity. Let $\overline{\boldsymbol{q}}_U \doteq \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{q}_U^i$ be the empirical mean of a signal whose samples ($i = 1, \ldots, N$) are given by $\boldsymbol{q}_U^i = [x_U^i, y_U^i]^\top$. We will use $U = G$ for the observed gaze and $U = S$ for the coordinates of the stimulus' (centre of gravity). Gaze inertia I was computed over the stimulus trajectories over a trial as follows:

$$I = \frac{1}{N}\sum_{i=1}^{N} \left(\boldsymbol{q}_G^i - \boldsymbol{q}_O^i\right)^\top \left(\boldsymbol{q}_G^i - \boldsymbol{q}_O^i\right) = \frac{1}{N}\sum_{i=1}^{N} \|\boldsymbol{q}_G^i - \boldsymbol{q}_O^i\|^2 \tag{2.1}$$

where N represents the total number of frames in the trial, $\boldsymbol{q}_G = [\boldsymbol{x}_G, \boldsymbol{y}_G]^\top$ the measured monocular bi-variate gaze signal coordinates and $\boldsymbol{q}_O = [\boldsymbol{x}_O, \boldsymbol{y}_O]^\top$ the origin reference point coordinates in the screen plane—however, one can compute inertia with respect to other points in space, e.g., stimulus centre of gravity or the fixation's

mean gaze position. Inertia quantifies gaze displacement as does BCEA (Epelboim and Kowler, 1993) and box-count measures (Engbert and Mergenthaler, 2006). Its key advantage over the former two is that inertia is a more intuitive measure of spatial displacement over a fixation period. The box-count metric is simple and provides similar insight in gaze dispersion over an epoch, it is dependent on the size of the box in space and time used for analysis. Hence, it corresponds to a down sampling measurement of inertia over a fixed time window. Finally, inertia provides the advantage of being a metric relative to a chosen origin or reference point—box count being independent of the origin—and thus it can be used to look at spatial displacement in the following three contexts: (1) absolute inertia (I_{screen}) is obtained by choosing the centre of screen as a reference (absolute, like box count; $\boldsymbol{q}_O = [0, 0]^\top$), (2) relative retinal image instability (I_{stimulus}) by choosing the stimulus' centre of gravity (for pursuit; $\boldsymbol{q}_O = \boldsymbol{q}_S = [\boldsymbol{x}_S, \boldsymbol{y}_S]^\top$), and (3) general relative FEM instability (I_{fixation}) by referring to the fixation centre of gravity (obtained by choosing $\boldsymbol{q}_O = \overline{\boldsymbol{q}}_G = [\overline{\boldsymbol{x}}_G, \overline{\boldsymbol{y}}_G]^\top$ with $\overline{\boldsymbol{q}}_G$, the empirical mean of the gaze for a N samples fixation epoch).

2.2.5 Measuring gaze-stimulus similarity with MPC

Though humans can intuitively express a qualitative judgement of similarity between two trajectories, obtaining a quantified and objective value for any two bi-variate signals is not as trivial as one might suppose. Gain, of gaze velocity over stimulus velocity, has been used as a metric in pursuit data analysis (Skinner et al., 2018), though the stimulus moved in a uni-variate context: either horizontal or vertical. In bi-variate signals, however, a gain will be obtained for each dimension of the signal, and hence some form of projection to obtain a scalar metric is required. Although similarities between the stimulus and gaze trajectories can be quantified with a diversity of metrics, we will here focus on a measure based on multi-variate statistical theory (T. Anderson, 2003; Muirhead, 2009), quantifying the interaction between the stimulus ($\boldsymbol{q}_S$) and gaze ($\boldsymbol{q}_G$), in order to infer on the similarity of their trajectories during fixations. We choose to determine the direction of the plane for which correlation between gaze and target within a fixation are maximised, and report the such obtained correlation value, which we call MPC. Our metric hence inherits the ease of interpretability from (Pearson) correlation values and has low computational costs (just as gain). In addition, for unidirectional motion (see, e.g., (Skinner et al., 2018)), this exactly corresponds to Pearson's correlation coefficient between the two time-series.

Let $\mathbf{\Sigma}_{SG} \doteq \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{q}_S^i {\boldsymbol{q}_G^i}^\top - \overline{\boldsymbol{q}}_S \overline{\boldsymbol{q}}_G^\top$ the empirical (variance-)covariance matrix between stimulus (S) and gaze (G). We then write ρ^* as the maximal absolute empirical correlation that can be obtained under simultaneous projections onto a one-dimensional space, i.e.,

$$\rho^* \doteq \max_{\boldsymbol{w}} \rho(\boldsymbol{w}) \quad \text{where} \quad \rho(\boldsymbol{w}) \doteq \frac{\boldsymbol{w}^\top \mathbf{\Sigma}_{SG} \boldsymbol{w}}{\sqrt{\boldsymbol{w}^\top \mathbf{\Sigma}_{SS} \boldsymbol{w}}\sqrt{\boldsymbol{w}^\top \mathbf{\Sigma}_{GG} \boldsymbol{w}}} \tag{2.2}$$

and $\boldsymbol{w}$ represents the coordinates of the vector onto which both the gaze and the stimulus signal are projected. This method projects the data in a new space, and provides a quantity bounded between -1 and 1, where 1 shows perfect correlation and -1 perfect anti-correlation. By construction, MPC is invariant with respect to scale and to a translation of either or both of the signals.

To summarise this section, in this work, inertia with respect to screen (I_{screen}) was used as a measure of gaze displacement. Inertia with respect to stimulus (I_{stimulus}) was used as a measure of retinal image displacement. Inertia with respect to fixation (I_{fixation}) was used as a measure of FEM displacement. And finally, MPC (ρ^*) was used as a measure of similarity between gaze and stimulus trajectory, during a fixation.

2.3 Main Experiment: Necker cube

Micro-pursuits were observed and systematically detected at first in an experiment in which a moving ambiguous Necker cube stimulus was presented and participants had to report their perceived orientation. They were instructed to keep their gaze fixed on a static fixation cross at the centre of the screen and report which side of the cube was perceived at the front; either lower-left or upper-right square. The main objectives of the experiment was to manipulate the rate of reversal by imposing different motion to the cube. In this chapter, we focus solely on the oculomotor analysis of this data set, because the manipulation failed to induce any change in the reversal rate between the percept nor any observable percept modulation.

2.3.1 Methods

Apparatus

The display used was a 40 cm by 30 cm (20 inches) VisionMaster Pro 513 screen of resolution 1024 by 768 pixels and a 75 Hz refresh rate, located 57 cm from the participants, with mean grey luminance at 68 cd.m^{-2}. Eye movements were recorded using the Eyelink 1000 (SR Research, Ottawa, Ontario, Canada). Both eyes were tracked with a 1000 Hz sampling rate. The head was stabilised using a chin rest. A nine-point calibration routine was carried out at the beginning of each task and was repeated at the beginning of each block (every 15 trials) or when drift correction, performed every 5 trials, reported a mean error superior to 0.5 deg.

Experimental paradigm & Participants

We imposed three type of motion to an ambiguous Necker cube of 2.6 by 2.5 deg (Fig. 2.1-A): (1) '**FX**' the control condition with no motion, (2) '**RW**' an unpredictable motion condition with a random walk and (3) '**LJ**' the predictable motion condition where the cube moved along Lissajous trajectories (see Fig. 2.1-B). **Random walk** trajectories were implemented by choosing at each time step an amplitude chosen from an exponential-Gaussian distribution and an orientation from a uniform distribution on $(-\pi, \pi)$. The exponential-Gaussian distribution was built from the sum of two independent variables, $\varrho = G + E$ where $G \sim \mathcal{N}(\mu = 1.1;\ \sigma = 0.2)$ is the Gaussian component, and $E \sim \mathcal{E}(\lambda = 0.1)$ is the exponential one—units are in pixels (pix) and the $\sim$ symbol stands for "distributed according to". A radial limit

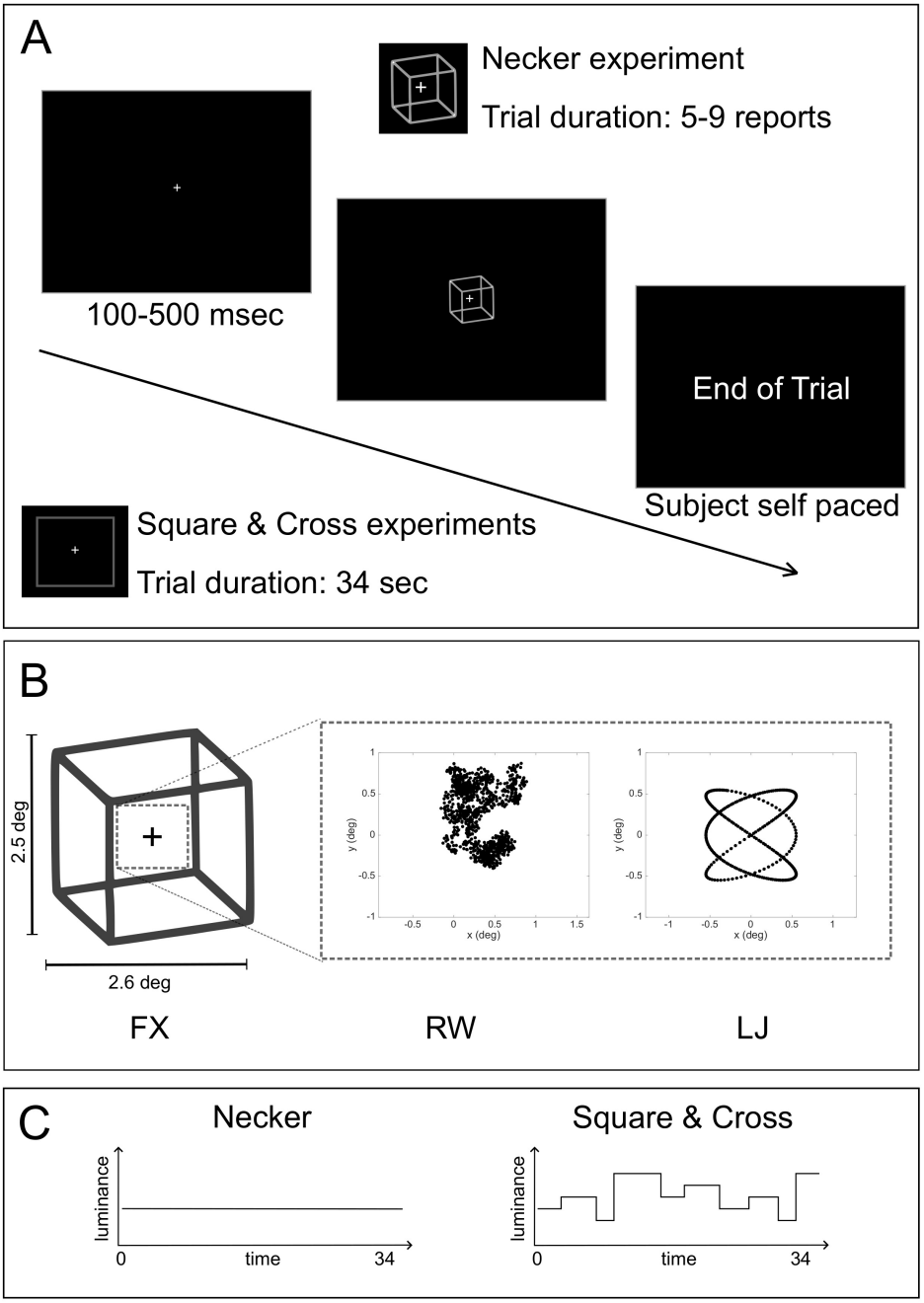


Figure 2.1. Experimental protocols.
A is a timeline of a trial for all three experiments (Necker, Square, Cross). For the Necker experiment, a Necker cube was displayed and the trial finished if the participant had reported a randomly picked number of perceptual reversals. For the Square and Cross experiments, a plain square was displayed and trial lasted approximately 34 seconds. A fixation cross was shown during a randomly chosen interval between 100 and 500 msec.
B shows the three different stimulus motion conditions; (1) FX, for the control no-motion condition, (2) RW, for the unpredictable random walk condition, and (3) LJ, for the predictable motion based on Lissajous trajectories.
C are representations of the stumuli's luminance. For the Square & Cross experiments, luminance changed randomly between 5 levels in order to provide the participants with a perceptual report task, while the Necker cube always kept a constant luminance.

of 10 pix (0.329 deg) with respect to the centre of the screen was implemented so that a step that would exceed the limit would have its orientation reversed such that the step would bounce back towards the centre. **Lissajous** trajectories in the LJ condition were defined by $x(t) = A\sin(c\theta t)$ and $y(t) = B\sin(d\theta t + \phi)$ with, in our setup, $A = B = 14$ pix (0.5497 deg) and $\phi = 0$ rad. The Lissajous ratio between signal frequencies randomly (uniformly) chosen across trials so that $(c, d) \in (2, 3), (3, 2), (-2, 3), (-3, 2)$ and $\theta = 2\pi\frac{(30/2.21)}{415} = 0.2$ Hz. The parameters' values were chosen empirically through *ad hoc* tests.

Stimulus spatial displacement due to movement was controlled across motion conditions. Indeed their inertia with respect to screen distribution were similar, with RW and LJ generating displacement of the same order of magnitude on average over trials ($\bar{I}^{RW}_{\text{screen}} = 0.2995 \pm 0.1988$, $\bar{I}^{LJ}_{\text{screen}} = 0.2747 \pm 0.1372$).

23 healthy adults participated in the experiment (15 females and 8 males; age range = 20–71 years, $\mu = 28.35 \pm 10.93$ years, whose tasks were two-fold:

- fixate a fixation cross at the centre of the screen for a random interval between 100 and 500 ms (uniform distribution);
- report percept reversals of an ambiguous Necker cube by pressing the arrows of a keyboard when perceptual changes occurred.

The experiment followed a continuous viewing paradigm in which trials had variable (random) durations ($\mu = 34.00 \pm 13.26$ sec, see Fig. 2.1-A) and ended based on which of the following condition happened first

number completion of a trial-based randomly (unifromly) set integer number ($n_{\text{rev}} \sim \mathcal{U}(5, 9)$) of perceptual reversals on the ambiguous stimulus (see Fig. 2.1-A);

time-out maximal percept duration of 20 sec.

The experiment was programmed using the *PsychToolBox* in MATLAB (Brainard, 1997). All participants gave their informed written consent before participating in the study, which was carried out in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki) for experiments involving humans and as approved by the ethics' committee of University Grenoble Alpes.

Data analysis

Data pre-processing: in our data analysis, only fixations of sufficient duration (> 80 ms) were considered. The duration threshold was set based on (1) the lack of significant fixations of interest in shorter time windows and (2) the necessity for the MPC metric to have a sufficient number of samples (see Appendix B.1). Gaze signals were first passed through a corrective process to adjust for pupil area deformation as described in Choe and colleague's work (Choe et al., 2016). As the gaze and stimulus signals were systematically compared and computed together, we then applied a Butterworth filter (second order low-pass filter with a cut-off frequency of $f_c = 35$ Hz) to smooth the gaze data and down-sampled the gaze signal at the same frequency as the refresh rate of the stimulus (75Hz). Thus, all analyses are done with data down sampled from 1000 Hz to 75 Hz. Fixations generating inertia with respect to screen values beyond two standard deviation from the mean or NaN (due to missing samples) were considered as samples with faulty or jittery gaze recording and were removed from analyses. Data for Fig. 2.3 and statistical tests had fixations with micro-saccades, detected—by the algorithm proposed by Engbert and Kliegl (Engbert and Kliegl, 2003) without the binocularity criterion, that uses relative thresholds based on median absolute deviation of the eye velocity, here over a fixation—and removed, while data for Fig. 2.2 and Fig. 2.5 was visualised with micro-saccades. Outliers were defined as data points[3] beyond two standard deviation from the mean, and were systematically removed from analyses. The results presented do not show these outliers, for better readability, but we also conduct the analyses with the outlier and found the same effects for all tests and experiences.

Statistical methods: statistical tests were conduct to assess difference between motion condition both within subjects (*group analysis*) and at the subject level (*individual analysis*). For both levels, we applied non parametric tests, since we did not have any priors on the data distribution for inertia and MPC. For group analysis, statistical tests were conducted using 10000 permutations on non parametric approximate (Monte Carlo) Friedman test for inertia, and if significant differences were inferred, approximate (Monte Carlo) Wilcoxon signed-rank tests were used for pairwise comparisons between conditions (with a decision criterion at $p = 0.05/3 = 0.017$). For MPC, a Wilcoxon signed-rank test was carried out All these tests were delivered using bootstraps based on 10000 permutations conditional on subjects for every experiments (Necker, Cross and Square) and metrics (I_{stimulus},

[3]Here, data points refer to a statistic of a fixation period, for a given experiment, subject and condition. We also have outliers subject (71 years old) that is not removed.

I_{fixation} and MPC) using the packages *coin* (Hothorn, Hornik, Van De Wiel, et al., 2006) and *rstatix* (Kassambara, 2020). Effect size were computed from the χ^2 statistics and using the transformation described by Tomczak and Tomczak (M. Tomczak and E. Tomczak, 2014) to get a Kendall W, that vary between 0 and 1, with 1 the maximum effect size:

$$W = \frac{\chi^2}{N(k-1)}. \tag{2.3}$$

With W, the Kendall's W value, χ^2 the Friedman test statistic value, N the sample size and k the number of measurements per subject. For each test, we report the χ^2 Friedman test statistic, with the p-value (p) computed with the bootstrap, it's effect size (Kendall W). For individual statistical analyses, we carried out an approximate Kruskal-Wallis test for inertia and an approximate Wilcoxon-Mann & Whitney test for MPC and pairwise comparisons using the same bootstrap package, with 10000 permutations. To compare experiments' data, Kruskal-Wallis tests were used over the three experiments' RW and LJ data, respectively, and Wilcoxon-Mann & Whitney tests were used to infer differences between pairs of experiment data-sets in each condition, with the same packages.

Additional analyses. Two analyses were conducted (after submission) and are presented here to provide further depth and insight on the data. First, we looked at MPC on the signals' velocities to verify that we replicated the results found on positions. Analyses were identical to position MPC, and the velocity was obtained by using the method proposed by Engbert and Kliegl (2003), for the Engber-Kliegl (EK) algorithm, by computing a moving average of velocities over 5 data samples, in order to reduce noise. The second addition analysis consisted in looking at the similarity between the two eyes' position signals, to verify whether the gaze data was conjugated or not. MPC scores were computed for each fixation between the directing and the non-directing eye, with the former acting as the reference, and statistical analysis was carried out, using the same procedure as for inertia, over all conditions.

2.3.2 Results—Corrected in Appendix C

When fixations with detected micro-saccades were kept, data pre-processing led to the removal of 12.32% of fixations for the Necker experiment based on fixation duration and outlier removal for inertia with respect to screen. When fixations with detected micro-saccades were removed, data pre-processing led to the removal of 54.53% of fixations. Results presented in this section were computed on the fixations

without micro-saccades, however when doing these analyses with fixations with micro-saccades, results led to the same conclusions.

Main sequence (and secondary sequence)

In the first column of Fig. 2.2, we described detected micro-saccades ($n = 27101$), using the algorithm from Engbert and Kliegl's work (Engbert and Kliegl, 2003), by amplitudes, peak velocities and rates. When plotting micro-saccades' amplitudes versus peak velocities, we observed two main sequences (Bahill et al., 1975) with a second main sequence with low velocity micro-saccades. The second main sequence appeared more in the LJ condition than in the RW. Furthermore, when using the MPC score, one can observe that the slow micro-saccades are mostly detected in fixations with gaze patterns highly similar to the stimulus', namely micro-pursuits, which is indicated by the dominance of yellow dots. Micro-saccade rates also seem to suggest that the LJ condition has more occurrences, over all amplitudes. This will be further analysed with the evaluation of dispersion of FEM and the similarity of their trajectories with the stimulus.

Inertia & MPC

We looked at the impact of the cube motion on eye movement and retinal image displacement. The former is made explicit through the inertia of gaze with respect to its average position within a fixation, see Fig. 2.3-B, whereas the latter is given by the inertia of the gaze with respect to the stimulus' centre of gravity, see Fig. 2.3-A. Descriptive statistics and statistical tests' summary are given in Table 2.1.

Dispersion of eye movements around the fixation, computed with median inertia of the eye with respect to mean fixation position (I_{fixation}; see Fig. 2.3-C) differed with motion condition ($\chi^2 = 36.261; p < 0.0001; W = 0.788$). Paired comparisons of I_{fixation} showed differences between FX, RW and LJ ($Z_{FX-RW} = -5.9052, p < 0.0001; Z_{RW-LJ} = -5.9052, p < 0.0001$ and $Z_{FX-LJ} = -5.9052; p < 0.0001$). Thus, when computing retinal image displacement, we found that the median inertia differed across cube motion conditions (see Fig. 2.3-A). Indeed, we find a difference in inertia computed with respect to the centre of gravity of the stimulus (I_{stimulus}) with motion condition ($\chi^2 = 28.783; p < 0.0001; W = 0.626$). Median inertia differed in the conditions where the stimulus was in motion ($Z_{FX-RW} = 2.890, p = 0.0032; Z_{FX-LJ} = 2.890, p = 0.0028$ and $Z_{LJ-RW} = 5.9052, p < 0.0001$).

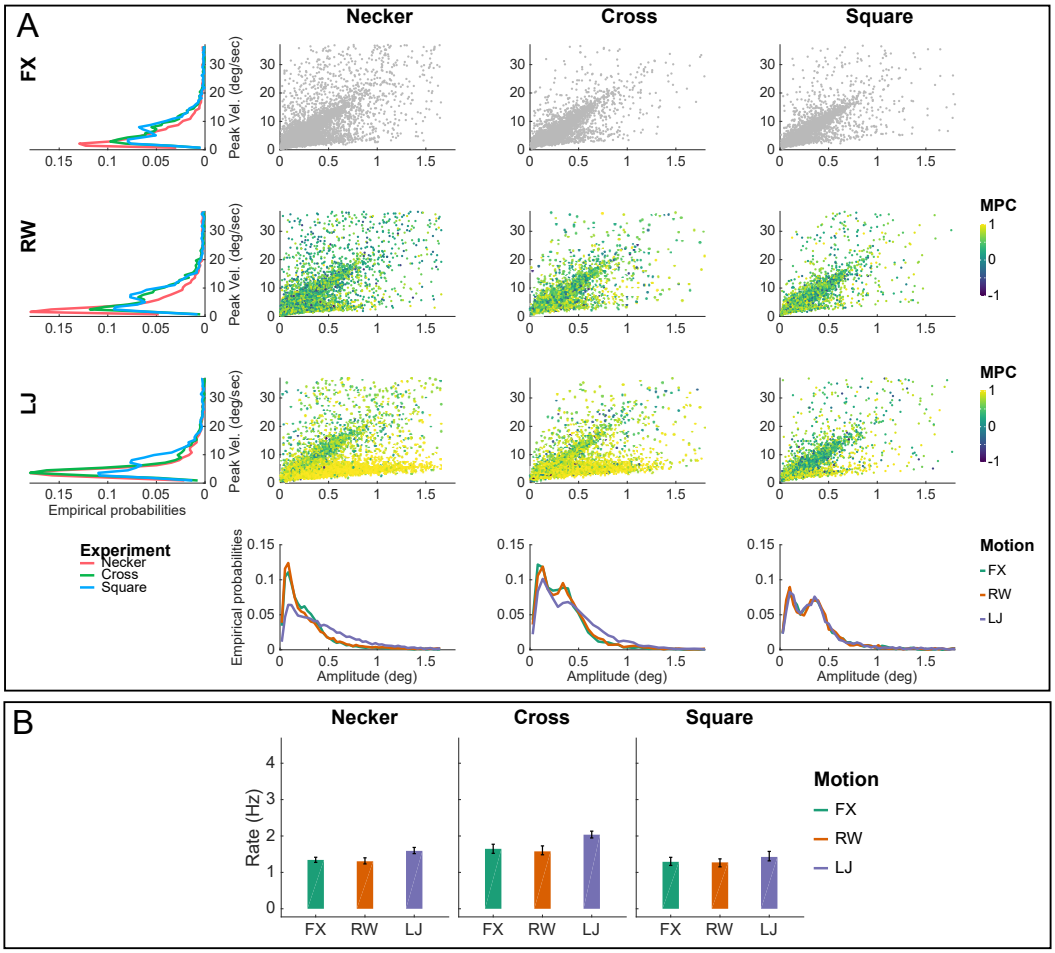


Figure 2.2. Micro-saccade analysis.
A shows the main sequences when plotting micro-saccades' amplitudes versus peak velocities for all three experiments (Necker, Cross and Square) and conditions. The colour codes for each the micro-saccade's fixation similarity score (using MPC) in the LJ and RW conditions. The LJ condition in Necker and Cross experiments shows a secondary main sequence correlated to micro-pursuit occurrences. *Left side*, marginal distributions of peak velocity depending on the experiment and condition are given, while *below*, marginal distributions for amplitudes are shown.
B shows mean micro-saccade rates over experiments and conditions with, in black, 95% confidence intervals computed using bootstrap ($n = 200$ iterations).

When considering that stimulus inertia was equivalent for both motion conditions ($\bar{I}^{RW}_{\text{screen}} = 0.2995 \pm 0.1988$,$\bar{I}^{LJ}_{\text{screen}} = 0.2747 \pm 0.1372$), the results suggest that both types of motion applied on the stimulus generated different effects on eye movements. Indeed, eye trajectories were more similar in the predictable LJ motion condition ($\tilde{\rho}^*_{LJ} = 0.869 \pm 0.081$) than in the unpredictable RW motion condition ($\tilde{\rho}^*_{RW} = 0.477 \pm 0.035$) with significant differences ($\chi^2 = 23; p < 0.0001; W = 1$ and $Z_{RW-LJ} = -5.9052; p < 0.0001$). The data is reported in Fig. 2.3-E. We evaluated the effect of the cube motion for every subject and found similar results (Fig. 2.3-B-D-F) that will be described in more details later.

$(N = 23)$	FX	RW	LJ	χ^2	p	W
$\tilde{I}_{\text{stimulus}}$	0.458 ± 0.151	194.404 ± 20.301	173.708 ± 12.018	42.348	< 0.0001	0.921
$\tilde{I}_{\text{fixation}}$	0.018 ± 0.010	0.020 ± 0.012	0.070 ± 0.042	36.261	< 0.0001	0.788
$\tilde{\rho}^*$	n/a	0.477 ± 0.035	0.869 ± 0.081	23	< 0.0001	1

Table 2.1. *Left*, Summary statistics of three measures for the Necker experiments in the FX, RW and LJ motion conditions; inertia w.r.t. stimulus centre of gravity (I_{stimulus}), inertia w.r.t. fixation centre of gravity (I_{fixation}), and MPC (ρ^*). For each condition in the Necker experiment, median values over participants' data are given with median absolute deviation (mad) following the $\pm$ sign.
Right, Approximate Friedman test results ($\chi^2; p$) and size effect (W) are given.

Binocularity & velocity

To confirm the MPC results on position signals, we proceeded to the same analysis with velocities. In fact, as for the position analysis, LJ's predictable motion ($\tilde{\rho}^*_{LJ} = 0.784 \pm 0.087$) led to velocities more similar between the eyes and the target than for RW's unpredictable motion ($\tilde{\rho}^*_{RW} = 0.207 \pm 0.042$) with significant differences ($\chi^2 = 23; p < 0.0001; W = 1$ and $Z_{RW-LJ} = -5.9052; p < 0.0001$). The data is reported in Fig. 2.4-C, along with analyses for each participants Fig. 2.4-D.

We also looked at the similarity of gaze between the directing and non-directing eye, to look at how conjugated the eyes were. We found overall differences across conditions ($\chi^2 = 39.39; p < 0.0001; W = 0.856$). Paired comparisons of eye *versus* eye similarity showed differences between FX, RW and LJ ($Z_{FX-LJ} = -5.9052, p < 0.0001; Z_{FX-RW} = -5.9052, p < 0.0001$ and $Z_{LJ-RW} = -5.9052, p < 0.0001$). Results are reported in Fig. 2.4-A, along with analyses for each participants Fig. 2.4-B.

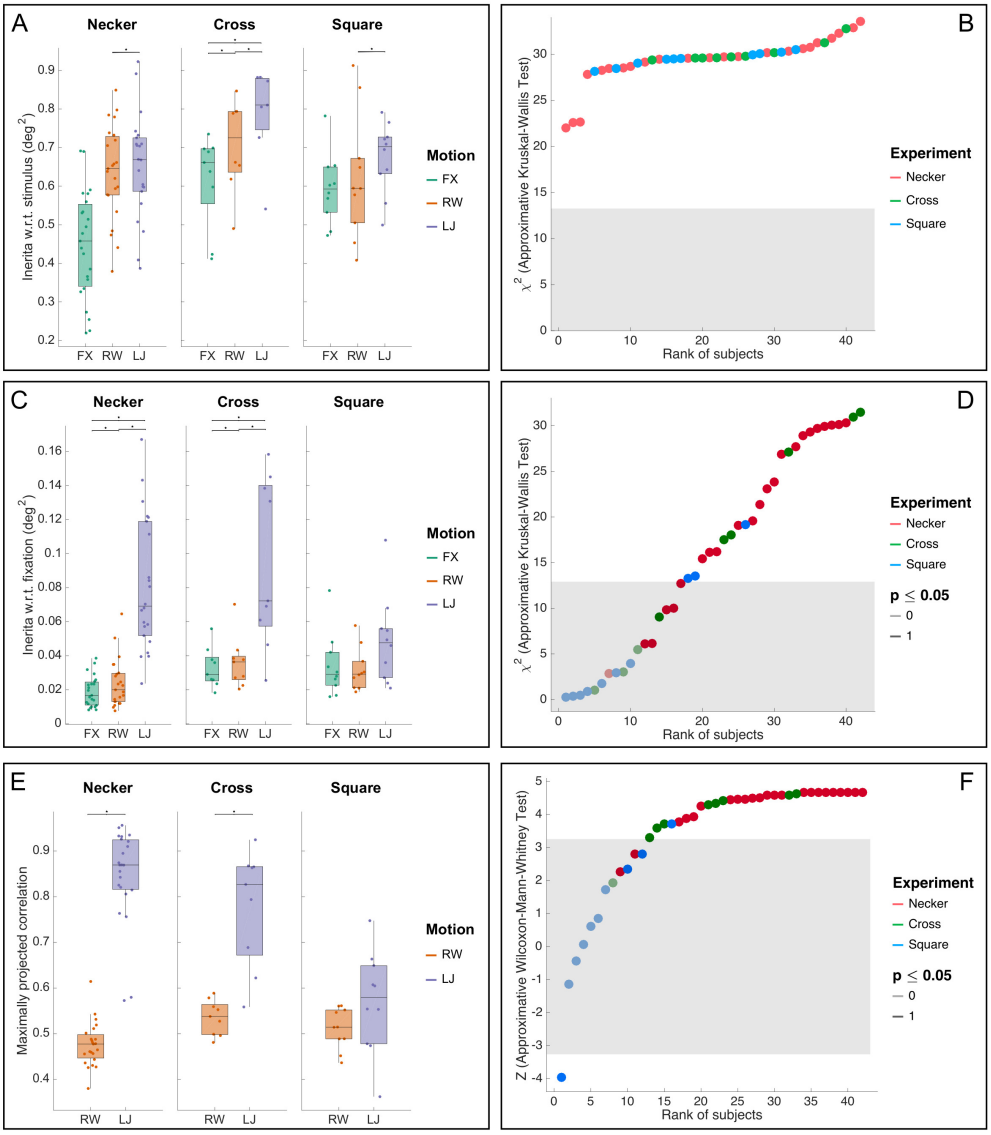


Figure 2.3. Micro-pursuit analysis.
A is a box plot of I_{stimulus} over the three experiments (Necker, Cross and Square) and three motion condition (FX, RW and LJ). Stars represent significant differences in pairwise comparisons using the Wilcoxon-Mann-Whitney test in a bootstrap. **B** plots the individual analysis results for I_{stimulus} in all three experiments' participants using an approximate Kruskal-Wallis test in a bootstrap. All the participant have significant ($p < 0.05$) results. For individual analysis, statistics (Z score or χ^2) that fall inside the 95 % confidence interval were drawn with light colour whereas statistics values outside the 95% confidence interval were drawn in plain colour. The grey area defines a conservative confidence interval corrected for multiple comparisons (Bonferroni), i.e. 42 comparisons for the 42 tests computed on each subjects. **C** is a box plot of I_{fixation} over all experiments and conditions. **D** plots the individual analysis results for I_{fixation}. The outcome of the statistical test per participant are given through different lightness value, with 1 (darker) meaning that $p \leq 0.05$ and 0 (lighter) the opposite. **E** is a box plot of MPC (ρ^*) over all experiments and the RW and LJ motion conditions. **F** plots the individual analysis results for ρ^* in all participants using an approximate Wilcoxon-Mann-Whitney test.

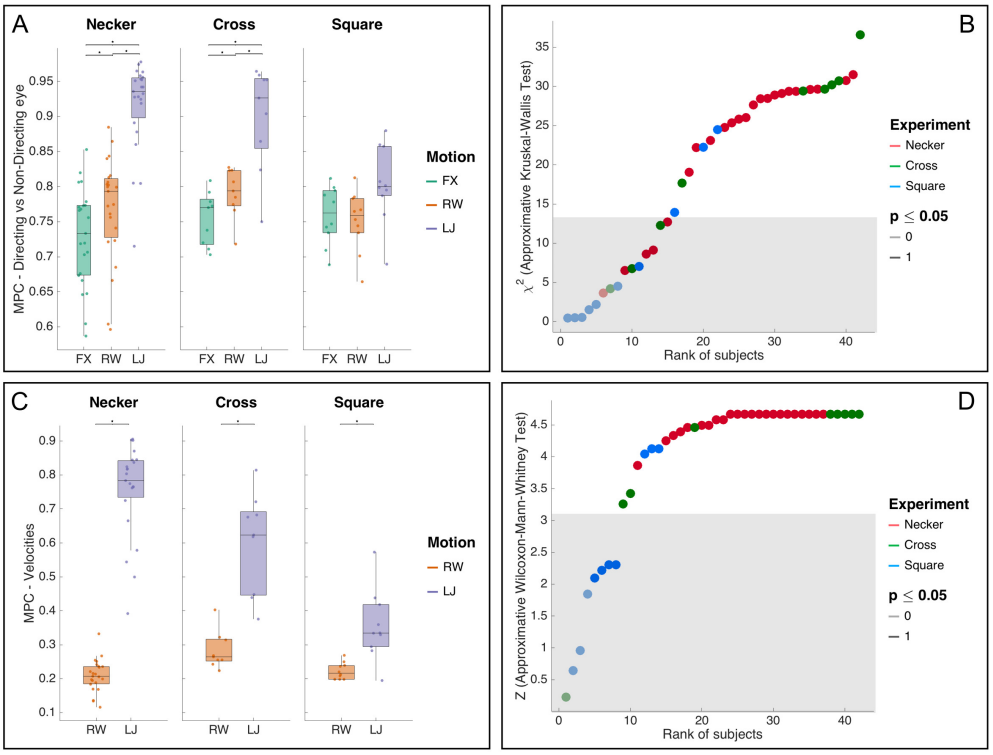


Figure 2.4. Micro-pursuit additional analysis.
A is a box plot of directing vs non-directing eye similarity over the three experiments (Necker, Cross and Square) and three motion condition (FX, RW and LJ). Stars represent significant differences in pairwise comparisons using the Wilcoxon-Mann-Whitney test in a bootstrap. **B** plots the individual analysis results for directing vs non-directing eye similarity in all three experiments' participants using an approximate Kruskal-Wallis test in a bootstrap. For individual analysis, statistics (Z score or χ^2) that fall inside the 95 % confidence interval were drawn with light colour whereas statistics values outside the 95% confidence interval were drawn in plain colour. The grey area defines a conservative confidence interval corrected for multiple comparisons (Bonferroni), i.e. 42 comparisons for the 42 tests computed on each subjects. **C** is a box plot of MPC based on velocity vectors over all experiments and conditions. **D** plots the individual analysis results for MPC based on velocity vectors. The outcome of the statistical test per participant are given through different lightness value, with 1 (darker) meaning that $p \leq 0.05$ and 0 (lighter) the opposite.

2.3.3 Intermediary discussion

Micro-pursuits. When looking at our descriptive statistics (Table 2.1 and Fig. 2.3, A-B-C), participants' median similarity based on MPC is centred on values of high correlation in the predictable motion condition (LJ) compared to the other motion condition (RW). This means that FEM gaze trajectories were, for most subjects, highly similar to that of the stimulus moving on screen. On the other hand, the unpredictable motion condition (RW) led to much lower similarity measurements; an observation that can be explained by the incapacity of the oculomotor system to predict the motion of the Necker cube as motion followed random walk dynamics.

(Un)attended motion. Therefore, globally, participants' gaze was influenced by the cube motion significantly more in LJ, where motion was predictable, than in RW, where motion was unpredictable, even though the oculomotor instructions were to fixate the cross in the middle of the screen for both. Moreover, the gaze in LJ showed similarity with the stimulus trajectories. All these measures were gathered on gaze data within fixation events and the difference between LJ and RW conditions show that oculomotor drift alone, as defined above, within FEM cannot account for this similarity. The oculomotor system would have to integrate visual information in order to quasi-systematically track the stimulus. We therefore refer to these detected FEM as micro-pursuits, in an effort to keep the analogy with the micro-saccades, while respecting the definition and metrics given above. Given the non-dedicated and unpredicted observation of the oculomotor phenomenon in the Necker experiment, we carried out a second set of experiments to replicate the generation of micro-pursuit using a simpler stimulus, and to verify that the phenomenon is not caused by the presence of a bi-stable stimulus—namely the Necker cube.

2.4 Replication Experiments: Square & Cross

The experimental protocol is similar to the previous one (Necker experiment) except that the Necker cube is replaced by a grey square and subjects have to report changes in luminance in either the fixation cross (Cross experiment) or the square (Square experiment). In the Cross experiment, we set the participants' tasks and stimuli such that they had to follow a moving cross and detect changes of luminance on it. In the Square experiment, the setup aimed to investigate whether a complete reproduction of the Necker experiment, with a square instead of the Necker cube would still lead to the observation of micro-pursuits.

2.4.1 Methods

Material and stimuli were identical to the previous experiment unless specified.

Apparatus

The stimulus was displayed on a 36 cm by 27.5 cm (19 inches) Dell M993s CRT screen of resolution 1280 by 1024 pixels and a 75 Hz refresh rate, located 57 cm from the participants, with white luminance at 70.89 $cd.m^{-2}$, black at 0.09 $cd.m^{-2}$ and mean grey at 15 $cd.m^{-2}$. Eye tracking was done using an EyeLink 1000+ (SR Research). Calibration was applied using a 5 points procedure between each block and if drift correction failed. Drift correction was applied between each trial. Participants had their head stabilised by sitting and resting their chin on a rest and their forehead against a bar.

Experimental paradigm & Participants

As in Experiment 1, we replicated the three motion conditions (FX, RW, & LJ) using the same parameters with balanced mean inertia. Trials lasted 34 seconds (the mean time duration of *Experiment 1: Necker Cube*) in which the same fixation cross was presented, and a moving object followed its trajectories depending on the condition (see Fig. 2.1-A).

The participants had to fixate a fixation cross surrounded by a square (2.5 deg by 2.5 deg), displayed in Fig. 2.1-A. They also had a perceptual task in which they had to report luminance changes using the same keys of the keyboard as in the Necker

Experiment. However, here the alternations were randomly selected among 5 levels of luminance (levels at 30%, 40%, 50%, 60% and 70% of white) and duration of a level was selected using a log-normal probability law $\mathrm{Log}{-}\mathcal{N} \sim (\mu = 1, \sigma = 1)$ seconds (see Fig. 2.1-C for a schematic representation of luminance over time). Two conditions were contrasted:

1. *Implicit pursuit - Square Attended*: fixate the fixation cross at the centre of screen, and report changes in luminance of the surrounding square moving with the three types of motions.

2. *Explicit pursuit - Cross Attended*: fixate the fixation cross and report changes in luminance of the fixation cross moving with the three types of motions. Do not pay attention to the surrounding square.

The 19 participants (17 females and 2 males; age range = 18-30 years, $\mu = 20.63 \pm 2.61$ years) were randomly oriented in one of the two experiments (Cross; $n = 9$, and Square; $n = 10$). We estimated the number of participants to be included in the protocol based on a power analysis using *g*power* (Faul et al., 2009) with $\alpha = 0.05$ and $1 - \beta = 0.95$. We found that we needed a minimum sample size of 9 participants (with 45 trials) to replicate the observations with a power of 0.95.

Data analysis

Data analyses were identical to the previous experiment.

2.4.2 Results—Corrected in Appendix C

The data was analysed by applying the same signal processing procedures and statistical methods as in the Necker experiment for inertia or MPC. When fixations with micro-saccades were kept, data pre-processing led to the removal of 8.79% and 9.23% of fixations for the Cross and Square experiments, respectively, based on fixation duration and outlier removal for inertia with respect to screen. When fixations with micro-saccades were removed as well, data pre-processing led to the removal of 52.88% and 52.59% of the data, in Cross and Square, respectively. Results presented in this section were computed on the fixations without micro-saccades, however when doing these analyses with fixations with micro-saccades, results led to the same conclusions.

Cross experiment: explicit micro-pursuits

When participants had to explicitly follow the fixation cross, on which the motion and luminance signals were applied, similar patterns to the Necker experiment were found for inertia of gaze. Dispersion of eye movements around the fixation, computed with median inertia of the eye with respect to mean fixation position (I_{fixation}; see Fig. 2.3-C) differed with motion condition ($\chi^2 = 9.556; p = 0.0071; W = 0.531$). Moreover, paired comparisons revealed differences between FX, RW and LJ ($Z_{FX-RW} = -3.7236, p < 0.0001; Z_{RW-LJ} = -3.7236; p < 0.0001$ and $Z_{FX-LJ} = -3.7236; p < 0.0001$). Retinal image displacement differed with cube motion (see Fig. 2.3-A). We also found a difference in inertia computed with respect to the centre of gravity of the stimulus (I_{stimulus}) with motion condition ($\chi^2 = 12.667; p = 0.0005; W = 0.704$). All pairwise comparisons showed differences too ($Z_{FX-RW} = -3.576, p < 0.0001; Z_{FX-LJ} = -3.576, p < 0.0001$ and $Z_{RW-LJ} = -3.7236, p < 0.0001$).

Given the fact that stimulus inertia was equivalent for both motion conditions, this suggests that motion of the stimulus generated different effects on eye movements. Indeed, eye trajectories were more similar in the predictable LJ motion condition ($\tilde{\rho}^*_{LJ} = 0.830 \pm 0.064$) than in the unpredictable RW motion condition ($\tilde{\rho}^*_{RW} = 0.535 \pm 0.056$) with significant differences ($\chi^2 = 9; p = 0.0039; W = 1$ and $Z_{RW-LJ} = -3.7236; p < 0.0001$). The data is visualised in Fig. 2.3-E. We evaluated the effect of the cube motion for every subject and found similar results (Fig. 2.3-F).

Square experiment: implicit micro-pursuits

Dispersion of eye movements around the fixation, computed with median inertia of the eye with respect to mean fixation position (I_{fixation}; see Fig. 2.3-C) did not differ with motion condition ($\chi^2 = 5.6; p = 0.0659; W = 0.28$). But retinal image displacement differed with cube motion (see Fig. 2.3-A). Indeed, we find a difference in inertia computed with respect to the centre of gravity of the stimulus (I_{stimulus}) with motion condition ($\chi^2 = 7.2; p = 0.03; W = 0.36$). Pairwise comparisons had a difference only between the motion conditions RW and LJ ($Z_{FX-RW} = 1.8666, p = 0.0654; Z_{FX-LJ} = 1.8666; p = 0.0645$ and $Z_{RW-LJ} = 3.9199, p < 0.0001$).

Given the fact that stimulus inertia was equivalent for both motion conditions, this suggests that motion of the stimulus did not generate different effects on eye movements. Unlike in the other experiments, eye trajectories were not more similar to stimulus trajectories in the predictable LJ motion condition ($\tilde{\rho}^*_{LJ} = 0.569 \pm 0.129$)

or in the unpredictable RW motion condition ($\tilde{\rho}^*_{RW} = 0.519 \pm 0.039$) with no inferred statistical difference ($\chi^2 = 1.6; p = 0.345; W = 0.16$). The data is visualised in Fig. 2.3-E. We evaluated the effect of the cube motion for every subject and found similar results (Fig. 2.3-F).

Individual analyses

We conducted the same analysis on every subject and results are displayed for the three experiments and three motion conditions in figure (Fig. 2.3-B-D-F). For every subject, we plotted the χ^2 or Z score statistics for the approximate Kruskal-Wallis and Wilcoxon-Mann-Whitney tests against their overall rank according to these statistics. For all subject we observed a main effect of inertia with reference to the stimulus (I_{stimulus}, with identical inertia between LJ and RW compare to FX. When looking at retinal displacement, we find the same pattern of result, i.e. a main effect of motion, with inertia with reference to the fixation (I_{fixation}) similar for FW and RW but lower to LJ for Necker and Cross experiments. For the Square experiment results were mixed within subject suggesting idiosyncratic behaviours. Finally, we observe more similar gaze pattern (high MPC) for the LJ condition both in the Necker and Cross experiments for every subject (except one out of nine in Cross) but mixed results for the square experiment. Thus individual analyses show that results observed at the group level are replicated at the subject level.

Binocularity & velocity

To confirm the MPC results on position signals, we proceeded to the same analysis with velocities. For the Cross experiment, LJ's predictable motion ($\tilde{\rho}^*_{LJ} = 0.623 \pm 0.146$) led to velocities more similar between the eyes and the target than for RW's unpredictable motion ($\tilde{\rho}^*_{RW} = 0.264 \pm 0.032$) with significant differences ($\chi^2 = 5.44; p = 0.0434; W = 0.605$ and $Z_{RW-LJ} = -3.7236; p < 0.0001$). For the Square experiment, LJ's predictable motion ($\tilde{\rho}^*_{LJ} = 0.335 \pm 0.068$) led to velocities more similar between the eyes and the target than for RW's unpredictable motion ($\tilde{\rho}^*_{RW} = 0.216 \pm 0.027$) with significant differences ($\chi^2 = 6.4; p = 0.0227; W = 0.64$ and $Z_{RW-LJ} = -3.9199; p < 0.0001$). The data is reported in Fig. 2.4-C, along with analyses for each participants Fig. 2.4-D.

We also looked at the similarity of gaze between the directing and non-directing eye, to look at how conjugated the eyes were. Group median values and median absolute deviations for FX (0.77 ± 0.048), RW (0.794 ± 0.041) and LJ (0.927 ± 0.048) describe

the most binocular similarity, in fixations, in LJ. We found overall differences across conditions for the Cross experiment ($\chi^2 = 14.889; p < 0.0001; W = 0.827$). Paired comparisons of eye *versus* eye similarity showed differences between FX, RW and LJ ($Z_{FX-LJ} = -3.723, p < 0.0001; Z_{FX-RW} = -3.723, p < 0.0001$ and $Z_{LJ-RW} = -3.723, p < 0.0001$). We did not find overall differences across conditions for the Square experiment ($\chi^2 = 5.6; p = 0.0709; W = 0.28$). Results are reported in Fig. 2.4-A, along with analyses for each participants Fig. 2.4-B.

2.5 Comparing Necker, Cross and Square experiments—Corrected in Appendix C

To summarise, descriptive statistics of detected micro-saccades in terms of main sequences (amplitude, peak velocity and MPC; see Fig. 2.2-A) and micro-saccade rates (Fig. 2.2-B) in all three experiments show similar patterns under the condition LJ for both Necker and Cross. The Square condition seems to exhibit a different behavior. More specifically, the secondary, slower—in terms of peak velocity—main sequence observed in LJ-Necker and LJ-Cross is less present in LJ-Square and all RW conditions. The micro-saccade rates also seem to be higher in LJ-Necker and LJ-Cross, over other conditions and experiments. These aspect of the detected micro-saccades over all experiments suggest that the predictable LJ condition under Necker and Cross led to a slow, small amplitude, oculomotor phenomenon.

Fig. 2.5 provides a focus on MPC for fixations in all data sets, as well as for some selected signals that showcase some typical examples of gaze-stimulus pairs for different values of MPC. Since one cannot track the RW movements, the distribution of MPC under this condition serves as a baseline or null hypothesis control distribution. It can be seen that under RW, the empirically observed MPC distributions for all three experiments are confounded, indicating independence of MPC with respect to the experiment. Furthermore, it is also possible to observe a bias—the distribution is skewed toward the maximum value of 1—introduced by (i) the maximisation of the correlation through the projection of the data into another coordinate system, and (ii) the RW movement being low-pass filtered by the observer, hence there exists a correlation at longer time scales. Indeed, the distribution under RW is not symmetric about 0 as would be the case for mere correlation between variables of multivariate independent Gaussian processes. On the other hand, under the LJ condition the distribution skews even further to one, resulting in a high probability for MPC values near one, specifically in Necker and Cross. This is less so in Square.

When we removed fixations with detected micro-saccades and carried out inertia and MPC analyses, we found a difference for MPC in the LJ condition across experiments ($\chi^2 = 19.078; p < 0.0001$). When looking at pairwise comparisons (subscripts N for Necker, C for Cross, and S for Square), no significant differences were found between Necker and Cross ($Z_{N-C} = -1.572; p = 0.121$), but Square differed from the other two ($Z_{S-C} = 2.939; p < 0.0025$ and $Z_{N-S} = 4.113; p < 0.0001$).

For RW inter-experiment comparisons, we found an overall difference ($\chi^2 = 10.617; p = 0.0036$). Paired comparisons showed a difference between Necker and

the two other experiments ($Z_{N-C} = 2.955; p = 0.0020$ and $Z_{N-S} = -2.076; p = 0.0350$) but none for Square versus Cross ($Z_{S-C} = 1.061; p = 0.3114$).

Finally, results for individual analyses show that most participants in the Square experiment had no significant differences between MPC in RW and LJ, while on the contrary, all 23 participants in the Necker and 8 out 9 participants in Cross do.

> **Overall, these results indicate that Cross did replicate the micro-pursuit phenomenon observed in the Necker experiment even with a smaller sample size, while Square did not.**

Median inertia with respect to the stimulus' centre of gravity ($\overline{I}_{\text{stimulus}}$) differed with motion conditions suggesting that the nature of stimulus motion, manipulated in each condition (fixed, unpredictable, and predictable) affects global spatio-temporal dynamics of FEM. Median inertia with respect to the fixation's mean gaze position ($\overline{I}_{\text{fixation}}$) showed the emerging pattern of a common oculomotor phenomenon occurring in Necker and Cross, where differences across conditions were measured. Again, this was not the case in Square (see Fig. 2.3-C). When looking at similarity between stimulus and gaze trajectories, integrated over fixation events using MPC, we found that the predictable motion condition (LJ) generated highly similar gaze trajectories in the Necker and Cross experiments, with large effect sizes. But we did not observe the same pattern for the Square experiment (see Fig. 2.3-E).

The contrast given by diverging results (Necker-LJ and Cross-LJ being different from Square-LJ) is interesting as it gives us a graduation of how likely, the same predictable motion (LJ) can make observers generates micro-pursuit. It also suggests that a coupling between the oculomotor and cognitive systems in the occurrence of micro-pursuits, which could be predicted and interpreted by a modelling framework we proposed when encountering the original observations. To go further, we propose a model, in Chapter 3, that can describe all FEM in a single mechanism and can take into account the competition between multiple stimuli.

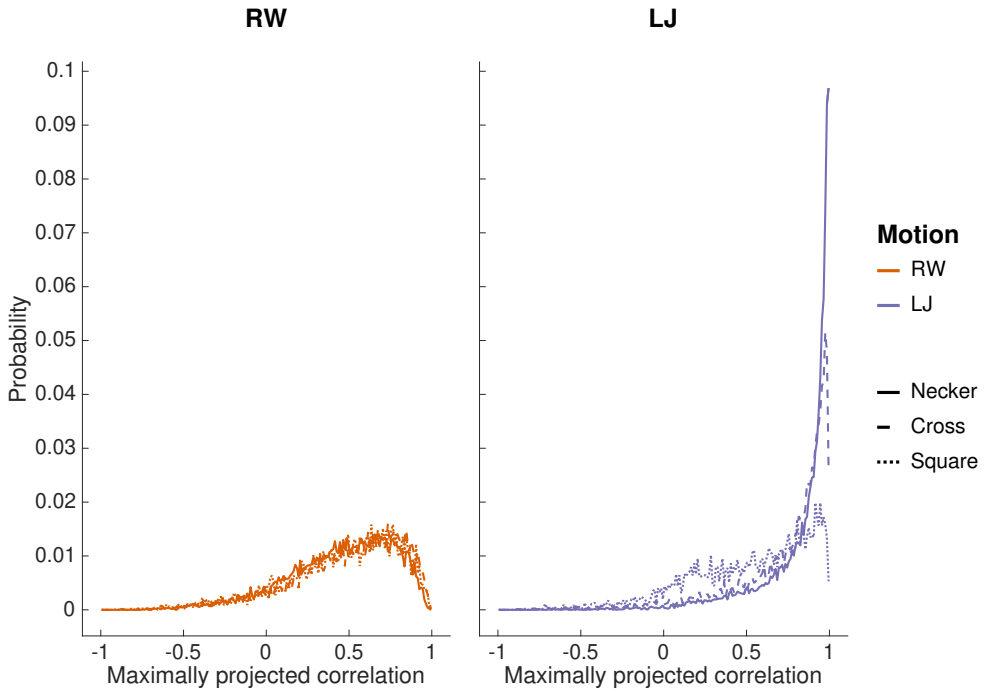


Figure 2.5. Focus on MPC results. Histogram of fixations by maximised correlation ρ^* (MPC) scores in the Necker Cube experiment. Illustrations of signals for values in some typical score intervals are presented to give a graphical intuition of the computed measure. We picked high similarity near a score of 1, no correlation near 0 and anti-correlation near -1. Dotted trajectories correspond to stimulus signals and continuous trajectories correspond to gaze signals. Temporal discourse is represented by lighter to darker samples.

2.6 Discussion—Corrected in Appendix C

2.6.1 Micro-pursuits

The proposed working definition of micro-pursuits in this chapter, is based on a class of fixations. Moreover, the MPC metric proposed can be applied directly to measure a fixation's similarity between gaze and stimulus, because it features scale, translation, and spatial offset invariances. Therefore, the results presented are based on entire fixations, rejecting those for which micro-saccades were detected. It is noteworthy that even when fixations containing micro-saccades are kept for analysis, our main results and trends, especially with respect to MPC are still valid at all three levels: individual, group and inter-experiment analyses. This can be partially observed from Fig. 2.5, where the MPC is obtained over all fixations, whether they contain micro-saccades or not. Indeed, we observe a clear difference between MPC distributions in

LJ versus RW. The trends in these observed differences were systematically obtained with (i) entire data including detected micro-saccades, (ii) when gaze samples of the detected micro-saccades were removed, or (iii) when considering only the fixations without detected micro-saccades—the reported data for analyses and Fig. 2.3 being the most conservative. In this work we thus focused on a proof of micro-pusuits' existence through the results obtained from the Necker experiment as well as results from the replication experiments (Cross or Square).

The additional analyses on binocular similarity and velocity similarity also provide more information on micro-pursuits. The former's results suggest that micro-pursuits tend to feature conjugated dynamics between both eyes since the LJ condition, in they were measured, showed more eye-to-eye similarity, in particular in the two data set where micro-pursuit are systematically measured (Fig. 2.4-A). Secondly, the replication of the MPC results on position and velocity signals adds weight to the reliability of the measure, with the same trends across experiments being observed (Fig. 2.3-E and Fig. 2.4-C).

A limitation of this data resides in the granularity of the working definition. Indeed, micro-pursuits should theoretically be treated as a class of FEM, at the same level as micro-saccades and drifts. However, to link the theoretical definition to our working definition, the experimental data would need to be recorded with more precise and more accurate systems than video-based eye-trackers (Wyatt, 2010; Choe et al., 2016). Hence, it is an invitation to research further into this oculomotor phenomenon. We provide a basic field guide to pursue investigation and characterisation of such a phenomenon, by providing a key metric, namely MPC which measures similarity of the gaze trajectory with respect to that of a target.

Eye movement research is gradually considering an *oculomotor continuum*. For instance, it is becoming less and less credible to consider a hard separation between micro-saccades and saccades because of their common neural origins in oculomotor programming (Krauzlis et al., 2017), their common properties, and mathematical models that can account for both (Sinn and Engbert, 2016). One may thus also consider that large amplitude smooth-pursuits share physical properties as well as neural correlates with micro-pursuits. Further work is needed to discriminate them from other FEM, like micro-saccades, and to develop detection algorithms, which will need to cleverly combine physical characteristics with a robust estimation of similarity between gaze and target based on very short episodes.

2.6.2 Micro-saccades define a main sequence

Micro-saccades, a class of FEM, have been characterised by (i) their ballistic properties—like saccades,—(ii) their small amplitudes, and (iii) the relationship between their peak velocity and amplitude. This resulted in the definition of a main sequence (Bahill et al., 1975). The latter stipulates that as micro-saccades have larger amplitudes, their associated, measured peak velocity will increase, and this relationship is linear. Indeed, this can be explained by the fast, ballistic nature of the micro-saccade, which, in essence, re-positions the fovea in the context of visual perception (Rolfs, 2009; Ko, Poletti, et al., 2010; Poletti and Rucci, 2016; Sinn and Engbert, 2016), similar to saccades at larger scales (i.e., not contained within FEM). The execution of this rapid movement is typically over a short period, under 80 ms. The physical properties of the oculomotor system constrain these ballistic motions of the eye to exhibit the linear velocity-amplitude relationship, characteristic of the classic main sequence.

The main sequence has been very reproducible, and appears in over decades of eye movement research (Rolfs, 2009; Hicheur et al., 2013). It has been used to develop robust micro-saccade detection algorithms such as the one proposed by Engbert and Kliegl (Engbert and Kliegl, 2003), which is the one also used in this work. Their detection is based on a lower relative velocity threshold computed from a sliding window—such that the detection threshold is dependent on the contextual oculomotor activity. Using this detection method, our data presented in Fig. 2.2 shows a clear secondary main sequence under the predictable motion condition (LJ) in both Necker and Cross experiments.

The detection of slow micro-saccades in our data set with this algorithm can be explained by the dependency of the detection algorithm on the fixations' mean velocities, and the adaptive threshold based thereupon. The conditions created by the use of small amplitude predictable Lissajous trajectories on an ambiguous Necker Cube on one hand (Necker), and an explicit pursuit task on the other hand (Cross), could explain the detection of *slow micro-saccades*.

This interpretation suggests that the classic main sequence can be composed of erratic micro-saccades such as under FX or act as catch-up micro-saccades for a micro-pursuit movement just as we might have at a macroscopic scale with smooth pursuit movements. For further understanding of this phenomenon, a more detailed break down of FEM data into gaze before and after micro-saccades is needed, which we consider outside the scope of this work.

2.6.3 Micro-pursuits might define a secondary sequence

Our graphical results in Fig. 2.2 suggest that we do have a secondary sequence under the LJ condition in both the Necker and Cross experiments. Moreover, detected micro-saccades that give high MPC scores are mainly associated with the secondary sequence and vice versa (see Fig. 2.2-A). Under the RW condition, however, the secondary sequence is less pronounced. Thus, a contrast can be observed between predictable and unpredictable motion conditions, suggesting that detected micro-saccades in FEM are more diverse than initially assumed. We propose to focus on this secondary sequence, i.e. the slow main sequence.

The differences in rate shown in Fig. 2.2-B also suggest that the secondary sequence may represent detected micro-saccades of different nature than those contained in the classical main sequence. Given the recent and older observations of small amplitude pursuit in monkey (Skinner et al., 2018) and human data (Heywood and Churcher, 1971; Heywood and Churcher, 1972; Martins et al., 1985; Poletti, Listorti, et al., 2013), as well as the contrast between unpredictable (RW) and predictable (LJ) stimulus motion in our experiments, a credible hypothesis is that the secondary sequence mainly covers an additional class of FEM, namely micro-pursuits. Alternative hypotheses might classify these as ocular drift (C.-Y. Chen and Hafed, 2013; M. Watanabe et al., 2019) or slow motor control (Kowler and Steinman, 1979a; Kowler and Steinman, 1979b; Kowler and Steinman, 1981; Kowler, Martins, et al., 1984; Epelboim and Kowler, 1993), although these do not track on a target (stimulus) and would thus contradict with the aforementioned correlation observed between membership of the secondary sequence and high MPC scores.

2.6.4 Micro-pursuits could be detected through similarity with a target

The MPC measure proposed to quantify similarity between two multi-variate signals—such as the bi-variate gaze and stimulus trajectories in the context of pursuit and micro-pursuit,—can be subjected to a finer analysis. It features robustness with respect to (i) additive noise degradation, (ii) scale, and (iii) spatial offset and translation, making it convenient for a study of similarity. It's limitation resides mostly in its variance and thus the number of (temporally correlated) samples needed to accurately measure similarity. This is illustrated through Fig. B.1 in Appendix B.1. While on the one hand, physical properties (amplitude, peak velocity, see discussion on the *secondary sequence*) can be used to discriminate micro-pursuits

from micro-saccades, on the other hand, functional characterisation will help provide discrimination between drift, slow motor control, and micro-pursuit. Indeed, the first two may be slow FEM, but have no requirement for target tracking, like pursuit, whereas the latter does (Martins et al., 1985; Spering and Montagnini, 2011). Furthermore, its link to visual perception remains speculative, though interpreting our data suggests that attention may lead to a tentative explanation (Spering and Montagnini, 2011).

2.6.5 Influence of attentional context on target locking

The three experiments (Necker, Cross and Square) presented show three different contexts of tasks and stimuli. In the Necker experiment, the ambiguous Necker cube was subject to motion and participants had to report how they perceived the cube's orientation while fixing a central cross. The context was challenging as perceptual changes were endogenous. The two tasks—oculomotor and perceptual—required a split of attention as one had to focus gaze on a central fixation while observing a moving, ambiguous cube (VanRullen, Reddy, et al., 2004). We consider this a difficult attentional context, which forces the observer to split attention between to elements, or perceptual objects, of the visual field, that can thus be considered competing (attentional) attractors. There have been previous reports showing that an ambiguous stimulus can reveal attentional modulations (Kohler et al., 2008). In the Cross experiment, we created an explicit context, where the fixation cross was moving as it underwent illumination changes. Participants were asked to fixate the cross, while a static, unchanging square remained in the background. One can consider that participants had to focus all their attention on the cross, and as it was moving in a predictable, tractable fashion (LJ), the cross induced micro-pursuits. In the Square experiment, the attention had to be split, like in the Necker experiment. However, the square was unambiguous, and its changes required less attention to detect. Thus, one can consider the Square experiment to have attention split between two attractors, as well, but given the results obtain (Fig. 2.3-E-F) and the lack of observed micro-pursuits, one can interpret the competition between attractors as unbalanced. This approach considers attention qualitatively, based on the manipulation of tasks and stimulus motion, but a more quantitative approach would provide a better view on the possible interactions—e.g. by means of an efference copy (Astrand et al., 2015)—of the oculomotor and attentional systems for micro-pursuits.

2.6.6 Modelling attention to generate gaze patterns

A first step towards a quantitative characterisation of how attention may influence oculomotor dynamics is proposed in Chapter 3, by a competing attractor model based on gravitational-like field model. The model simplifies the visual stimulation by considering perceptual objects as gravitational attractors with dynamically varying masses modelling the attention whereas gaze position is modelled through a unit-mass particle evolving in time subject to the gravitational field, and subject to additive velocity noise (Langevin dynamics). By manipulating the attractor's positions, masses, and the curvature of their *energy potential*, it is possible to generate (micro-)saccades, (micro-)pursuits, fixations, and drift. This mathematical model offers a quantitative method that may be interpreted in terms of spatial attentional loads with respect to oculomotor programming and execution. It is an extension of some models already proposed in the field of FEM modelling based on energy potential (Engbert, Mergenthaler, et al., 2011; C. Herrmann et al., 2017) as well as modelling work on bi-stable perception and processes (Moreno-Bote, Rinzel, et al., 2007; Shpiro, Moreno-Bote, et al., 2009; Moreno-Bote, Knill, et al., 2011; Moreno-Bote and Drugowitsch, 2015), to incorporate the influence of, e.g., ambiguous figures like the Necker cube.

2.6.7 Future works

Further studies based on quantitative approaches to spatial attention (Corbetta et al., 1998; Cavanagh and Alvarez, 2005; Engbert, Trukenbrod, et al., 2015; Gide, Karam, et al., 2017; Esterman and Rothlein, 2019), combined with the use of MPC on the gaze and salient points of the stimulus—rather than the simpler centre of gravity –, should constitute promising methods to investigate the relationship between attention and micro-pursuits.

Our model, as well as existing and competing models, feature stochastic processes (Gammaitoni et al., 1998; Kim et al., 2006; Engbert and Kliegl, 2004; J. Braun and Mattia, 2010) which could provide insights to further understand how FEM act as noise. The question remains on how that noise might impact perception, such as through hyper-acuity phenomenon (Rucci, Iovin, et al., 2007; Zozor et al., 2009; Rucci and Victor, 2015). An extension of the presented modelling approach in Chapter 3 incorporating additional, competing attractors living in a complex, possibly high-dimensional, perceptual space coupled with the already existing oculomotor model, constitute a framework to study the interaction between

oculomotor, attentional, and visual systems in a goal-oriented complex system like the brain (Kelso, 2012; Schwartz et al., 2012).

2.7 Conclusion

In this work, micro-pursuits are proposed as a type of FEM occurring at small amplitude, within a fixation, as the gaze follows a target. We proposed two metrics: inertia and MPC to measure gaze displacement within a fixation and to quantify gaze-target trajectory similarity, respectively. When searching for micro-saccades, our data showed the presence of a secondary sequence, contrasting with the well-known main sequence exhibited by micro-saccades. Detected micro-saccades that belonged to this secondary sequence showed lower peak velocities as well as higher similarity with the target, which has led us to classify these movements as micro-pursuits. Upon further inspection of the data, both the Necker experiment and the Cross experiment showed fixations with high similarity values under predictable target trajectories. Micro-pursuit here is presented as a class of fixation, but further research is needed to identify the physical properties and distinguish it from other FEM. Moreover, this work calls for further investigation on the functional role of micro-pursuits, and how the oculomotor and perceptual systems interact during such movements.

3 Modelling eye movements & multi-stable perception

> *Between my consciousness and my body as I experience it, between the phenomenal body of mine and that of another as I see it from the outside, there exists an internal relation which causes the other to appear as the completion of the system. The possibility of another person's being self-evident is owed to the fact that I am not transparent for myself, and that my subjectivity draws its body in its wake.*
>
> **— Maurice Merleau-Ponty**
> **"Phenomenology of Perception"**
> **1945.**

Eye movements and multi-stable perception have been further understood and deciphered with the study of mathematical models capable to produce analogous behaviours. Models, however, offer a key advantage: they can be manipulated and understood in fine details through theoretical studies and numerical simulations. In this thesis, we proposed a model based on gravitational energy potentials to generate eye movements. This approach is described in this chapter by presenting numerical empirical results. Furthermore, perceptual tri-stability was studied in order to begin an investigation of how one can expand some of the results found for bi-stability, and generalise them for multi-stability. The proposed framework is a first step towards the construction of formal models that bind perception and action; here, in particular, ambiguous perception and oculomotricity.

Publication.
The work presented in this chapter is exploratory and partially completed. Its aim is to provide the reader with some understanding of the theoretical framework used to develop the experimental work in the following chapters.

Contents

3.1 Gravitational fixational eye movements

Models come in a variety of forms, depending on the mathematical framework used to formalise and compute their mechanics. Two main families can be differentiated: descriptive statistical and generative mechanistic models. Here, we focus on the latter. The motivation is the following: generative models can produce simulated and synthetic results that can be compared to observed empirical data. The model can then be studied and decomposed such that each internal force can be characterised, and their functional role in creating the analogous behaviour can be investigated. All together, models remain key to understand a phenomenon and make predictions for empirical and experimental work. We focused here on FEM in an attempt to explain and understand the data observed and reported in Chapter 2. Eye movements are modelled by various techniques.

Probabilistic modelling - scan-paths

For instance, gaze data in scene exploration task, also known as *scan paths*, have been studied through probabilistic models of spatio-temporal dynamics (Marat et al., 2009; Ho-Phuoc et al., 2009; Tatler et al., 2011). These models are often based on *saliency maps*, which draw the empirical two dimensional distribution of fixation location convoluted with a spread function (Gide, Karam, et al., 2017), and have provided researchers with a better understanding of natural image's statistics for the perceptual, oculomotor and attentional systems. Scan path models add a temporal dependency to saliency maps, and allow the generation of synthetic gaze patterns linked to an image or video's observation. However, they often neglect the role of intention and task (see Fig. 1.1 in Chapter 1), as they are estimated bottom-up from gathered data.

Accumulation processes - saccade generation

Another known approach is derived from the field studying decision making in which, the cognitive system is considered to accumulate evidence or information up to a threshold. When the threshold is reached, a decision is made. For eye movements, the decision is materialised by the execution of a saccade, for instance (Orquin and Loose, 2013). Accumulation processes are implemented by *drift diffusion* models, in which, a particle influenced by a deterministic force, e.g., the evidence, and a stochastic force, e.g., the internal noise, drifts in a space until a threshold is

reached. This type of models are also close to *integrate and fire* neuronal models, in which information is accumulated, in weights that correspond to dendrites, until a threshold is reached and synaptic spiking occurs (Gerstner et al., 2014). This approach is effective to understand how intentions can be linked to a decision, and subsequently to an action. However, they often neglect the role of attention and the stimulus' statistical properties.

Energy potential models

Some research teams have attempted to merge both paradigms by proposing models using energy potential models. In fact, FEM phenomenology offers an interesting perspective for such unification because FEM have historically been considered as unconscious ocular events linked to the visual stimulus on the fovea (Thaler et al., 2013), but is also known to reflect conscious perception, attention and expectation (Kowler, Martins, et al., 1984; Laubrock et al., 2008). Recently, Engbert and colleagues (Engbert, Mergenthaler, et al., 2011) proposed a generative model that could reproduce the statistical properties of FEM stationary displacement, namely the short term persistence and long term anti-persistence of drift and tremors. They used a self-avoiding walk (Freund and Grassberger, 1992) in a discretised quadratic energy potential: at each iteration, the gaze, represented by a particle in the energy potential landscape, can either go left, right, up or down. The walker will choose the slot with the lowest energy. Once a step is made, the slot of the previous iteration is set to a high energy value, and the entire energy landscape follows a linear relaxation law. Hence, FEM bottom-up dynamics can be reproduced. Furthermore, the model also proposed to integrate micro-saccade generation by a threshold rule: when the particle is surrounded by options with energy higher than the threshold, it jumps to the global minimum of the energy landscape. Here, the authors provide an accumulation process linked to a global integration of the oculomotor field.

The integrated FEM model described above is a key foundation to bridge the oculomotor modelling communities and accounts for multiple FEM phenomena (e.g., drift displacement, micro-saccade, spatial orientation biases). However, it did not possess a mechanism to account for micro-pursuit, as these are hardly studied and reported. The observation of micro-pursuits presented in Chapter 2 implies that the dynamics of the gaze within a fixation can be affected and attracted by motion of a perceptual object in or nearby the foveal field. Therefore, we propose modelling approach, gravitational fixational eye movements (GraFEM), inspired by gravitational energy field theory to model motion of eye movements (Parisot, Chauvin, Guérin-Dugué,

et al., 2017; Parisot, Chauvin, Phlypo, et al., 2018) and derived from the work on integrating FEM in energy potential models (Engbert, Mergenthaler, et al., 2011).

3.1.1 Gravitational potential energy field modelling

Integrated and generative FEM models make use of energy potentials to generate self-avoiding walks, constrain the walks and replicate oculomotor biases (Engbert, Mergenthaler, et al., 2011). In fact, the latter is used to constrain the pseudo random walk's spatial horizon. Furthermore, it can be considered as an attractor of the energy landscape. Thus, the use of the particle in an energy potential framework can be adjusted to provide biases of the stimulus on the FEM generation. Combining attractors in the energy fields, that increase the probabilities of having the gaze at some spatial coordinates, and adding stochasticity to the movement of the particle can provide a simple mechanism for FEM generation.

Attractors.
Attractors in energy landscape are local minimums into which a particle will roll by gradient descend, i.e., gravity. However, if one considers the problem from a probabilistic point of view, energy landscape correspond to negative inverted probability distributions, with minimums corresponding to modes with higher probabilities. The advantage of considering the gaze dynamics as a physics problem is that it offers an intuition of the dynamics when the attractors' masses and positions are known, since humans all have an intuitive understanding of Newtonian gravity by experiencing life on earth.

Model description

The attractors' properties can be manipulated over time to affect the energy field and thus dynamics of the FEM generated. The energy field that is mapped to the visual field can be populated by an arbitrary number of n attractors of varying strength (see Fig. 3.1a). Inspired by the formalism of gravitational fields, one can generate fields with the following equations. Let Φ_i represent the field generated by the i^{th} attractor given by:

$$\Phi_i(\boldsymbol{q}, t) = -\frac{1}{\|\boldsymbol{q}(t) - \boldsymbol{a}_i(t)\|^{2\beta_i(t)} + \delta_i(t)} \tag{3.1}$$

with $\boldsymbol{q}$ and $\boldsymbol{a}_i$ corresponding to the spatial x-y coordinates (at time t) of the observer's gaze position and the i^{th} attractor, respectively. The potential landscape can be fine tuned according to assumptions on attentional attributes of the stimulus and the tasks. First, it is necessary to set how many attractors are present and give them

spatial coordinates in the plane over time. Secondly, it is possible to handle the mass of those attractors and their subsequent force of attraction and distortion of the field by tuning two parameters; δ for the depth of the well and β for the concavity of its slope. Summation and normalisation of the field allow for the fusion of the multiple attractors.

$$\Phi(\boldsymbol{q}, t) = \sum_{i=1}^{n} \Phi_i(\boldsymbol{q}, t) \tag{3.2}$$

A logarithmic attenuation is added to allow the possibilities of exploring high energy areas of the visual/foveal field, giving the energy E:

$$E = -\ln(-\Phi) \tag{3.3}$$

Memory of attractor motion (Fig. 3.1b) are modelled by adding a moving average (MA) process (Hannan, 2009) on the field at a given time t:

$$E_{FEM}(\boldsymbol{q}, t) = E(\boldsymbol{q}, t) + \sum_{k=1}^{K} \frac{\lambda}{k+1} E(\boldsymbol{q}, t - k\Delta t) \tag{3.4}$$

where K is the temporal parameter limiting how far in time will the fields be summed over and with λ the relaxation rate parameter and Δt is the temporal step size. It is also possible to set the impact of memory and anticipation through parameters that define the iteration window over which the field is deformed using traces of the attractor in the past of a given current iteration and the rate λ at which the deformation affects for a given lag.

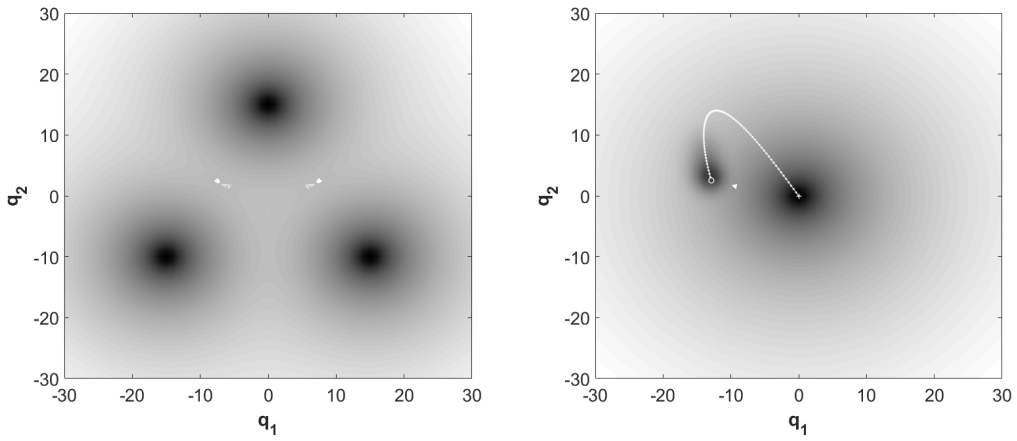
A particle of position ($\boldsymbol{q}$) with negligible mass (or with very high friction) is dropped in the field and is disturbed by noisy force, in order to generate and simulate gaze dynamics. Therefore, given the fundamental relation for dynamics, where the accelerating second order component is neglected, the gaze particle's motion is derived by the Langevin equation (Langevin, 1908), in which $m\ddot{\boldsymbol{q}}$ is equal to the sum of forces applied to the particle, and can be rewritten as follows:

$$m\ddot{\boldsymbol{q}} = -\gamma\dot{\boldsymbol{q}} - \nabla E_{FEM}(\boldsymbol{q}, t) + \xi(t) \tag{3.5}$$

with m the negligible mass, γ the friction and where ξ is an external force, here an oculomotor noise (η) applied to the gaze, such that $\eta(t) = \frac{\xi(t)}{\gamma}$. With the assumption of low mass and after normalisation[1], such that $E_{FEM} = \frac{E_{FEM}}{\gamma}$, the dynamics can be expressed as:

$$\dot{\boldsymbol{q}} = -\nabla E_{FEM}(\boldsymbol{q}, t) + \eta(t) \tag{3.6}$$

[1]Note that in the next equation, we use E_{FEM} with the same notation as above, which is not exact writing though it simplifies reading. We refer to a normalised term by γ in the next equation.



(a) GraFEM with 3 attractors.

(b) GraFEM with 2 attractors.

Figure 3.1. Examples of energy landscape surface plots setup using the GraFEM model for:
(a) 3 attractors ($n = 3$) with all attractors i having no motion and the following parameters: $\beta = 2;\ \delta = 1;\ K = 5;\ \lambda = 0.9$.
(b) 2 attractors ($n = 2$) with all attractors i having the following parameters: $\beta_1 = 2;\ \beta_2 = 4;\ \delta_1 = \delta_2 = 1;\ K = 15;\ \lambda = 0.9$ and attractor motion computed with the following arbitrary sinusoidal motion: $\boldsymbol{a}_1(t) = (0, 0);\ \boldsymbol{a}_2(t) = \boldsymbol{a}_2(0) + (-5\sin(2t), 5\sin(3t))$, with a time step of $\Delta t = 13$ ms. The figure is extracted from on the 75$^{\text{th}}$ time step corresponding to the 975 ms into the simulation. The motion of $\boldsymbol{a}_2$ is shown in white, with $\boldsymbol{a}_2(0) = (0, 0)$.
Though the model has many parameters, those manipulated in this work's results are exclusively the depth δ (or mass) of the attractors and the slope β by affecting the concavity of the attractors' field. White spots between attractors are rendering errors by the visualisation method.

The evolution of the gaze particle's dynamics can be computed by making the problem a discrete one using the Euler-Maruyama method (Kloeden and Platen, 2013), for instance.

3.1.2 Model simulations: what are the parameters corresponding to ocular events & interpretation?

Fixations of 3.5 seconds, with a discretisation Euler-Maruyama step $\Delta t = 13$ ms equal to the time step, were simulated using the GraFEM model with two attractors, $\boldsymbol{a}_{cross}$ corresponding to the attractor of a fixation cross at the centre and $\boldsymbol{a}_{stim}$, the attractor representing the stimulus, with a Lissajous motion: $\boldsymbol{a}_{stim} = (\sin(2t), \sin(3t))$. Only the slope and depth parameters were manipulated: $\beta_{stim} \in [0; 15]$ and $\delta_{stim} \in [0; 500]$. All other parameters were kept constant with the other attractor position at $\boldsymbol{a}_{cross} = (0, 0)$ with $\beta_{cross} = 1$ and $\delta_{cross} = 100$, the relaxation rate parameter $\lambda = 0.9$, the memory temporal limit $K = 5$ and noise $\xi \sim \mathcal{U}[-0.5; 0.5]$. These

simulated fixations were then analysed using the measures presented in this article, namely, inertia, MPC and micro-saccade detection using the EK algorithm based on relative velocity thresholds (Engbert and Kliegl, 2003). Fig. 3.3a shows that higher inertia follows a diagonal region along the $\{\beta_{stim}, \delta_{stim}\}$ space. When looking at Fig. 3.3b, one can see that the same area in the parameter space has systematically high MPC. Finally, the EK algorithm was applied (without the binocularity criterion) to measure detected micro-saccades, and summed over the time of a fixation. The results (Fig. 3.3c) show that micro-saccades are detected when concavity is high due to a larger β_{stim} parameter.

3.1.3 Discussion and perspectives: attentional, oculomotor and perceptual multi-stability

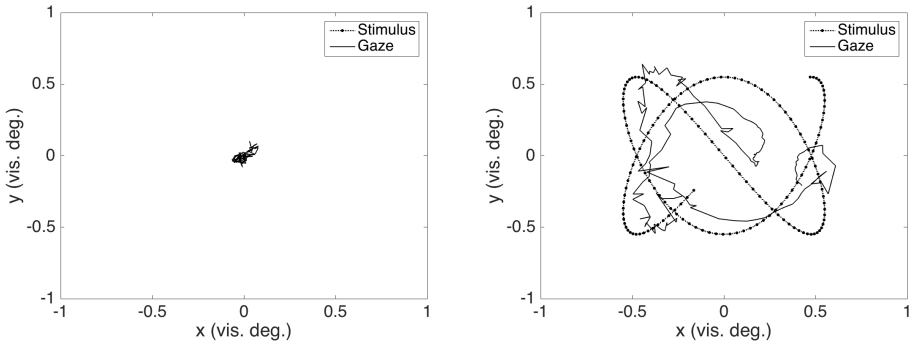
The simulation results presented above show:

- Fixations' dynamics can be modelled including a variety of FEM such as drift, tremors, micro-saccades and micro-pursuits.
- Attractor dynamics can be intuitively manipulated by two parameters that control their slope and depth, hence imposing, by gravity, faster or slower dynamics on the gaze-particle.
- Generalisation to more complex stimuli or tasks can be manoeuvred by such a model as attractors can be multiplied, if necessary.

However, this work remains preliminary and calls for further investigation. Such perspectives are discussed in the following paragraphs.

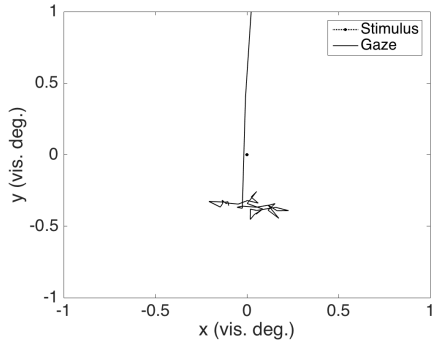
Model interpretation for eye movements

The GraFEM model proposed in this paper is capable of generating micro-saccade, drift and tremor FEM (see Fig. 3.2) as classified in the literature (Martinez-Conde, Macknik, and D. Hubel, 2004) as well as the micro-pursuits presented and detected in Chapter 2 as reported in Fig. 3.3 & Fig. 3.4. By using classified data (observations), the parameters of the model that allow the generation of these FEM could be inferred, and insights on the mechanics of micro-saccades, micro-pursuit, drift and tremor generation and their interaction can be studied. The diagram Fig. 3.4 already gives a useful and overall understanding of the model, with respect to the manipulated



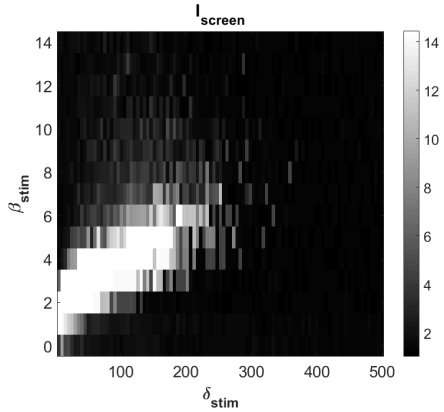
(a) Example of a simulated fixation with GraFEM.

(b) Example of a simulated micro-pursuit with GraFEM.

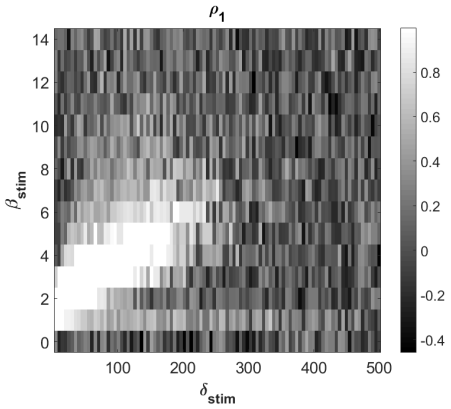


(c) Example of a simulated micro-saccade with GraFEM.

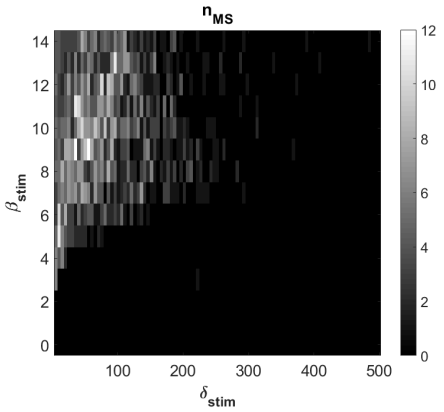
Figure 3.2. Simulation examples generated with the GraFEM model.
Simulations of fixations of 3.5 seconds with Euler-Maruyama time steps of $\Delta t = 13$ ms, with variable fixation dynamics generated through by manipulating of δ_{stim} and β_{stim} parameters. Constant parameters of the model were: number of attractors $n = 2$, with one for the fixation cross $\boldsymbol{a}_{cross} = (0, 0)$ and another for the motion of the stimulus following a Lissajous trajectories with the same parameters as in the three experiments from Chapter 2; $\boldsymbol{a}_{stim}(t) = \big(\sin(2t), \sin(3t)\big)$. The relaxation rate parameter $\lambda = 0.9$, memory temporal limit parameter $K = 5$ and noise $\xi \sim \mathcal{U}[-0.5; 0.5]$ were used.
(a) A simulated fixation with stable fixation dynamics with $\delta_{cross} = 100;\ \delta_{stim} = 100;\ \beta_{cross} = 1;\ \beta_{stim} = 1$.
(b) A simulated fixation with micro-pursuit dynamics with $\delta_{cross} = 100;\ \delta_{stim} = 25;\ \beta_{cross} = 1;\ \beta_{stim} = 1$.
(c) A simulated fixation with micro-saccade dynamics with $\delta_{cross} = 100;\ \delta_{stim} = 25;\ \beta_{cross} = 1;\ \beta_{stim} = 12$ and detected using the EK algorithm for micro-saccade detection.



(a) Inertia over the GraFEM simulations.



(b) Similarity (MPC) over the GraFEM simulations.



(c) Micro-saccades detected over the GraFEM simulations.

Figure 3.3. Simulations and analyses of the GraFEM model.
Simulations of fixations of 3.5 seconds with Euler-Maruyama time steps of $\Delta t = 13$ ms, with variable fixation dynamics generated through the variation of $\delta_{stim} \in [0; 500]$ and $\beta_{stim} \in [0; 15]$ parameters. Constant parameters of the model were: number of attractors $n = 2$, with one for the fixation cross ($\boldsymbol{a}_{cross} = [0, 0]$) and another for the motion of the stimulus following Lissajous trajectories with the same parameters as in the Necker cube experiment (Chapter 2): $\boldsymbol{a}_{stim}(t) = \big(\sin(2t), \sin(3t)\big)$. The relaxation rate parameter $\lambda = 0.9$, memory temporal limit parameter $K = 5$ and noise $\xi \sim \mathcal{U}[-0.5; 0.5]$ were used.
(a) Behaviour of inertia over the parameter space of the GraFEM model.
(b) Behaviour of the similarity between stimulus and simulated fixation motion using the MPC ρ_1.
(c) Number of micro-saccades detected by the EK micro-saccade detection algorithm.

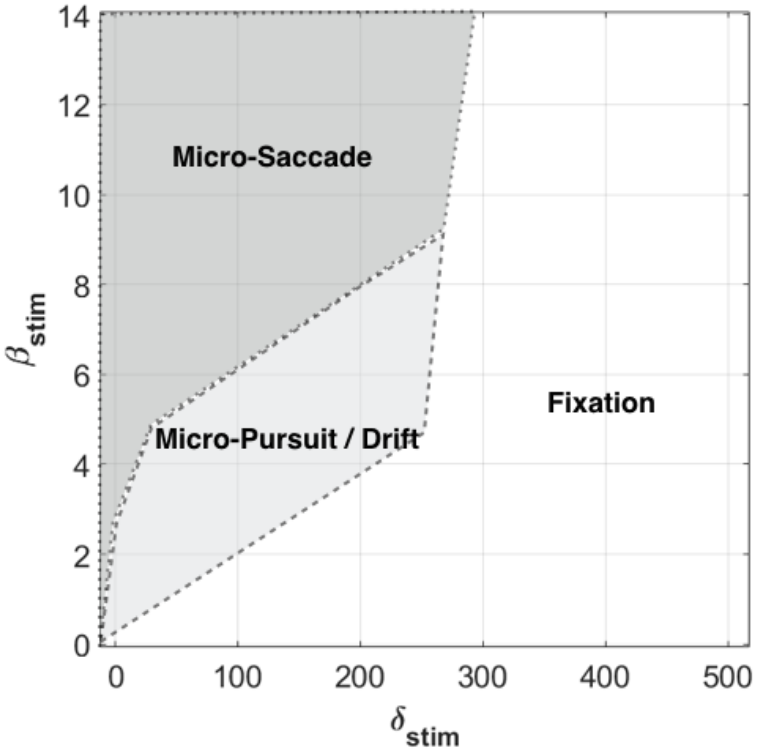


Figure 3.4. GraFEM oculomotor interpretation. Schematic interpretation of oculomotor dynamics generated in the parameter space of the GraFEM model with manipulation of the δ and β parameters of the stimulus attractor, while keeping all other parameters constant.

parameters, but the work on parameter inference should be addressed in a near future in more details.

Given the observed data and the proposed model to account for it, questions and perspectives can be redefined with a novel angle for interpretation of fixational eye movements. Inversion and a full analysis of a model, like GraFEM, with multiple free parameters is a complex task out of the scope of this thesis but should be tackled and reported in a near future.

The model presented here gives a mathematical framework in which eye movement phenomena can be generated and interpreted. Attractors are interesting as tools to explain and interpret cognitive and physiological behaviours as they allow an intuitive understanding of the evolution of dynamical systems (T. Watanabe et al., 2014; Kelso, 2012). Furthermore, complex learning systems—i.e., neural networks—are known to develop such properties as the parameters of their processes tend to learn the statistics of the environment by creating attractors in the parameter space (Moreno-Bote, Rinzel, et al., 2007; Shpiro, Moreno-Bote, et al., 2009; Moreno-

Bote, Knill, et al., 2011; Moreno-Bote and Drugowitsch, 2015). With this modelling framework, the FEM classification of the literature can be described and interpreted in terms of attractor spatio-temporal dynamics (Fig. 3.3 & 3.4).

A stable fixation (Fig. 3.2a) in the GraFEM model corresponds to a stabilisation of an attractor with the energy landscape having little change. The gaze-particle is stuck and only the noise affecting its position may lead to small random movements of the eyes, as in other generative FEM models (Engbert, Mergenthaler, et al., 2011; C. Herrmann et al., 2017). In these models, constraints to the energy field of the fixation are used in an analogous fashion to reflect the higher probabilities of having FEM in horizontal and vertical directions. A fixation attractor can thus be predicted by the task or the stimulus controlled by the experiments, and its parameters can be inferred by *a priori* information and data. Hence, the model gives predictive capabilities that can be tested and requires assimilation of data to constrain its range of possibilities.

Micro-saccades (Fig. 3.2c) correspond to sudden changes in the energy depth of attractors, with a new one emerging or deepening while the attractor of fixation has suddenly disappeared. They are likely to emerge as the gaze-particle rushes down a gradient to the centre of an attractor, giving it sufficient velocity. The depth and slope of the attractor can be manipulated (following the dynamics described in Fig. 3.4), thus making it possible to infer, based on observed velocities and amplitudes, the saliency of that attractor. The GraFEM model does not use an explicit and separated mechanism for micro-saccade generation—as the model presented in Engbert, Mergenthaler, et al. (2011)—though it is not incompatible.

Drifts correspond to a stability of the gaze-particle with respect to the attractor by which it is transported. However, the attractor might itself slowly drift away in the visual space (independently from the target motion) or alternatively, the shape of the well might get larger (by manipulating the parameter β), allowing for the noisy gaze-particle to explore further. These are two hypotheses that could be tested, in future work, by inferring the model parameters given sufficient data. These FEM are known to help reduce visual redundancy and extract features in complex visual stimuli (Kuang et al., 2012) but are mostly considered to be consequences of the eye muscles and their neural control properties. Therefore, they have mostly been considered as independent processes from the visual stimulus presented. The micro-pursuits detected and described in Chapter 2 could be interpreted as a form of stimulus related drift, as its signal dynamics place it in similar ranges, and is capture by the proposed metric; namely MPC. Consequently, this argues in favour of our proposition that drifts are composed of two categories—stimulus independent

and dependent—and micro-pursuits logically fall within visually dependent ocular drifts. This dependency can be interpreted as the interference of bottom-up salient elements interrupting the top-down task of fixation.

Micro-pursuits (Fig. 3.2b) are therefore close to drifts in the energy landscape dynamics.

Model mechanics

The model sets the gaze as a particle in an evolving gravitational energy potential field. When the system has no dynamics added to the potentials' landscapes, the particle will fall into its nearest local minimum. In this implementation, at each iteration—here a discrete time step using Equation (3.6) –, the first derivative is computed to update the position of the particle in the plane, corresponding to the screen. Noise is then added to the deterministic dynamics and can drive fixational oculomotor decision-making with respect to attractors if its amplitude is sufficiently large (Shpiro, Moreno-Bote, et al., 2009; Moreno-Bote, Rinzel, et al., 2007). This mechanism is similar to bi-stable energy potential models, though it extends on the dimensions of the system. In a set of simulations reported in Fig. 3.3, we show that through two continuous parameters applied to a target attractor, it is possible to generate and interpret oculomotor dynamics observed in FEM. However, here, there is no prior requiring the existence of different systems for each class of movements observed (Liversedge et al., 2011). FEM dynamics can be reproduced through a unique mechanism as shown by the simulated examples in Fig. 3.2.

Top-down intention processes can be tested and simulated, given the context of a task, by applying changes in the model's β and δ parameters. Fig. 3.4 can be used as a road map of the oculomotor dynamics and regimes expected, depending on parameter values. Moreover, bottom-up attentional effects can also be taken into account. This can be done with simpler assumptions, such as the ones presented here for the task used in Chapter 2, but can be more complex if using natural scene tasks, for instance. An interesting and practical perspective in this context lays in investigating how salience models, which derive probability distribution based on the statistics of images or videos, can be integrated such that only attractors are fed into a GraFEM oculomotor execution system.

How would this be implemented in the brain?

Anatomically, oculomotor programming has been shown to be highly correlated and linked to a network of areas involving neural activity in the SC, the FEF and the LIP cortex (Hafed, Goffart, et al., 2009; Krauzlis, 2004; Krauzlis, 2005; Krauzlis et al., 2017; Astrand et al., 2015; Peel et al., 2016; Taouali et al., 2015). There are inter-individual differences in anatomy and behaviour for fixational eye movements measuring and observed dynamics. For instance, it has been shown that not only oculomotor behaviour between trained and untrained participants vary a lot, but that drift accounts for more fixation correction motion than micro-saccades (Cherici et al., 2012). The observations of micro-pursuits presented in Chapter 2 suggest that the dynamics of the gaze within a fixation can be affected and attracted by motion of an object in or nearby the foveal field.

However, rather than having an attractor with a pseudo-random displacement, its motion follows a deterministic and predictable trajectory, that can be computed and estimated by the oculomotor system. Moreover, that attractor is, given our observations so far, only related to a target motion. This could, for instance, be implemented in the brain by the means of an efference copy (Astrand et al., 2015), though this idea remains speculative and further modelling and neuro-physiological research is needed. The low energy attributed to a decoded and perceived object moving across space encourages the oculomotor system to track it as it tries to minimise the energy of the gaze-particle. Finally, tremors are generated and explained by the noise given to the particle over all FEM events.

This model complements the eye movement field of research with the possibility to program intentions, salience, and their effects on the gaze dynamics by simply using attractors and setting out their dynamics in terms of motion on the visual field, depth and memory. For instance, the model can predict the different dynamics reported based on the eccentricity of an attractor corresponding to an afterimage, as observed in (Heywood and Churcher, 1972). Thus, one can use the model to generate statistical predictions of eye movement dynamics. Given an understanding of the visual attention effects of their stimulus and take into account all the associated intentions to the tasks that participants are required to be operated during a trial, it is possible to use this modelling to generate quantitative predictions on the oculomotor dynamics. Moreover, the generative properties makes it possible to work on simulated data and extract dynamics' statistics in terms of eye movements, and this is possible using the traditional algorithms for eye movements classification. Inversely, obtaining the parameters of the model that replicate the dynamics of

observations could help understand better the internal processes that drive eye movements.

3.1.4 Perspectives: towards oculomotor multi-stability.

A key aspect of this family of models is that it showcases *multi-stability* regarding their attractors. This phenomenon can emerge in many complex biological systems and is present in many cognitive processes (Schwartz et al., 2012). It is linked to coordination dynamics between sub-systems which have varying levels of coupling, leading to mono-stable, multi-stable or meta-stable dynamics (Kelso, 2012). The consequent interpretation is that the oculomotor system could have multi-stable dynamics with respect to visual attractors. In this case, the oculomotor dynamics are likely driven by noisy signals (J. Braun and Mattia, 2010) representing other interfering systems, such as perception, attention, and other cognitive systems. This framework connects to the growing body of studies linking perceptual decisions and multi-stable system dynamics. It also creates a link for motor systems to studies of noise as a component that helps a perceptual system operate through stochastic resonance[2] (Gammaitoni et al., 1998; Patel and Kosko, 2005; Kim et al., 2006).

[2]Stochastic resonance refers to phenomena in which a system is able to detect a weak signal because noise boosts it, by providing the energy needed for the signal's frequencies to resonate mutually.

3.2 Multi-stable perception

As suggested by the conclusion of the previous section, the GraFEM models can be interpreted as multi-stable oculomotor models. One of the aims of this thesis was to propose a framework in which both action and perception can be linked, in order to investigate how visual perception, when stimulation is ambiguous, is dependent and constrained by active oculomotor behaviour. Moreover, works on modelling multi-stable perception has been extensively investigated for two state systems, e.g., bi-stability, (Shpiro, Curtu, et al., 2007; Shpiro, Moreno-Bote, et al., 2009; Moreno-Bote, Knill, et al., 2011) but less is known for states with three or more quasi-stable states. This section uses the framework and model developed above to look at perceptual multi-stability.

3.2.1 Extension for multi-stability

An advantage of the GraFEM model's formalism is that it can gain in dimension for its energy space and the number of attractors quite easily. Indeed, Equation (3.1) can be summed and normalised over n attractors in a two-dimensional space. But if $\boldsymbol{q}$ and $\boldsymbol{a}_i$, the coordinates vector of the particle and the i^{th} attractor, respectively, are considered as two-dimensional vectors with (x, y) coordinates on the screen plane for GraFEM, it can be extended to p-dimensional vectors in the perceptual space. We then have, for $(\boldsymbol{q}, \boldsymbol{a}_i) \in \mathbb{R}^p$, for n attractors and by computing the l-norm[3]:

$$\Phi_i(\boldsymbol{q}, t) = -\frac{1}{\|\boldsymbol{q}(t) - \boldsymbol{a}_i(t)\|_p^{2\beta_i(t)} + \delta_i(t)} \tag{3.7}$$

Therefore, the model extends the classic bi-stable model (Gammaitoni et al., 1998). However, the larger p-dimensional space remains abstract and symbolic, as computing such a model with a realistic level of complexity comparable to the human visual system would formulate the computations in a highly dimensional space—see Appendix A.1.2. The perceptual model also integrates all attractors' energy fields by summation and normalisation, and the logarithmic attenuation is applied such that:

$$V(\boldsymbol{q}, t) = -\ln\Big[-\sum_{i=1}^{n} \Phi_i(\boldsymbol{q}, t)\Big] \tag{3.8}$$

[3]The formalism with the l-norm is presented as such to show the flexibility of such energy potential models such as the one presented here, which could, in theory, allow to approach complexity levels close to the human visual system. However, in the rest of this section, given we have no hypotheses on that complexity, we work on $p = 2$ and $l = 2$.

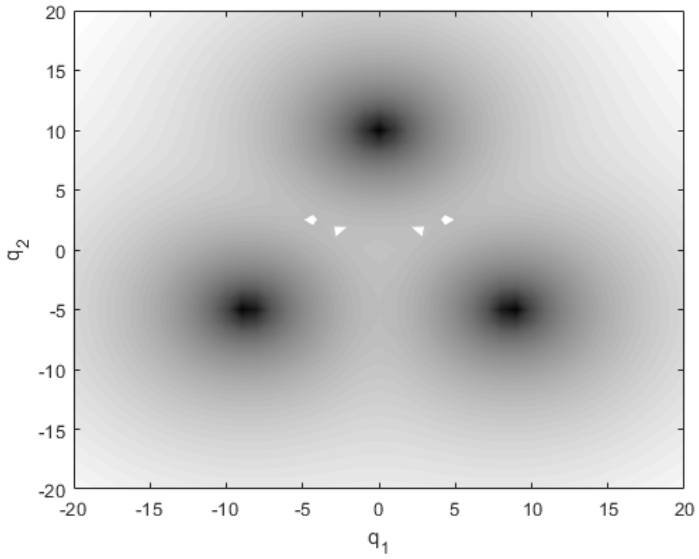


Figure 3.5. Tri-stable perception space. Visualisation of the simulated tri-stable perceptual space with three equi-distant attractors. Note that the white forms are due to errors in the visualisation method.

However, unlike the GraFEM model, we do not apply the MA process as attractors did not move in our simulations and no perceptual phenomenon, to the author's knowledge, justifies for now, adding this computational cost. Finally, as for GraFEM, the particle's gradient was computed by iterations and a noise (η) was forced upon its first order derivative:

$$\dot{\boldsymbol{q}} = -\nabla V(\boldsymbol{q}, t) + \eta(t) \tag{3.9}$$

Particle trajectories in the perceptual energy space then need to be processed and interpreted as percepts.

3.2.2 Simulation methods

We simulated 100 bi-stable systems in $\mathbb{R}^2$ with attractors $\|\boldsymbol{a}_1 - \boldsymbol{a}_2\| = 22.06$ (arbitrary units) over 40 seconds per simulation, with a Euler-Maruyama time step of 1 ms. Attractors did not move in the energy space. The following parameters, kept constant over time and simulations, were used; $\boldsymbol{\beta} = (3, 3)$, $\boldsymbol{\delta} = (0.01, 0.01)$. Noise samples were drawn from a Gaussian distribution with mean $\mu = 0$ and standard deviation $\sigma = 1$ and samples were then multiplied by a noise amplitude value that varied across simulations. Only noise amplitude varied over simulations ranging from 0.45% to 13.6% as a ratio of attractor distance $\|\boldsymbol{a}_1 - \boldsymbol{a}_2\|$. The particle was initiated outside the two attractors at a distance of $\|\boldsymbol{a}(0) - \boldsymbol{q}(0)\| = 8.7633$. The generated

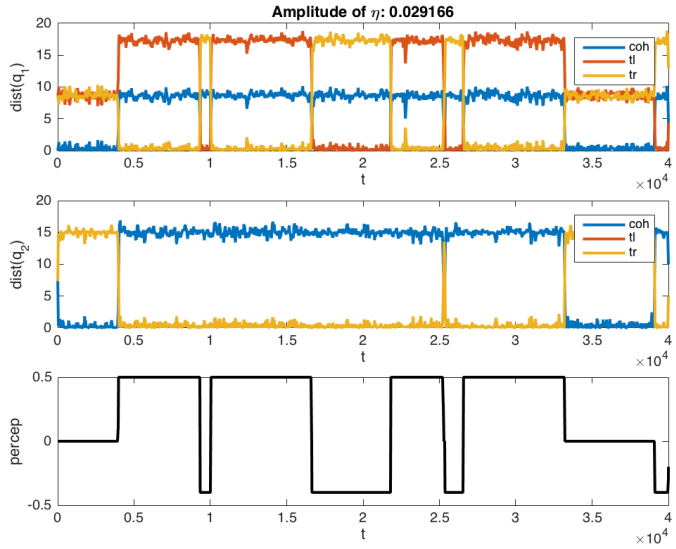


Figure 3.6. Tri-stable perception interpreter. Example the tri-stable interpreter based on finding the minimal distance between the particle's samples of the filtered trajectories. In *colour*, distances to the three simulated moving plaid percept (for more details on the stimulus, please refer to Section 4.1) are plotted and in *black* the interpreter's time series with 0 for coherency (coh), 0.5 for right transparency (tr) and -0.5 for left transparency (tl).

particle trajectories were processed to interpret them as percepts. First, noise was removed using a 100$^{\text{th}}$ order median filter (see Fig. 3.7). The signals were then interpreted into percepts p associated with each attractor, by finding the attractor with closest distance over each iteration (Fig. 3.6) using the following rule:

$$p_i = \arg\min_{i=1,n} \|\boldsymbol{a}_i - \boldsymbol{q}(t)\| \tag{3.10}$$

over n the total number of attractors.

The obtained time series were then analysed using the following procedure. A first order derivative was used to find moments of simulated perceptual change—the first and last 50 iterations were removed to avoid boundary effects. Percept durations were thus extracted and normalised as fractions of the total simulation time (40s). They were then fed to the default *MATLAB* function estimating parameters to fit Gamma and a Log-normal distributions, using maximum likelihood. We used Kolmogorov-Smirnov goodness-of-fit tests (Massey Jr, 1951), which quantifies the distance between empirical distribution function of the sample and the cumulative distribution function of the referred distribution, at $\alpha = 0.05$ to determine whether the estimated Gamma and Log-Normal distributions could be considered as adequate for analysis. One test was done for the Gamma distribution as reference, and one

test was applied with the Log-Normal distribution as reference. Noise amplitude varied across all 100 simulated bi-stable systems.

The same analysis was carried for tri-stability, i.e., 3 attractors, with constant parameter values identical to the bi-stable case described above. The third attractor was placed in a equi-distant manner (Fig. 3.5) and simulations were initiated in that attractor to reflect first percept biases of the moving plaid stimulus—see Chapter 1 and Section 4.1 in Chapter 4 for more details on the stimulus. The results are visualised together with bi-stability in order to facilitate comparison.

3.2.3 Bi-stability

The simulations presented show how the model can change its regime of stability: the system may become unstable, even after the interpretation step, or it can show only one stable state, i.e., mono-stability, or it can have two stable states, i.e., bi-stability. Hence, multi-stability is not the only regime. In Fig. 3.8, one can see how, for low noise levels, the system displays detected changes only at initiation, or at the end (suggesting a boundary effect) and can be considered as mono-stable. The particle is stuck in the first attractor it dives into and remains so. On the other hand, when noise is high, the system becomes so unstable that the percept decoder cannot interpret stable states. This is represented by the decoder's erratic trajectory in the bottom part of Fig. 3.8, where the system enters an unstable regime. In between, however, we do observe bi-stability, with number of switches observed increasing as noise increases in the system.

Because we are interested in multi-stable behaviour, we report Gamma and Log-normal distribution fits only on the simulations for noise amplitude levels of 1.91% to 5.61%—the others are considered mono-stable or unstable simulations. As described in the literature (reviewed in Chapter 1 and Appendix A.4), the percept durations observed in the experimental data can fit to Gamma and Log-Normal distribution functions—however, further goodness of fit tests need to be applied to solidify these results—, as these distribution are typically used in the multi-stability literature (see Chapter 1). Fig. 3.10-*bottom* reports the spread of the estimated distribution parameters, while Fig. 3.10-*middle* shows the computed log-likelihood and the number of switches as noise in the system increases. Both functions seem to perform similarly to describe the simulated data, based on the log-likelihood values reported and none of the analysed simulation were rejected by the Kolmogorov-Smirnov goodness-of-fit tests for both Gamma and Log-Normal estimations. As expected, higher system noise amplitude leads to an increase in the system's number of

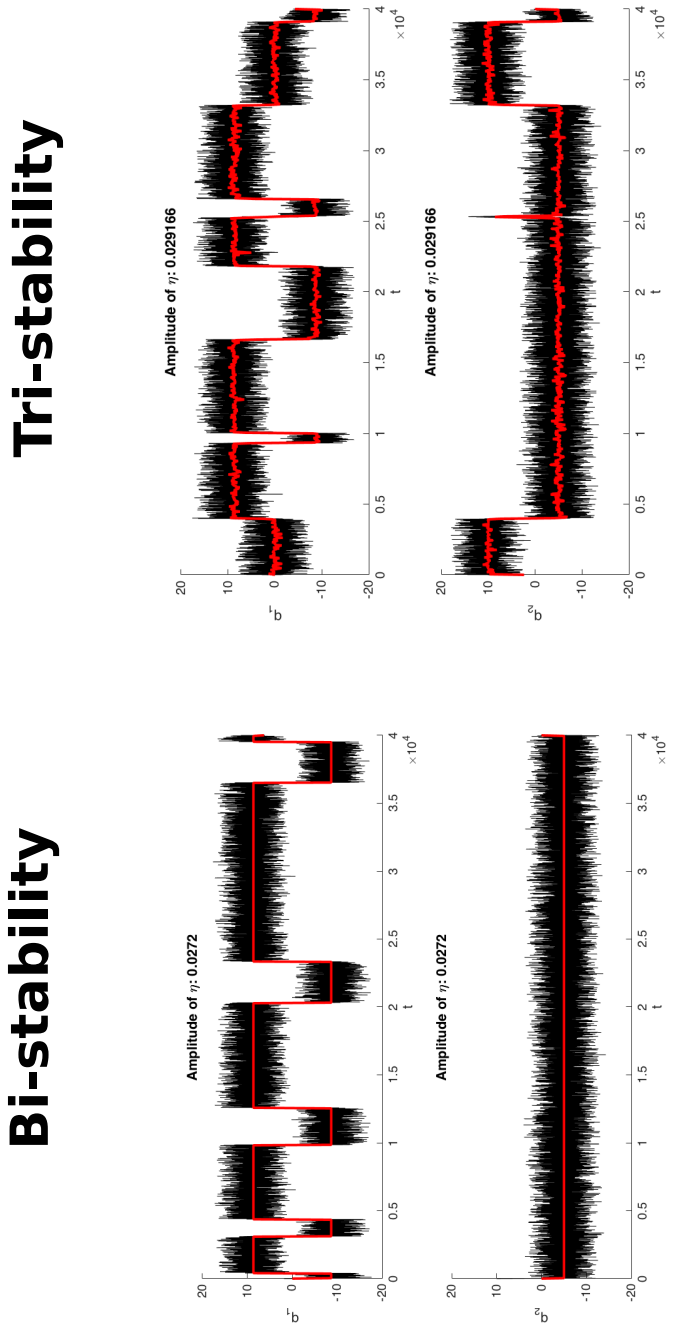


Figure 3.7. Figures are both rotated 90° anti-clockwise.
Bi-stable perception simulation. Example of bi-stable simulation of 40000 iterations, with $\beta = (3, 3)$, $\boldsymbol{\delta} = (0.01, 0.01)$ and noise amplitude of 2.72% of attractor distance. In *black*, the raw signals from particle simulation and in *red*, the signal filtered by a 100$^{\text{th}}$ order median filter.
Tri-stable perception simulation. Example of tri-stable simulation of 40000 iterations, with $\beta = (3, 3, 3)$, $\boldsymbol{\delta} = (0.01, 0.01, 0.01)$ and noise amplitude of 2.9166% of attractor distance. In *black*, the raw signals from particle simulation and in *red*, the signal filtered by a 100$^{\text{th}}$ order median filter.

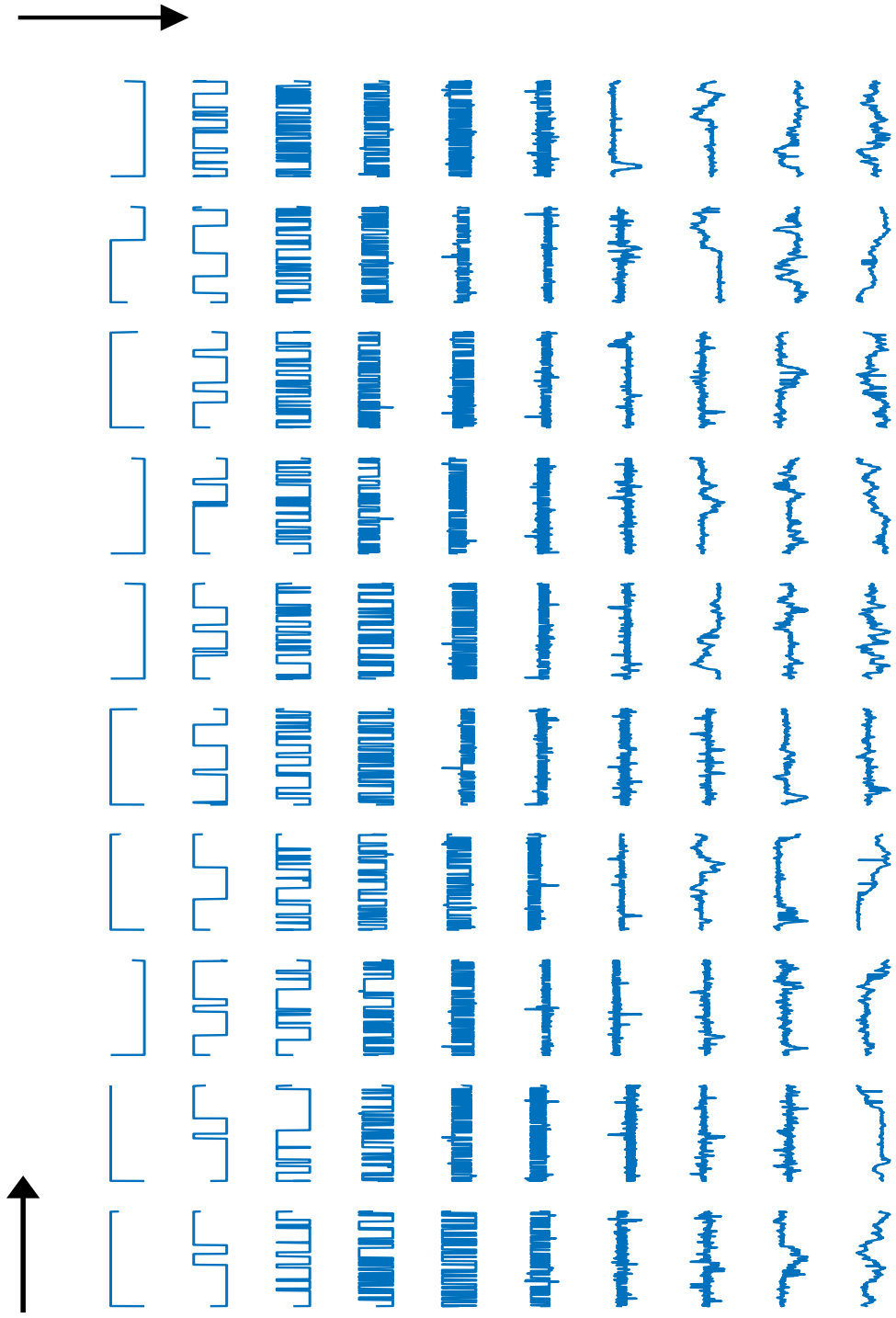


Figure 3.8. Bi-stability simulations. Figure is rotated by 90° anti-clockwise. The arrows show the direction, when rotated, of increasing noise levels along lines and down columns, respectively (*read like a matrix)*). All 100 simulation results, after interpretation of percepts. This figure shows (qualitatively) how noise impacts the model, with simulations with lowest (*top-left, when rotating*) to highest (*corner bottom-right, when rotating*) noise amplitudes.

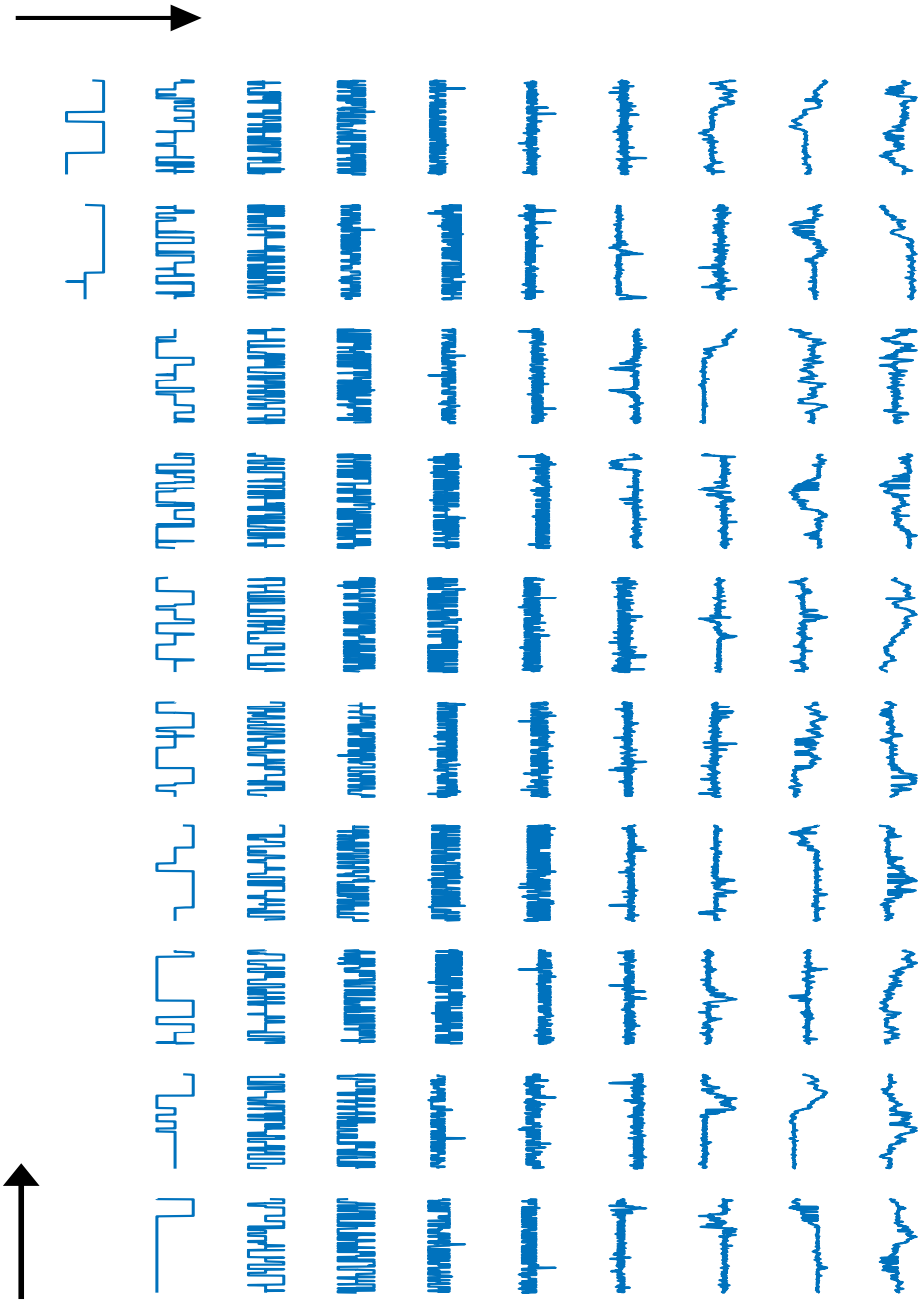


Figure 3.9. Tri-stability simulations. Figure is rotated by 90° anti-clockwise. The arrows show the direction, when rotated, of increasing noise levels along lines and down columns, respectively (*read like a matrix)*). All 100 simulation results (some low noise simulations did not vary enough to yield interpretations), after interpretation of percepts. This figure shows (qualitatively) how noise impacts the model, with simulations with lowest (*top-left, when rotating*) to highest (*corner bottom-right, when rotating*) noise amplitudes.

switches (Fig. 3.10-*middle*), thus replicating the observations from previous studies on the role of noise in bi-stable systems. This appears to be also correlated to less performing distribution fits (see Fig. 3.10-*middle*).

3.2.4 Tri-stability

Here as well as in bi-stability, we observe a reproduction of system stability regimes, shown in Fig. 3.9. For tri-stability, the particle was initiated in the coherency attractor. Again, when noise is high, the system becomes so unstable that the percept decoder cannot interpret stable states. This is represented by the decoder's erratic trajectory in the bottom part of Fig. 3.9, where the system enters an unstable regime. In between, however, we do observe tri-stability, with number of switches observed increasing as noise increases in the system.

Because we are interested in multi-stable behaviour, we report Gamma and Log-Normal distribution fits only on the simulations for noise amplitude levels of 1.91% to 5.61%—the others are considered mono-stable or unstable simulations. Fig. 3.10-*bottom* reports the spread of the estimated distribution parameters and the evolution of computed log-likelihood and the number of switches as noise in the system increases. Gamma and Log-Normal distribution fit the tri-stable data similarly, across noise levels, alike to bi-stable simulations (Fig. 3.10-*middle*). However, 14.82% of estimations were rejected by the Kolmogorov-Smirnov test for the Log-Normal distribution across the considered noise interval, while no rejections were observed for the Gamma distribution. Moreover, higher system noise amplitude appears to be correlated to a large increase in the system's number of switches (Fig. 3.10-*middle*), thus generalising the observations from previous studies on the role of noise in bi-stable systems to tri-stability.

3.2.5 Discussion

With these simulations we have provided:

- A description of empirical computational works on a multi-stable model based on gravitational energy potentials.
- A generalisation from bi-stable models to multi-stable models, with a showcase of tri-stability simulations as a first step.

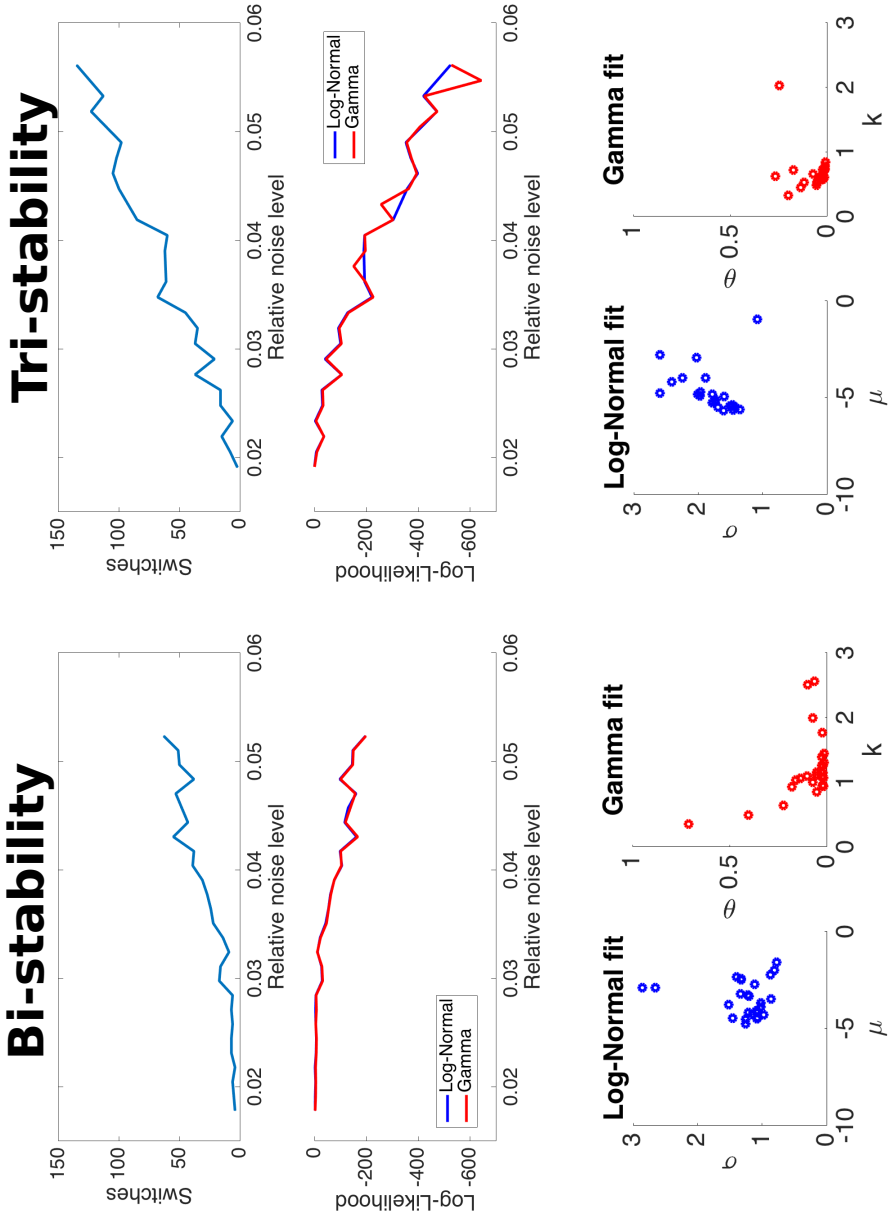


Figure 3.10. Figure is rotated by 90° anti-clockwise.
Bi-stability analysis. *Top*: Number of perceptual switches as a function of noise level in the system. *Middle*: Computed log-likelihood values for estimated distributions as a function of noise level in the system, for the Gamma (red) and Log-Normal (blue) distributions. *Bottom*: Scatter of estimated parameters for Log-Normal and Gamma distribution fits. Kolmogorov-Smirnov goodness of fit tests were computed and estimations rejected at $\alpha = 0.05$ are not displayed; here 0% of the data for both Gamma and Log-Normal.
Tri-stability simulations. *Top*: Number of perceptual switches as a function of noise level in the system. *Middle*: Computed log-likelihood values for estimated distributions as a function of noise level in the system. *Bottom*: Scatter of estimated parameters for Log-Normal and Gamma distribution fits. Kolmogorov-Smirnov goodness of fit tests were computed and estimations rejected at $\alpha = 0.05$ are not displayed; here 0% of the data for Gamma and 14.82% for Log-Normal.

- A framework based on oculomotor modelling that can be applied to perceptual ambiguity.

Differences between bi-stability and tri-stability.

The simulations presented above provide a computational view of the differences that may arise in multi-stable systems, going bi-stability to tri-stability. Both Gamma and Log-Normal fitted parameters are concentrated in small sub-spaces in tri-stability and bi-stability (Fig. 3.10). Also, while both Gamma and Log-Normal described the data in a similar fashion in the bi-stable system, we observed a difference between the two functions' performance for tri-stable system, with the Log-Normal distribution estimations being rejected by a Kolmogorov-Smirnov goodness-of-fit test for 14% of the considered simulations. Finally, the tri-stable system, for the same noise levels, goes through more perceptual changes than the bi-stable system, thus suggesting that as the number of attractors increases, residual durations in attractors decrease and system changes are more frequent.

In fact, this last point raises question on the view proposed by (Kelso, 2012) that mono-stability is a regime in the region between multi-stability and meta-stability (see Fig. 1.21 in Chapter 1). If meta-stability typically produces unstable behaviours, our study links it to noise amplitude in the system; i.e., for both bi-stability and tri-stability, the models produced highly unstable behaviours as noise took over the gradient force and allowed the particle to roam (see dynamics presented in the bottom half of the visualisation matrix in Figs. 3.8 and 3.9), regardless of the slope in the energy field. Furthermore, given the results for numbers of detected perceptual changes shown in Fig. 3.10-*top*, one can expect that increasing the number of attractors in a multi-stable system will generate less stable observations. However, these would differ from instability driven solely by noise, as we can expect short term dynamics to show persistence, and longer ones anti-persistence.

Model fusion and predictions

The last point links back to the data and model proposed for FEM that inspired our framework (Engbert, Mergenthaler, et al., 2011). Indeed, an aim of this work has been to propose a methodology that can combine eye movements and perception. If an agent's next oculomotor action can be motivated by top-down intentional (decision making) and bottom-up salient (attention) processes, multi-stable perception provides an interesting paradigm to study how such interactions emerge and

operate. In fact, the data observed in the Necker experiment (Chapter 2) suggested such interactions too. Here, both models were studied separately, but a future work would consist in investigating how motion attractors in the oculomotor space can act as perceptual attractors. For instance, this can be done with motion based multi-stable stimuli such as the moving plaid, where perceived motion directions alternate (Moreno-Bote, Shpiro, et al., 2008; Moreno-Bote, Shpiro, et al., 2010). Interestingly, works have shown that energy potential bi-stable systems provide an elegant mechanism to simulate Bayesian sampling (Moreno-Bote, Knill, et al., 2011). And this argument has been investigated with the moving plaid stimulus, strictly from a perceptual viewpoint. A promising perspective is to formulate such a problem as an active vision one.

Perspectives: investigating the role of adaptation

Lastly, a force often described in bi-stable perception model is adaptation. However its meaning is somewhat unclear, and sometimes it is referred as synaptic depression for percept competition, at other times, it is linked to neural adaptation. Nevertheless, it gives a system a deterministic force related to the current state, namely the attractor, the system is in. In this work, adaptation was not manipulated and only the noise's role was considered. But an interesting future work would be to test how, depending on the defined adaptation function applied on the attractors' depths, the results vary. And furthermore, the presented model provides a method to investigate adaptation's extension and generalisation to multi-stability. This form of computational investigation could derive key insight on the role of adaptation in multi-stable perception.

3.3 Synthesis

In this chapter, we have looked at computational methods for eye movements, in particular FEM, and for perceptual multi-stability. Models based on the same framework, gravitational energy potential fields, have been use to generate and study both eye movement and perceptual dynamics. Thus, an active vision modelling of multi-stable perception taking into account oculomotor dynamics can be rendered and studied. While the data presented in Chapter 2 suggested that oculomotor dynamics can produce multi-stable behaviours, the work presented in Chapter 4 aims to investigate this experimentally, based on the theoretical understanding and gains from this chapter's work.

4 Multi-stability: manipulating perceptual ambiguity

> *Many people would accept that we do not really have knowledge of the world; we have knowledge only of our representations of the world. Yet we seem condemned by our constitution to treat these representations as if they were the world, for our everyday experience feels as if it were of a given and immediate world.*
>
> — ***Francisco Varela***
> **The Embodied Mind, 1991.**

In this chapter, the results of multiple experiments are presented. The ideas and design are driven from the models proposed in Chapter 3, and aim to investigate the multi-stable dynamics of perception with the moving plaid stimulus on one hand, and the dynamics of the oculomotor system on the other. The emergence of multi-stability is the result of combining a stimulus with specific signal properties and competing inferences for the visual system. It is possible to stabilise or bias these perceptions by changing the stimulus through expectations or task manipulation or also by modifying the oculomotor dynamics. The perceptual and oculomotor systems being interlinked, acting on one of them has an impact on the other. This principle will be shown and described empirically in this chapter (and the next), with two experiments in which ambiguity is manipulated. The aim being to reach a situation where it is possible to show that motor control can be a physiological marker of perceptual content, in a no-report paradigm. The aim of this chapter is also to show that the models and theoretical ideas covered in Chapter 3 can be transcribed into an empirical and experimental investigation.

Contents

4.1 Hypotheses

In this section, we present the reasoning and rationale making possible the derivation of the major hypotheses.

> **Using the theoretical and computational frameworks developed in this thesis (Chapter 3), namely the gravitational energy field models for eye movements and perception, and by applying them to the moving plaid, an experimental design is conceived and proposed. Its aim is to investigate how active vision theories can be applied to link oculomotor and perceptual dynamics for the moving plaid.**

The experimental design is driven from theoretical work, inspired from premotor theories of attention (Rizzolatti, Riggio, Dascola, et al., 1987; Rizzolatti, Craighero, et al., 1998), linking motor control, attention, and visual perception as three interacting spaces using attractor models.

4.1.1 Building a framework combining theoretical and experimental methods

Eye movements expected: (micro-)pursuits and (micro-)saccades

In Chapter 3, the model for multi-stable perception extended on current bi-stable models based on a particle in an energy potential framework (Moreno-Bote, Rinzel, et al., 2007; Shpiro, Moreno-Bote, et al., 2009; Moreno-Bote, Knill, et al., 2011; Engbert, Mergenthaler, et al., 2011). Furthermore, the proposed GraFEM model has multi-stable dynamics with respect to attractors in the visual fields. The oculomotor and perceptual models, are similar in dynamics and mechanisms, however, they differ in terms of interpretation and physiology. While the first model remains a large simplification of a highly multi-dimensional neuronal space in which a conscious percept is coded for the visual system, the latter could relate to efference copies in the SC, the FEF or the LIP cortex (Keating, 1991; Krauzlis, 2005; Krauzlis, 2004; Hafed, Goffart, et al., 2009; Astrand et al., 2015; Krauzlis et al., 2017). In the context of enacted and embodied cognition, the markers of perception could be detected and interpreted in the active mechanisms of the observer. Therefore, eye movements could partly reflect the internal state of perception.

The GraFEM model can reproduce pursuit mechanisms by having an attentional attractor, implemented on a gravitational energy field, draw the gaze particle, as

described in Chapter 3. If combined with a multi-stable stimulus in which the percepts are related to ambiguous motion decoding, a simple hypothesis drawn from an enacted cognition view (see Appendix A.1) can be formulated such that an attractor in the observable eye movement space corresponds to an attractor in the less accessible space of perception. The main hypothesis in this work can be formulated as follows:

> **Eye movements dynamics are driven by underlying visual multi-stability over plausible and competing motion percepts. Thus, eye movements may reflect, at times, the content of perception.**

In this framework, the above statement translates in the merging of the interpretations of attractors in the perceptual and the visual spaces. In other words, when one perceives a motion[1], one's gaze dynamics are highly influenced by its corresponding visual attractor, thus leading to oculomotor dynamics correlated to the motion. Hence, the detection of pursuits and micro-pursuits could be linked to the detection of motion perception. As seen in Chapter 3, when a visual attractor becomes more influential with by generating a deeper field or widening its read, changes will be applied to the observed gaze dynamics; one can expect switching between attractors to produce saccades and pursuits, based on the properties of the attractors. Furthermore, pursuits can be associated with catch-up saccades and micro-saccades, as the former precedes the latter in tasks where gaze should remain at the centre of screen and visual motion is displayed. They could be predictors of changes in dynamics (L. C. v. Dam and Ee, 2006b; L. v. Dam and Ee, 2005; Laubrock et al., 2008; Rolfs, 2009; Spering and Montagnini, 2011; Engbert and Kliegl, 2003; Hicheur et al., 2013) and these eye movements would be considered as physiological and active markers of the internal state of perception (Aleshin et al., 2019; Frässle et al., 2014; Einhäuser, Thomassen, et al., 2017).

Moreover, our hypothesis was that eye movements will be marked by maximal perceptual information at the moment of perceptual reversal, and in the moments that follow up. Based on the description and explanation of the oculomotor model given in Chapter 3, one can predict that a change of oculomotor attractor (e.g., moving in a different direction) brings about new motion for the gaze for an epoch. Another possibility is that the switch of attractor will generate a saccade as the gaze particle is rushed to the new attractor. With the hypothesis given above, the same applies to perceptual attractors as they are merged with the oculomotor ones. As the perceptual state particle reaches rapidly a new perceptual attractor, the inhibition of competing perception is at its maximum.

[1] i.e., a perceptual attractor for the visual system

Finally, though this approach showcases an association between the oculomotor and the perceptual spaces, it is possible, through experimental manipulation, to observe the influence of one system on the other.

Therefore, the following precise hypotheses can be derived:

1. **Percepts**: in visual multi-stability, they correspond to stable states of the perceptual system compete for dominance, i.e., perceived motion direction, for instance, and they thus act as perceptual attractors;

2. **Ambiguity**: perceptual attractors can be manipulated by controlling the stimulus' parameters and thus the dynamics of visual multi-stable perception can be influenced (Hupé and Rubin, 2003; Moreno-Bote, Shpiro, et al., 2008);

3. **Eye movements**: markers such as pursuits followed by corrective saccades can be detected to infer perceptual dynamics;

In the context of visual ambiguity leading to multi-stable perception, one can manipulate the visual properties of the stimulus to bias a percept over another, and observe whether the changes in oculomotor dynamics correspond to the associated predictions. Alternatively, it should theoretically be possible to manipulate the oculomotor dynamics and measure how this affects the perceptual dynamics. Investigating causal links and temporal correlations between the oculomotor and perceptual sub-systems and their respective attractors also raises question regarding organisation and hierarchy.

Overall, the framework proposed and applied here, gives an understanding of visual multi-stability as a multi-system phenomenon, here involving eye movements and perception. Its characterisation will be done by identifying the interactions and dynamics between sub-systems as well as those between the attractors within each sub-system. In this work, the focus is on the oculomotor system and the perceptual system, when facing a physically stationary ambiguous stimulus that leads to perceptual shifts. Among the possible known stimuli from the literature, one particularly fits the criteria needed to implement this framework: the *moving plaid*.

4.1.2 The multi-stable moving plaid stimulus

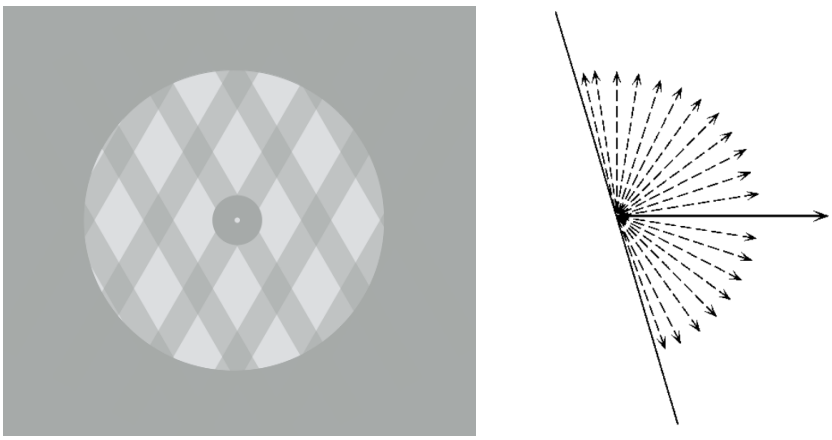
Among the variety of multi-stable stimuli studied in psychology and neurosciences (presented in Chapter 1), the **moving plaid** (Fig. 4.1a), a stimulus consisting of two superimposed gratings with different angular orientation moving in opposite directions presented through an aperture, presents many features that can be exploited in our framework. Here, the motion ambiguity and its link to perception shall be presented, followed by a description of the moving plaid's parameters and finally, empirical results, specific to it, from psycho-physics, neurosciences and eye movement studies are reported.

Motion ambiguity

The ambiguity of the moving plaid stimulus is related to motion perception, segmentation and integration (Welch, 1989; Hupé and Rubin, 2003; Hupé and Rubin, 2004; Rubin, Hupé, et al., 2005; Moreno-Bote, Shpiro, et al., 2008; Moreno-Bote, Shpiro, et al., 2010). Essentially the direction (Hupé and Rubin, 2004) and the depth ordering (Moreno-Bote, Shpiro, et al., 2008) of the scene cannot be clearly interpreted by the visual system—illustrated in Fig. 4.1b; figure taken from (Wuerger et al., 1996). Because the observer cannot see the edges of the gratings' bar move (see Fig. 4.1a), its perceived motion is ambiguous. All the dashed arrows in Fig. 4.1b show the possible interpretations for a real physical motion shown by the continuous arrow (which can also be perceived). This motion ambiguity is further mixed with a depth ordering ambiguity dictated by the gratings' bar transparency properties. Depending on how the parameters are set for the stimulus, it is possible to generate more or less stable states of perception. For example, the grating with higher spatial frequency will be mostly likely perceived as behind, and the same goes for the grating with the smaller duty cycle (Moreno-Bote, Shpiro, et al., 2008). If the gratings have different velocities, the faster one will tend to be more perceived as behind too.

The percepts are defined and usually reported by the participants using keys from the keyboard referring to each percept.

> **The depth ordering competition corresponds to the *transparency* percepts, in which the gratings are seen as two different objects moving in opposite direction, with one being on closer to the observer than the other. When the grating moving rightward is perceived as closer, the percept is said to be *transparency right*. And when it is moving leftward, the percept is called *transparency left*.**



(a) Moving plaid. **(b)** Motion ambiguity.

Figure 4.1. Motion ambiguity in the moving plaid.
(a) Static image of the moving plaid, showing the superimposed gratings displayed behind an aperture.
(b) Diagram of the relationship between the direction of motion and the interpreted sensations of motion for the aperture problem. Figure taken from (Wuerger et al., 1996).

The bi-stable dynamics of the moving plaid have been deeply studied (**castelo2000neural hupe2003dynamics**; Stone et al., 1990; Stoner and Albright, 1992; Moreno-Bote, Shpiro, et al., 2008; Moreno-Bote, Shpiro, et al., 2010) and is mostly generated when the relative angle is relatively small (less than 45° or $\frac{\pi}{4}$ radians) or with non square wave gratings.

> **The moving plaid has also been studied for its tri-stability in which the two transparency percepts are present. The third percept is named *coherency* and is experienced as the fusion of the two gratings into a single grid, typically moving upwards.**

Tri-stability also raises questions on hierarchy of competition—*more details on hierarchy of percepts in the box below*.

More generally its direction is the sum of the motion vectors of the transparency percepts (Welch, 1989; Gorea and Lorenceau, 1991). More formally, one can consider the problem to be defined as follows. The perception is a random variable X, that can become one of the three percepts of coherency (c), transparency left (l) and right (r)—i.e., $X = x$ with $x \in \{c, l, r\}$.

Is there a hierarchy for percepts? And for spaces?
Supposing that oculomotor and perceptual attractors are tightly linked, in an enactive or embodied approach to cognition context for instance, the perceptual and oculomotor manipulation can provide insights on the relationships between percepts in their respective spaces. The question arises when moving beyond bi-stability to multi-stability. A hierarchy of percepts implies that the emerging percepts do not compete within the same neural network or the same visual decoding process. The question has been of interest in the literature (Friston and Kiebel, 2009; Huguet et al., 2014; Megumi et al., 2015; Hupé, Signorelli, et al., 2019). The approach presented in this chapter addresses this issue by measuring asymmetrical relationships between perceptual dynamics—i.e., a percept acts as a systematic transition, or can never be suppressed by stimulus manipulation, etc. Or, on the contrary, if the data collected can be modelled with attractors of identical properties, this would form a body of evidence against a hierarchical view. A more neutral and less biased conceptualisation is to consider that attractors can be considered to influence each other directly, in a unique neural network, with no hierarchical organisation. On the other hand, an indirect relationship would suggest that the percepts can be coded over multiple networks interlinked by relatively reduced interdependent connections. The consequence is that these networks could be seen to synergetic or competitive behaviours. The more isolated and specialised the networks, the more likely they are to generate a hierarchical organisation.

Parameters description

The moving plaid relies on the aperture problem to generate ambiguity (Wuerger et al., 1996). The gratings are composed of transparent square waves (Fig. 4.2).

It is composed of two grids, generated as gratings stimuli—the one moving leftwards (G_L) and rightwards (G_R) from the observer's perspective, respectively—with a relative angle θ_{diff} computed from the difference of orientation of these gratings ($\theta_L, \theta_R \in [-180; 180]^2$ in degrees or $\in [-\pi; \pi]^2$ in radians) so that $\theta_{\text{diff}} = |\theta_L - \theta_R|$. The luminance of the gratings is set by parameters L_L, L_R but are eventually modified by manipulating their transparency. The gratings also have levels of transparency ($\alpha_L, \alpha_R \in [0; 1]^2$) that generate further ambiguity on depth ordering of the objects and change the physically observed luminance and contrasts. The perception changes are based on the direction of the structure of perceived movements. The motions of the gratings have velocities (v_L, v_R). Other aspects that can be controlled are the spatial frequencies of the grating ($f_L = 1/\lambda_L, f_R = 1/\lambda_R$ with λ_L, λ_R the wavelengths) and the associated duty cycles ($D_L =, D_R$). The aperture radius r_A can also be manipulated, as well as the fixation circle area's radius r_F. Background and aperture luminances can also be controlled with L_{BG} and L_A, respectively. The

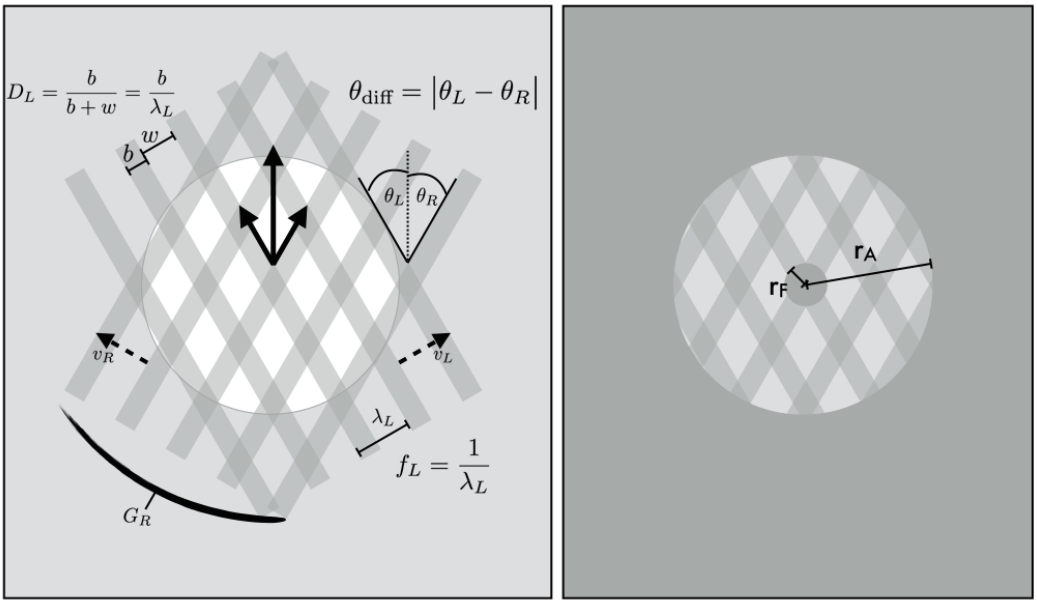


Figure 4.2. Moving plaid parameters.
Left: diagram of a moving plaid stimulus with parameter visualisation for the parameters from Tab. 4.1. Luminances $\boldsymbol{L}$ and transparencies $\boldsymbol{\alpha}$ are not represented spatially as they refer to pixel brightness. The three continuous arrows represent the three percept of moving plaid's tri-stability. The gratings' orientation $\boldsymbol{\theta}$ are shown relative to the vertical axis. The gratings' motion velocity is shown as the horizontal dashed lines v_L and v_R. Finally, duty cycle D_L is shown as well as spatial frequencies' wavelengths λ_L.
Right: the moving plaid, when presented is occluded by a layer with an aperture hole, such that the bars' edges are not visible, thus creating an ambiguous motion. r_A and r_F define the radius for the aperture and the fixation area, respectively.

parameters described here, are summarised in Tab. 4.1 and Fig. 4.2 with further information on their values for our experiments.

There are a number of reasons why the moving plaid was chosen as a stimulus to investigate the links between perceptual and oculomotor multi-stability. First, this stimulus has a high number of parameters that can be manipulated. This gives more possibilities to control ambiguity and experimental conditions over more classical complex images such as the rabbit/duck, old/young women or Rubin's Vase (Chapter 1). But it also make parameter space exploration a problem with high complexity due to the high number of possible combinations. Unlike these cited stimuli where the bi-stability seems to occur at higher abstraction level of interpretation (*for more details, read the box below*), the ambiguity in the moving plaid is generated in visual signals that are simple, straightforward and not dependent on semantic competition, but on motion vectors that can be identified in time and space. The stimulus can be

Formal symbol	Description	Values	Units
$\boldsymbol{\theta}$	orientation	± 30	deg
$\boldsymbol{L}$	base-line luminance	0.35	$\varnothing$
$\boldsymbol{\alpha}$	transparency	$(0, 1)$	$\varnothing$
$\boldsymbol{v}$	velocity	± 1.5	deg.s^{-1}
$\boldsymbol{f}$	spatial frequency	0.01	Hz.pix^{-1}
$\boldsymbol{D}$	duty cycle	0.35	$\varnothing$
r_A	aperture radius	6.37	deg
r_F	fixation disk radius	1.25	deg

Table 4.1. Moving plaid parameters. Parameters manipulated to generate a single square wave grating. Background (L_{BG}), gratings' (L_L,L_R) and aperture (L_A) luminances are provided as normalised numerical values which are then translated into physical light emitted by the screen an measured in candela for a surface (cd.m^{-2}). Their numerical values change according to computations described in Section 4.2.2. $\varnothing$: ratios with no units.

manipulated in subtle fashion, by tweaking the difference of grating orientations (θ_{diff}) or the transparency of the gratings ($\boldsymbol{\alpha}$). The moving plaid can also be manipulated to generate multiple states of perception, notably a tri-stability that has been studied (Huguet et al., 2014). Fig. 4.3 shows a schematic view of the percepts presented in (Hupé and Rubin, 2003).

Where does bi-stable competition occur in the visual system?
The level in visual processing at which the bi-stable competition is thought to occur is at later stages of the visual pathways. This can be explained by the semantic competition of the percepts. In fact, cross-modal studies of bi-stability have shown how giving a cue in audio on a visual rivalry can modulate the perceptual dynamics, thus suggesting that for complex and semantic perceptual objects, the brain accumulates evidence from different sensory inputs and binds them (Y.-C. Chen et al., 2011). A review of perceptual binding and cross-modal work on multi-stable perception can be read for further details (Schwartz et al., 2012).

The moving plaid has been studied extensively in the context of perceptual bi-stability (Rubin, Hupé, et al., 2005). Hupé and Rubin (2003) showed that the duration of the first percept can be linked to the relative dominance coherency perception over time, in a first experiment—see comparison between both measures in Fig. 4.4**A**. Their results were obtained for a bi-stability task, where transparency (t) and coherency (c) percepts were reported. They provided three key observations: (i) over time, the distribution of percept durations are stable, (ii) coherency (c) was systematically the first percept, and (iii) the first percept was considerably longer than the other ones. In fact, in a first experiment, they showed that the first percept's log-duration $\ln(\text{RTransp})$ was linearly correlated to the trials' empirical dominance of coherency $C/[C+T]$—where C is the sum of durations in which coherency

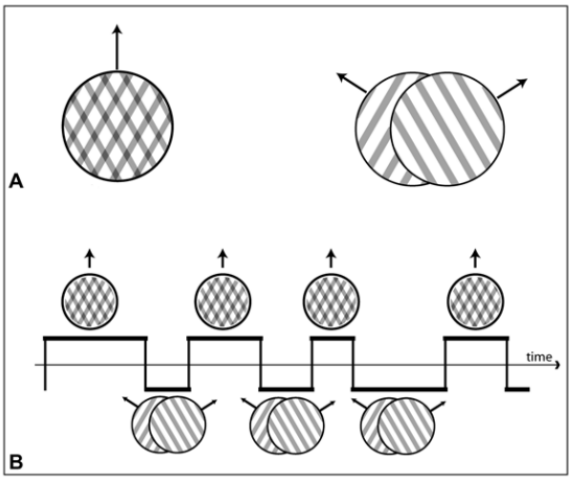


Figure 4.3. Moving plaid perception. Diagrams presenting the perceptual dynamics of the multi-stable moving plaid stimulus.
A: diagram showing the two different perceptual interpretation considered in the case of the bi-stable plaid.
B: diagram showing a schematic example of the temporal discourse of perception.
Figure taken from (Rubin, Hupé, et al., 2005).

was perceived, without counting the first percept duration, and T the same for the transparency percept.

Based on this method, they showed, in a second experiment, that the difference of angles between the gratings (θ_{diff}) impacts the dominance of the coherency; the smaller θ_{diff}, the more the transparency percepts dominate (Fig. 4.4**B** *left*). Likewise, when manipulating both gratings' velocities ($\boldsymbol{v}$), they showed the following effect: coherency dominance decreased as velocity increases, though the effect was weaker than for angles (Fig. 4.4**B** *right*). Finally they manipulated both duty cycles ($\boldsymbol{D}$) simultaneously and showed a reinforcement of coherency dominance as duty cycle tends towards 0.5 (Fig. 4.4**B** *bottom*). In their third experiment, spatial frequency ($\boldsymbol{f}$) and the wave functions—e.g., square, sinusoidal or rectangular wave pattern for the gratings—were investigated as control parameters on perception, and $\boldsymbol{f}$ was reported to impact the intercept of the monotonic decrease function for the rate of coherency dominance along the θ_{diff} variations while the wave shape changed the rate of decrease (Fig. 4.4**C**).

In a subsequent study, Hupé and Rubin (2004) showed that simultaneous direction changes of the gratings' motion led to percepts' motion direction changes as shown in Fig. 4.5. Furthermore, the relationship between coherency dominance and θ_{diff} was replicated while investigating the global motion orientation changes. The plaid's depth order for transparency was characterised in a series of experiments

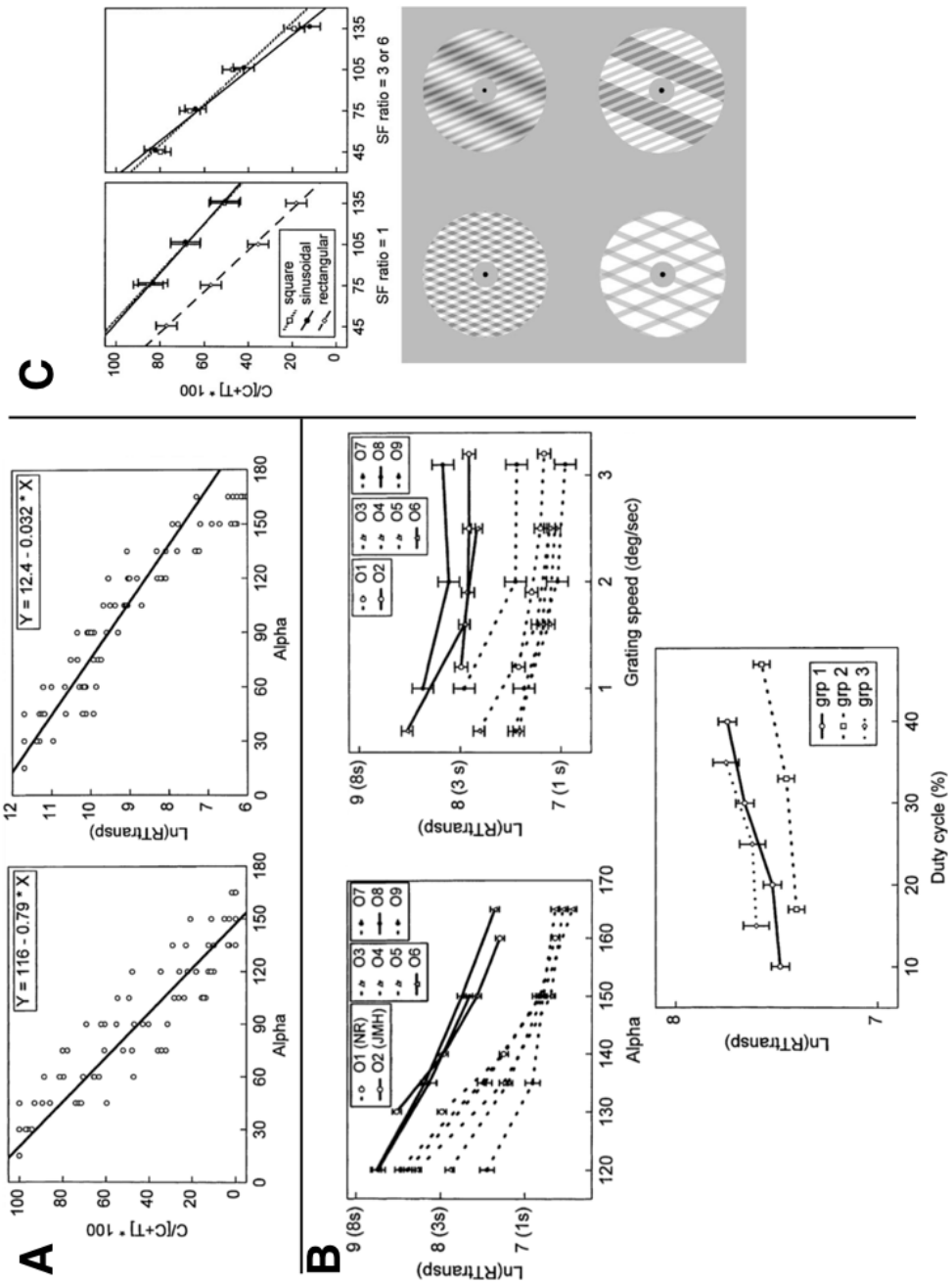


Figure 4.4. Moving plaid literature results.
A: the relative dominance of coherency ($C/[C+T]$ with C and T the coherency and transparency percepts' total durations over a trial, respectively) follows the same dynamics as the logarithm of the first percept duration (RTransp) over variations of relative orientation differences—here named Alpha, but referred to as θ_{diff} in this thesis, not to be confused with the transparency parameters.
B *left* shows the decrease of RTransp over Alpha (θ_{diff}) for three groups of participants tested over different plaid parameter values, while *right* shows the effect of grating velocity, referred to as velocity v in this thesis. *Bottom*, shows how dominance of coherency increases gradually as duty cycle tends towards 0.5.
C: provides the results for different spatial frequencies (SF here, but f in this thesis) as Alpha (θ_{diff}) is manipulated, while below illustrations of the plaid with different function patterns are given.
Figures were taken from Hupé and Rubin (2003).

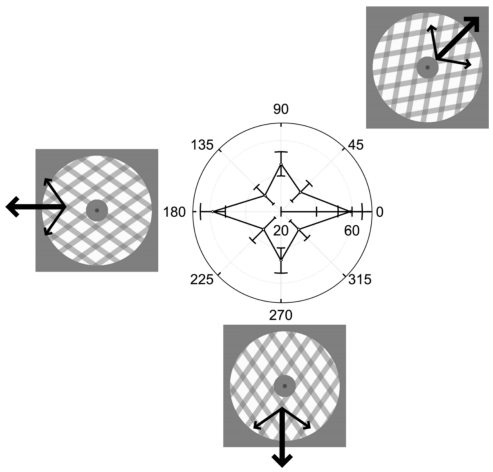


Figure 4.5. Motion orientation in the moving plaid. Polar plot showing the coherency dominance ($C/[C+T]$) as a function of orientations (0°, 45°, 90°, 135°, 180°, 225°, 270° and 315°). Examples of perceived directions at orientations (45°, 180° and 270°) are shown, with global motion direction being shifted. Taken from Hupé and Rubin (2004).

that focused on disassociating the parameter values for each grating (Moreno-Bote, Shpiro, et al., 2008). The authors showed that the grating most likely to appear behind was the one with (i) the higher spatial frequency, (ii) the smaller duty cycle, and (iii) the higher velocity. As depth ordering was studied, the tri-stability of the moving plaid was demonstrated (Moreno-Bote, Shpiro, et al., 2010). The authors showed a moving plaid on one hand in binocular rivalry and on the other hand as an ambiguous dynamic figure with two competitions; coherency *vs* transparency and depth ordering within transparency—e.g., transparency left *vs* right—and showed that perceptual reversal velocity is maximised at equi-dominance for all three bi-stability. They also showed that this feature was predictable by double well bi-stable models.

Furthermore, an argument for choosing the moving plaid to study multi-stability with eye-tracking methods is that these percepts are based on perceived motion. And as motion can be tracked by the eyes in the form of pursuit eye movements, the moving plaid provides an interesting setup to investigate both oculomotor and perceptual multi-stability. A review of works on oculomotor studies with the moving plaid stimulus is provided in Appendix A.6.

Finally, the moving plaid stimulus has been studied in the context of energy potential models (Moreno-Bote, Shpiro, et al., 2010; Moreno-Bote, Knill, et al., 2011), a theoretical approach also used in our modelling work (Chapter 3). The authors also used fraction of time dominance—similar to the percept dominance discussed

above—for each percepts, which they interpreted as a measure of "*percept strength*", to study the dynamics binocular rivalry, plaids (coherency C vs transparency T) and depth reversals (right transparency r vs left transparency l). They also compared their results to double well energy models and neural competition models with cross-inhibition and input normalisation, and showed that equi-dominance—i.e., percept equi-probability, when each percept occur over time as much as the other—led to maximum changes of percepts. Furthermore, the perceptual change rate against dominance results showed symmetry around the equi-dominance point. Although they provide perspectives for generalisation to multi-stability beyond bi-stability, the expected symmetry around the equi-dominance point(s) and the relationship between perceptual switch rate and perceptual dominance is less clear.

Choosing a control parameter: the transparency of gratings

As described at the beginning of this chapter, many parameters can be used to control the properties of the moving plaid stimulus. Thus, it is necessary to isolate a restrained number of parameters that can manipulate perceptual dynamics efficiently, in order to minimise the number of manipulated variables.

> **The approach undertaken in this work was to focus on the gratings' *transparency* parameters ($\boldsymbol{\alpha} = [\alpha_L, \alpha_G]$) and to verify that all percepts can be biased in order to control ambiguity levels (Castelo-Branco et al., 2000). Transparency was chosen as it does not affect the geometry of the stimulus, unlike the orientation. It also does not impact the spatio-temporal dynamics and remains subtle.**

The luminance and contrast are however affected, though the global mean luminance of the stimulus can be modelled and compensated by augmenting the background luminance for instance—we provide a solution in Section 4.2.2. The transparency parameters are normalised and their values can vary in the following space: $\boldsymbol{\alpha} \in [0; 1]^2$.

The aim, in the manipulation of a visual parameter of the stimulus, is to contrast conditions in which trials are considered ambiguous and non-ambiguous for the observer. This has been done with Necker cube lattices (Ehm et al., 2011; Kornmeier and Bach, 2012; Kornmeier and Bach, 2014) where the authors could contrast:

- *endogenous* perceptual reversals, in which the observer changes perception through internal cognitive processes and,

- *exogenous* perceptual reversals, in which changes of perception is induced by controlled experimental manipulation.

The manipulation of the transparency of a grating changes the contrast between gratings and therefore, the probability of seeing that grating's motion. It can be interpreted as a manipulation of percept strength, as is done in binocular rivalry (Levelt, 1966; Chopin, 2012), especially for the transparency rivalry l vs. r, as demonstrated in Moreno-Bote, Shpiro, et al. (2010). Indeed, as transparency impacts the luminance of each grating, it impacts the strength of the percept.

4.1.3 Hypotheses & experimental protocol designs for perceptual and oculomotor investigation of multi-stability

In Chapter 3, the models proposed for eye movements and perception are based on gravitational energy fields and the phenomena are dependent on attractors' dynamics and noise. A first simple prediction to test is whether perception leaves markers in oculomotor dynamics. This corresponds to having a direct link between spaces of oculomotor programming and perception. A minimal set of predictions can be articulated as follows:

- **perceptual changes generate oculomotor markers**—as perception changes, the saliency of features related to the new percept will drive attentional focus and thus, the active visual system will generate eye movements linked to the perceived motion (here pursuits or micro-saccades),
- **oculomotor dynamics lead to perceptual changes**—if gaze dynamics follows a visual attractor, it will accumulate evidence for a percept, and gaze dynamics will precede a perceptual change.

The approach to test these predictions was to design an experiment (presented further in Appendix A.8) in which different conditions are applied:

- **ambiguity** (A) versus **non-ambiguity** ($\overline{A}$)—manipulate the transparency parameters to bias the level ambiguity so that we maximise ambiguity in one condition and bias a chosen percept in the other; this will allow to test the impact of perceptual changes on eye movements,
- **oculomotor control** (F) versus **free oculomotor dynamics** ($\overline{F}$)—vary the oculomotor restriction so as to manipulate eye movements implicitly; this will allow to test the impact of oculomotor dynamics changes on multi-stable perception;

- **fixation** (F) versus **no fixation** ($\overline{F}$)—vary the oculomotor task so that participants restrict their eye movements or let them be free to follow their percepts, to test the scale of oculomotor influences on ambiguity;
- **report** (R) versus **no report** ($\overline{R}$)—vary the perceptual task so that participants are asked to report their subjective perception using key press or not, to test the impact of motor action on perception and eye movement dynamics; the goal here is to be able to predict perception when there are no key presses.

Four pilot experiments (in this chapter, only the first two are presented, the latter two are presented in Chapter 5) were carried out to test each condition and hypothesis:

- the *Percepts* experiment to verify that all three percepts could be identified with transparency manipulation,
- the *Ambiguity* experiment for ambiguity control using transparency,
- the *Eye Movements* experiment (Chapter 5) to verify that oculomotor dynamics can be influenced implicitly and,
- the *Noisy Motor Events* experiment (Chapter 5) to verify that perceptual dynamics can be inferred from oculomotor dynamics.

The diagram Fig. 4.6 shows an overview of the experiments, the studied systems and the observed phenomena.

The methods developed to estimate the maximal level of ambiguity and ambiguity control will be described in details in the *Ambiguity* experiment. The methods detecting changes in the oculomotor dynamics and thus allowing the inference of perception with eye movement data will be presented in the *Eye Movements* and *Noisy Motor Events* experiments. The complete experiment was designed to combine these separate investigation, with the addition of EEG recording for neural correlates exploration—*the experimental design is presented in Appendix A.8*.

> **Overall, based on the models proposed in this work (Chapter 3) and our proposed interpretations, a set of hypotheses, defining a relationship between eye movements and visual signal motion, were applied on the multi-stable moving plaid stimulus. This approach enabled the conception and design of an experimental protocol that will provide data that may characterise the relationships between perception and action.**

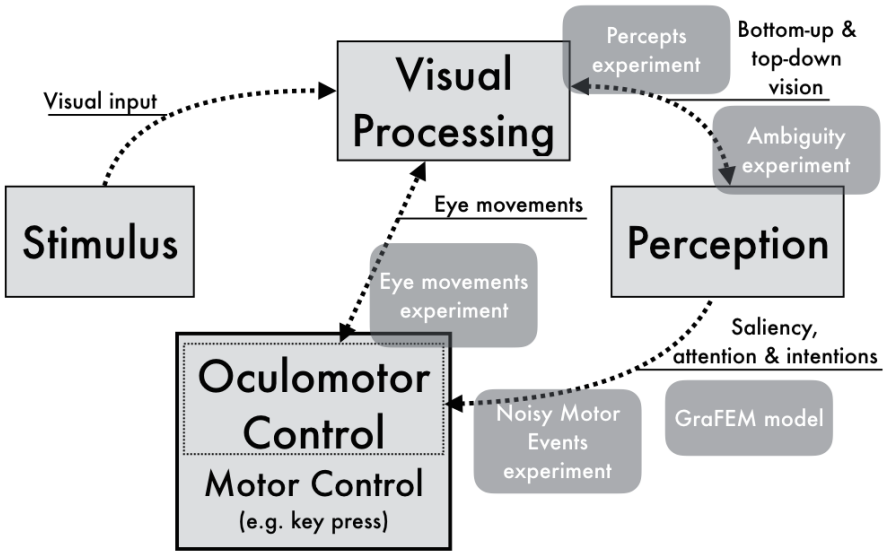


Figure 4.6. Active multi-stable perception. Diagram showing a simplified representation of the motor and perceptual systems, and their interactions for visual perception. The experiments listed are placed as they test specific interactions when observing a multi-stable moving plaid stimulus.

Note.
In this chapter, only the *Percepts* and *Ambiguity* experiments are reported because the analyses for the other experiments did not yield clear results and further investigation is needed. However, some preliminary results from the *Eye Movements* and *Noisy Motor Events* experiments are discussed in Chapter 5.

4.2 Percepts experiment: identifying the motion percepts

The basis of the empirical investigation was to identify the percepts of the moving plaid by finding recurrent states of perceived motion. For that stimulus, it has been studied in depth as a bi-stability problem, but few studies have investigated its tri-stable features.

> **Our main hypothesis for this experiment was that three stable states of perceived motion could be controlled, for the first percept, by manipulating the transparency of the gratings.**

If the visual motion can be easily characterised because it is controlled by the experiment, the perceived motion should be estimated as an elementary step before further manipulation can be applied.

Publication.
This section's work and results were published in a *GRETSI* conference paper in 2019, in french: *Modélisation de l'ambiguïté d'une multi-stabilité visuelle.*, Kevin Parisot, Alan Chauvin, Ronald Phlypo, & Steeve Zozor, *GRETSI, 2019* (Parisot, Chauvin, Phlypo, et al., 2019). The results presented here, however, expand on this publication.

4.2.1 Motivation

The moving plaid stimulus is known to be ambiguous in terms of motion direction perceived (Hupé and Rubin, 2003; Hupé and Rubin, 2004). However, as discussed in the previous section, the ambiguity emerging from an aperture problem may lead to an infinite number of directions being interpreted by the visual system (Wuerger et al., 1996). This first experiment's aim was to collect data showing that the percepts experienced are stable across participants. Another aspect was to verify that the link to the keys used on the keyboard in the literature could be justified in the case of a tri-stable plaid and subjective reporting task. Finally, a third element was to investigate the transparency parameter manipulation, for each grating.

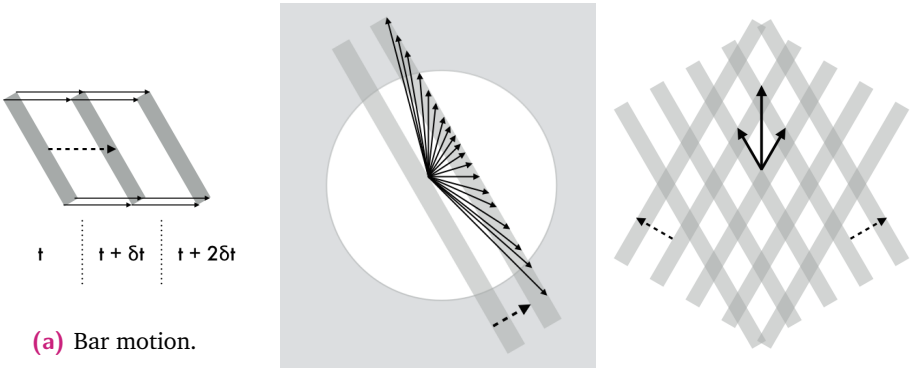
As introduced in Section 4.1, motion ambiguity arises when the most salient points, based on contrast, of a moving object are made hidden, thus reducing the capacities of the brain to infer direction of movement. In Fig. 4.7, the problem of motion ambiguity and motion direction inference in the context of the moving plaid stimulus is presented in more details. The visual system decodes a perceptual object's motion direction by using that object's higher contrast points' motion to infer overall motion

direction. For the bars composing the gratings of the moving plaid stimulus, these are the corners shown by the continuous arrows of Fig. 4.7a, but they remain hidden when the bars are viewed through an aperture, as shown by Fig. 4.7b. In the latter case, the possible inferred directions are based on the wave front, and they are numerous and link direction and velocity. Moreover, when two gratings are overlaid, as shown in Fig. 4.7c, the expected perceived motion directions are dependent on perceptual organisation: if two grids are perceived, each have the most likely direction, namely the direction orthogonal to the wave front, and if one grid is perceived, the perceived motion direction vectors are summed. This is the expected tri-stability with the left (l) and right (r) transparency percepts on one hand, and the coherency (c) percept on the other.

> **Thus our hypothesis for this experiment is that we can find three regular states in the observers' responses to perceived motion direction that follow the rules listed above. And furthermore, these states' probabilities of being perceived at stimulus onset can be manipulated by the gratings' transparencies as they impact the luminance and contrast of the visual stimulus.**

Finally, the first percept of the moving plaid stimulus is known to have particular dynamics (Hupé and Rubin, 2003). The coherency percept (c) is known to be more likely to be perceived at stimulus onset. However, it cannot dominate when the α of a grating is too low and the grating nearly disappears. Also, Hupé and Rubin (2004) showed that overall motion orientation affects the dominance of coherency in what they called the *oblique plaid effect*. Moreover, orientation preferences or biases in the visual processing of motion are also known to exist (Werkhoven et al., 1990). Thus overall motion orientation of the moving plaid should be considered to verify that reported perceived motion are not due to biases.

The transparency of both gratings is controlled by the parameters (α_L, α_R), each corresponding to the level of transparency (or alternatively of opacity) for the objects generated using the *PsychToolBox* in MATLAB (Brainard, 1997). The subscripts refer to the leftward and rightward moving gratings, respectively. When varying their value between 0 and 1, one can make the grating fully transparent and opaque, respectively (see Fig. 4.8). The manipulation of transparency can be defined in the following parameter space: $\boldsymbol{\alpha} = (\alpha_L, \alpha_R) \in [0; 1]^2$. Our hypothesis is the following; it is possible to control and bias each of the three percepts (c, l, r) by varying only the transparency parameters. However, the total luminance and contrast should be maintained constant across the screen of stimulus presentation. Indeed, this is essential as the strength of the visual signal is known to be linked to the perceptual



(a) Bar motion.

(b) Bar motion with aperture.

(c) Gratings' motion.

Figure 4.7. Grating motion.
For all these figures, dashed lines show physical motion and full lines show perceived motion.
(a) Diagram showing the motion of a bar, horizontally as shown by the dashed arrow, over three time steps δ. The continuous black arrows show how the visual system will identify the corners as the most salient points, based on contrast, and will track their motion to infer the perceptual object's motion.
(b) Diagram showing the motion of a bar implemented in the experimental code, as shown by the dashed arrow, over one iteration. A circular hole is drawn to describe the aperture problem: when the visual system has no information on the hidden corners of the bar, it uses the square wave front, as it is the most contrasted area of the stimulus, to estimate and infer motion direction. However, as shown by the multiple arrows at the central point, without the corners, the brain can infer multiple directions, with velocity of perceived motion varying depending on the direction. Note that this occurs along the entire wave front.
(c) Diagram showing the physical motion of two gratings, with different directions as shown by the dashed arrows, over one iteration. Here the most salient points, based on contrast, are the edges of the diamonds formed at the relatively transparent gratings. Two types of motion can be predicted based on perceptual depth organisation: (1) if the system infers two distinct grating objects, the system will infer motion going along the wave front with the lowest velocity as it will be the most present and central vector in the inference problem, and (2) if the system infers a single uniform grid object, the addition of the two transparent grating vectors will provide a unique displacement vector with higher velocity.

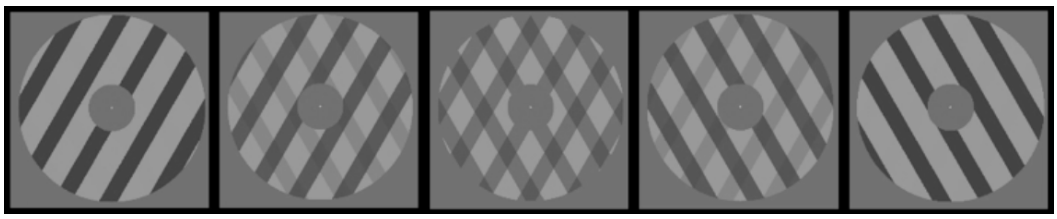


Figure 4.8. Moving plaids over α. Examples of plaids generated with varying values for transparencies: $\alpha_L = [1, 0.75, 0.5, 0.25, 0]$ and $\alpha_R = 1 - \alpha_L$.

dynamics (Levelt, 1966; Levelt, 1967). Hence, as the luminance relates directly to the strength of the visual signal and transparency affects the final luminance, the latter is a parameter for perceptual strength manipulation.

A key bias that requires verification is linked to the keys chosen to be used for subjective perceptual reports in the literature. This is linked to the absolute orientation of the stimulus and the motion. In most studies, the configuration used is the following; the gratings are set to move with constant velocities of $v_L = 1.5$ deg.s^{-1} and $v_R = -1.5$ deg.s^{-1} along the orthogonal directions of the square wave fronts (see dashed arrows in Fig. 4.7c.). Meanwhile the orientation of each gratings (θ_L, θ_R) are usually set between 0° and ±45° in a symmetric fashion about the vertical axis[2] (*for more details on gratings' orientations, read the box below, in the Methods*.

> **To summarise, in this experiment, the following two hypotheses were tested: (i) the α parameters control the first percept such that a tri-stability can be observed, and (ii) overall motion orientation of the moving plaid setup does not affect the observation of a tri-stable perception.**

4.2.2 Methods

Stimulus

The stimulus, the moving plaid, was generated using the *PsychToolBox* on MATLAB (Brainard, 1997), with a code produced and developed by the author, from scratch. The values of the parameters for the two gratings were kept constant (provided in Tab. 4.1) except for transparency (α) which was the controlled variable

[2] At $\theta = \pm 45°$ the coherency percept dominates so much that the multi-stability phenomenon disappears for most inexperienced observers. Going beyond the gratings' 45° rotation means that the coherency becomes the mono-stable percept. One can test this with the online moving plaid demo developed by J.M. Hupé: http://www.cerco.ups-tlse.fr/~hupe/plaid_demo/demo_plaids.html

and gratings' orientations ($\boldsymbol{\theta}$). To explore the entire α parameter space—in practice, we did not pick very low values for $\boldsymbol{\alpha}$ as this would translate into imperceptible gratings –, their values were selected from a uniform distribution independently for each trial:

$$\alpha_L \sim \mathcal{U}(0.1; 1)$$
$$\alpha_R \sim \mathcal{U}(0.1; 1)$$

Therefore, some trials had very contrasted gratings, such as the ones shown on the extreme left and right examples of Fig. 4.8.

Luminance and contrast control

For reminders, we refer to luminance L as the amount of light—as a numeric value of the screen's candela output for a surface area (cd.m^{-2})—generated by the screen at a spatial location, namely a pixel. Contrast C is the spatial difference between two points.

Because the manipulation of $\boldsymbol{\alpha}$ leads to variations in luminance and given there were no relationship between α_L and α_R to keep an equilibrium, an algorithm was implemented to compute the total luminance of the stimulus and was used to compensate the manipulation on the gratings by adjusting the luminance of the background (L_{BG}) and contrast. The area for each structure of the stimulus were computed. Given a pair of $\boldsymbol{\alpha}$ values, we can define the sum of the vector as: $S = \alpha_L + \alpha_R$ and product of the vector as: $P = \alpha_L \alpha_R$. Hence, the global luminance L_{all} was computed as follows:

$$L_{all} = L_{BG} + D(P \times D - S)(L_{BG} - L_G)$$

with L_G a vector of the grating mean luminance[3], L_{BG} the luminance value for the background and D the duty cycle of the grating square wave. Contrast (C) was then estimated using the geometrical properties of each object of the stimulus. The various contrast components computed are defined as follows:

$$C_G = D^2\big((1 + P - S)L_{BG} + (S - P)L_G\big)^2$$
$$C_{LR} = (1 - D)D\Big(\big(L_{BG} + \alpha_L(L_G - L_{BG})\big)^2 + \big(L_{BG} + \alpha_R(L_G - L_{BG})\big)^2\Big)$$
$$C_{BG} = (1 - D)^2 L_{BG} - L_{all}$$

[3] More precisely, a central value between the light and dark values from the square wave was given, such that the luminance of the dark component of the square was $L_- = LG - k$ and the light component was $L_+ = L_G + k$. In our experiments, we set $k = 0.15$ and $L_G = 0.35$ for the normalised luminance values, before the control was applied.

with C_G corresponding to the gratings' contributions, C_{LR} the areas of intersections' contributions and C_{BG} the background contribution. These are all compounded in the following equation.

$$C_{all} = \sqrt{C_G + C_{LR} + C_{BG}} \tag{4.1}$$

The luminance and contrast values (L_{all}, C_{all}) are then fed into an equalising function with reference values (L_{ref}, C_{ref}):

$$\Gamma = C_{ref}/C_{all} \tag{4.2}$$

that is applied to the background to compensate the variations on mean stimulus global contrast by manipulating transparency. A luminance offset (Ω) is also obtained with:

$$\Omega = L_{ref} - \Gamma L_{all} \tag{4.3}$$

The new luminance values for the two gratings and the moving plaid background are calculated using the relationship:

$$L' = \Gamma L + \Omega \tag{4.4}$$

Though the local contrasts varied, the global luminance and contrast generated remained the same in the aperture. The foreground, the fixation disk at the centre, the mouse response circle, the fixation dot and the mouse dot were kept constant.

The other variable manipulated was the orientation offset (θ_{offset}). The offset was thus also selected from a uniform distribution for each trial:

$$\theta_{\text{offset}} \sim \mathcal{U}[-\pi; \pi]$$

Given the changes of absolute orientation due to the variability of the offset value over each trial (*see the box below for more details*), the use of the keyboard was rendered impracticable to collect fine data on motion direct. Therefore, a circle was added on the periphery of the stimulus (as shown in Fig. 4.9) where the participants had to click to respond.

Gratings' orientations.
We choose $\theta_L \in [0; 45]$ degrees as orientation shifts applied on the gratings. Therefore, we define relationship of the orientation of the gratings to be symmetric such that:

$$\theta_L = -\theta_R \tag{4.5}$$

Therefore we can define the gratings' orientation difference θ_{diff} as follows:

$$\theta_{\text{diff}} = \left|\theta_L - \theta_R\right| = 2\left|\theta_L\right| \tag{4.6}$$

where 0° is set on the vertical axis in the direction of the top of the screen for simplicity of interpretation. The absolute overall orientation of the moving plaid system can be manipulated by computing the relative angles (θ_{diff}) from the gratings' angles (θ_L, θ_R) by adding an arbitrary polar offset:

$$\theta_{\text{abs}} = \theta_{\text{diff}} + \theta_{\text{offset}} \tag{4.7}$$

which can also be expressed as follows.

$$\theta_{L,\text{abs}} = \theta_L + \theta_{\text{offset}}$$
$$\theta_{R,\text{abs}} = \theta_R + \theta_{\text{offset}}$$

Procedure

Participants were asked to click with the mouse's left button in the circle area at the location in which they saw the perceived motion going, as soon as they were confident. The experiment was composed of 10 blocks of 50 trials restricted to the first percept, giving us 500 data points per participants in the $\boldsymbol{\alpha}$ space. Participants had up to 5 seconds (s) to respond on their first percept and stimulus presentation ended once the mouse button was pressed. The moving plaid stimulus appeared after an interval in which, only the central fixation area was presented and which lasted between 900 ms and 1300 ms (Fig. 4.9). This interval was established using a random distribution (in seconds):

$$t_{\text{fix}} \sim \mathcal{U}[0.9; 1.3]$$

The mouse's location on screen was displayed as a dot and reset at the beginning of trial in the centre of the screen (see Fig. 4.9). The experiment was carried out on 11 participants (6 naive, 5 women, mean age 33.27 years old with standard deviation 9.89 years), with 500 trials each. Between blocks, participants were given short breaks to rest.

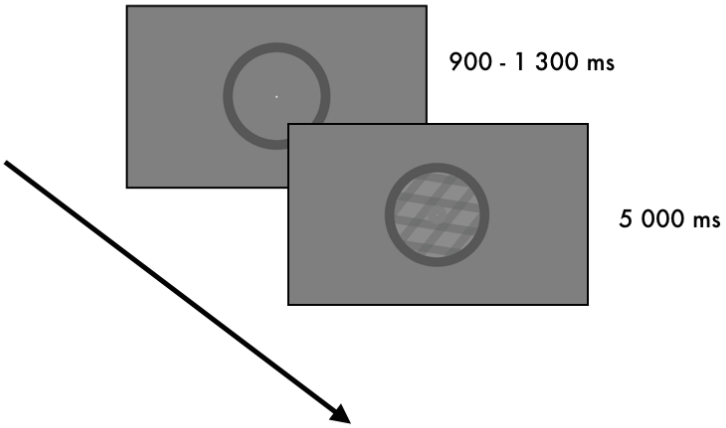


Figure 4.9. Protocol. Experimental trial protocol structure for the *Percepts* experiment. The arrow represents the direction of time.

Analysis

The mouse key press responses were pre-processed such that their coordinates in the screen plane were recovered and the angular component extracted. The random orientation offset θ_{offset} introduced was then subtracted. A Rayleigh test was carried out and if rejected, a three Von Mises distribution mixture was estimated for the data. The data was then fitted to a mixture of Von Mises distribution mix using an *expectation maximisation* method (Agostinelli and Lund, 2017; Bee, 2020):

$$f(x|\mu,\kappa) = \sum_{k=1}^{3} \beta_k \frac{e^{\kappa_k \cos(x-\mu_k)}}{2\pi I_0(\kappa_k)} \tag{4.8}$$

with $I_0(\kappa)$ the modified Bessel function of order 0, μ_k the parameter relating to the k distribution's location, κ_k the parameter relating to the dispersion of the distribution, and $\beta_k \in [0;1]$ the weight of each Von Mises in the mixture such that $\sum_k \beta_k = 1$. We used the following criteria to stop the algorithm: a maximum of 10^5 iterations or a maximum difference in log-likelihood between successive iterations of 10^{-6}. We imposed the algorithm to search for a mixture of three Von Mises distributions as expected from a tri-stable motion perception. We call estimated threshold values the approximate angular direction where two estimated Von Mises function meet.

For group data, we carried statistical tests on the estimated Von Mises mean parameters by grouping them along their orientation with respect to the vertical axis of the screen: left (l), centre (c) and right (r). We conducted a non-parametric approximate (Monte Carlo) Friedman test using a bootstrap method with 10000

permutations. We also computed approximate (Monte Carlo) Wilcoxon signed rank tests for pairwise comparisons (with a decision criterion at $\alpha = 0.05/3 = 0.017$).

Expectation maximisation algorithm.
We consider a sample of $\boldsymbol{x} = (\boldsymbol{x_1}, ..., \boldsymbol{x_n})$ of n independent observations, following a probability density function $f(x_i, \Theta)$, with Θ its unknown parameters. The aim is to determine the values of Θ by using the maximum log-likelihood given by:

$$L(\boldsymbol{x}; \Theta) = \sum_{i=1}^{n} \ln f(\boldsymbol{x_i}, \Theta)$$

Based on the assumption that an unknown data vector $\boldsymbol{z} = (z_1, ..., z_n)$ exists, we can define the complete log-likelihood as:

$$L((\boldsymbol{x}, \boldsymbol{z}); \Theta) = \sum_{i=1}^{n} \Big[\ln f(z_i | \boldsymbol{x_i}, \Theta) + \ln f(\boldsymbol{x_i}, \Theta) \Big]$$

and thus,

$$L(\boldsymbol{x}; \Theta) = L((\boldsymbol{x}, \boldsymbol{z}); \Theta) - \sum_{i=1}^{n} \ln f(z_i | \boldsymbol{x_i}, \Theta)$$

Then, the expectation maximisation algorithm is an iterative procedure based on the updated completed data with a current parameter set Θ_c and can be written as:

$$E\Big[L(\boldsymbol{x}; \Theta) | \Theta_c\Big] = E\Big[L((\boldsymbol{x}, \boldsymbol{z}); \Theta) | \Theta_c\Big] - E\Big[\sum_{i=1}^{n} \ln f(z_i | \boldsymbol{x_i}, \Theta) | \Theta_c\Big]$$

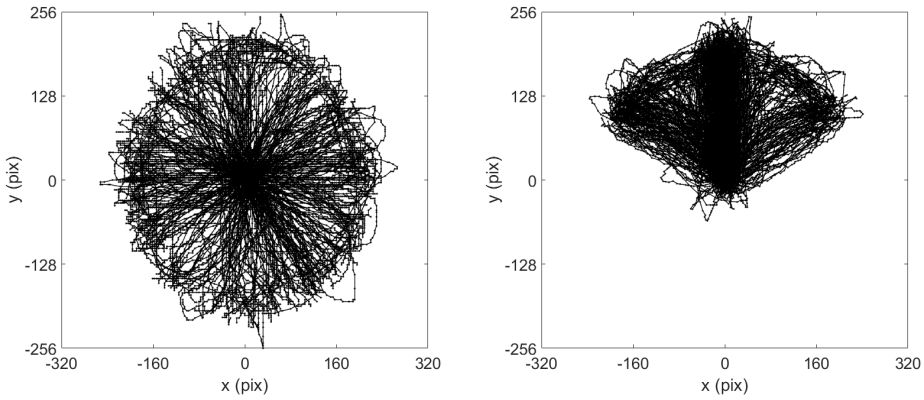
The maximisation is operated as follows:

$$\Theta_{c+1} = \arg\max_{\Theta} \Big(E\big[L((\boldsymbol{x}, \boldsymbol{z}); \Theta) | \Theta_c\big] \Big)$$

The expectation maximisation algorithm is often used in data classification, machine learning or artificial vision problems, for instance, where latent variables are the labels of the classes.

4.2.3 Results

The mouse tracking responses were analysed in terms of direction of the mouse button press with respect to the centre of screen and stimulus. A participant's response signals are shown as an example in Fig. 4.10 where one can see the trajectories of the mouse over the trial until button press before (Fig. 4.10a) and after (Fig. 4.10b) correction for θ_{off}. Interestingly, the data shows that some paths to response were highly deviated by other motion components perceived.



(a) Raw mouse responses (S1). **(b)** Corrected mouse responses (S1).

Figure 4.10. Mouse responses.
(a) Plot of all raw mouse tracking responses for S1, showing the dispersion over all orientation due to the uniform random sampling of θ_{offset}.
(b) Plot of all the mouse tracking responses for S1 once θ_{offset} is compensated. We can see that the responses concentrate in the upper part of the screen, with three main paths being systematically taken.

Von Mises density estimation

A Rayleigh test of circular uniformity was carried out and the null hypothesis was rejected ($T = 0.8436; p < 0.0001$). The estimation was done for a mixture of three Von Mises distributions (Fig. 4.11b) and we obtained the following estimated Von Mises parameters[4]: ($\mu_1 = 55.62°$; $\kappa_1 = 30.823$; $\beta_1 = 0.166$), ($\mu_2 = -0.72°$; $\kappa_2 = 40.799$; $\beta_2 = 0.666$) and ($\mu_3 = -55.62°$; $\kappa_3 = 14.751$; $\beta_3 = 0.168$). Therefore, the Von Mises mixture fit from the expectation maximisation algorithm provides a good estimation of the data in terms of how many functions are mixed and whether the use of a mixture of Von Mises distributions is justified.

Thresholds between percepts were inferred by using the relative probability density of percepts; i.e., finding the points in Fig. 4.11a where the probability densities are equal for two neighbouring percepts. Here we found the transitions—once shifted by -90°—to be at 25.56°, 142.38° and -22.50°.

A clear result that was expected and is present in this data set, is the dominance of the coherency percept: $\beta_2 = 0.666$ which is associated to the Von Mises near 0° ($\mu = -0.72°$) weights more that the other two. Another key result is the measured direction of motion perceived for the transparency percepts. Though the physical motion is horizontal in the implementation, the aperture driven ambiguity impedes

[4]Values are given in degrees and corrected for the -90° shift, such that 0° corresponds to the top of the vertical axis.

any possibility to infer the true direction of motion. However, the visual system interprets this motion's direction at -55.62°, -0.72° and 55.62°, which corresponds to a quasi perpendicular shift from the direction orientations ($\boldsymbol{\theta} = [-30, 30]$) for the transparency percepts, corresponding to the shortest line of motion in Fig. 4.7b. Meanwhile, coherency corresponds to the sum of motion vectors for both transparency percepts (longest arrow in Fig. 4.7c).

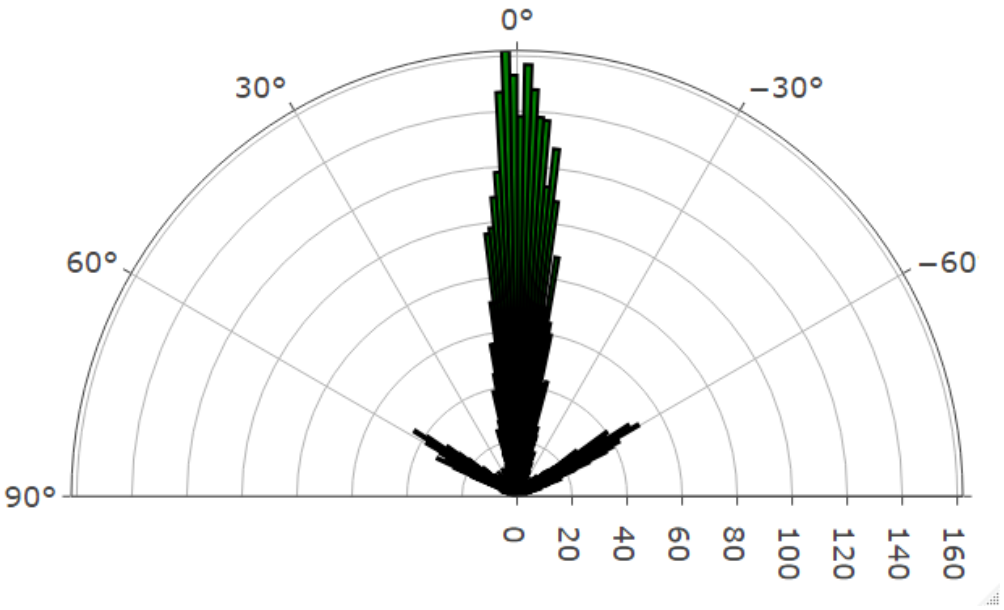
Individual & group data

We also looked at individual's data; Von Mises estimated parameters, estimated thresholds between distributions and Rayleigh tests are reported in Tab. 4.2. All participants showed non-uniform responses, based on Rayleigh tests results. However, the Von Mises estimation varied across participants with S10 reporting responses so centred near 0° that fitting three Von Mises distributions with our method led to asymmetric means, and for S11, the algorithm did not find an adequate mixture. Specifically, S11 had responses that were centred around 0°, and did not have other mode. In fact, when doing the estimation for one Von Mises distribution, the algorithm yielded the following estimated parameters: $\mu = -0.43°$; $\kappa = 45.39$. S10's data had a similar uni-modal aspect and when applying the estimation for one Von Mises distribution, the estimation yielded the following parameters: $\mu = -3.42°$; $\kappa = 11.42$.

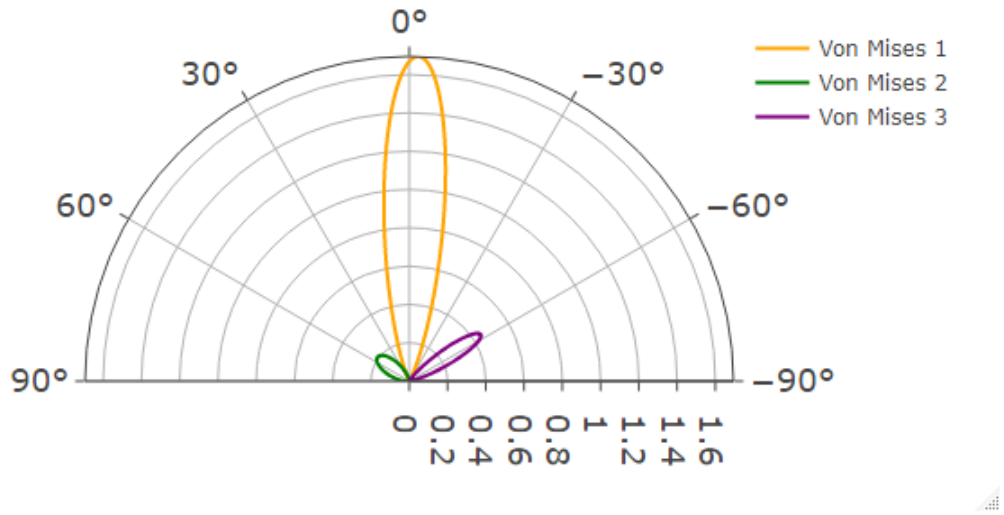
When averaging[5] participant's estimated Von Mises parameters (μ_r, μ_c, μ_l correspond to the mode on the *right*, *centre* and *left* of the vertical axis), the group had more variance over the non vertical upwards (r and l): $\overline{\mu_r} = -49.6 \pm 18.4$, $\overline{\mu_c} = -1.37 \pm 1.41$ and $\overline{\mu_l} = 65.77 \pm 44.54$. In fact, we carried out an approximate Friedman Test and found differences ($\chi^2 = 16; p < 0.0001$) when grouping the data according to clusters corresponding to the expected percepts. More specifically, approximate Wilcoxon-Mann-Whitney tests showed differences for all contrasts ($Z_{lr} = 3.3607; p < 0.0001$, $Z_{cr} = 3.3607; p = 0.0002$, $Z_{lc} = 3.3607; p < 0.0001$).

Overall, group data shows, on a small sample ($N = 8$), three different modes can be estimated, even though some participants showed unexpected responses (S10 and S11).

[5] $\overline{\cdot}$ correspond to means and numbers following $\pm$ correspond to standard deviations.



(a) Angular histogram of all participants' data.



(b) Estimated Von Mises density functions.

Figure 4.11. Von Mises estimation.
(a) Histogram of participants' responses' orientations (count for 9 participants with 500 trials each) with respect to the centre of screen. Angles have been corrected of the randomly picked orientation offsets θ_{offset}.
(b) Von Mises functions estimated (no units) using the expectation maximisation algorithm with estimated parameters ($\mu_1 = 55.62°$; $\kappa_1 = 30.823$), ($\mu_2 = -0.72°$; $\kappa_2 = 40.799$) and ($\mu_3 = -55.62°$; $\kappa_3 = 14.751$).

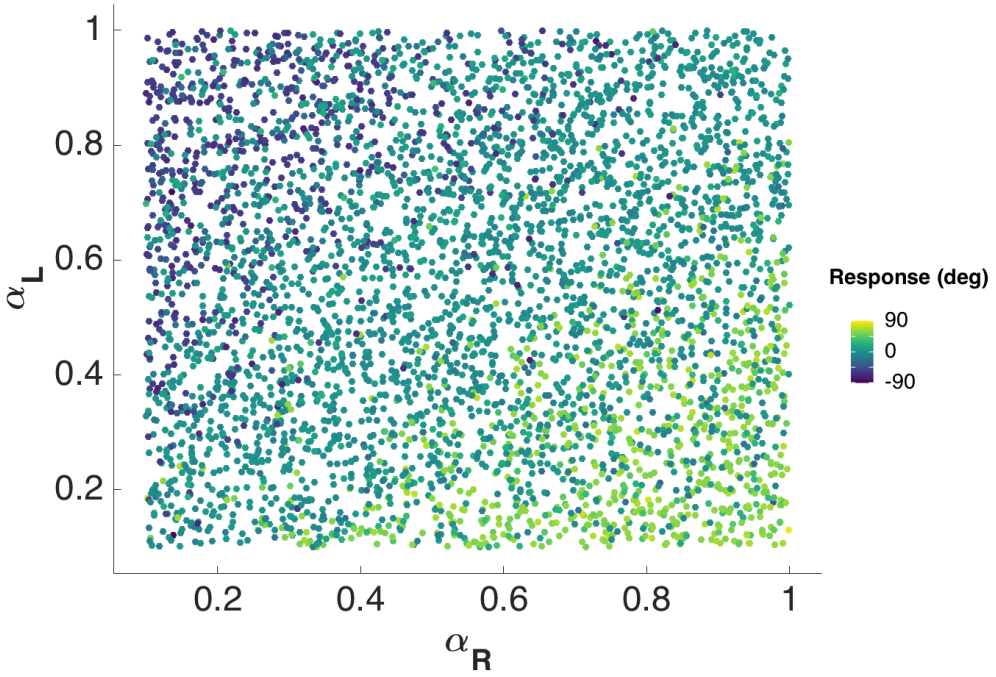
Subj.	(μ_r, μ_c, μ_l) (deg)	$(\kappa_r, \kappa_c, \kappa_l)$	Est. Thres. (deg)	Rayleigh test (T)
S1	$(-58.45, -1.74, 57.64)$	$(89.87, 88.02, 60.58)$	$(-30.23, 25.27)$	$0.778; p < 0.0001$
S4	$(-52.90, -2.53, 53.05)$	$(43.23, 48.23, 41.73)$	$(-27.09, 24.33)$	$0.9274; p < 0.0001$
S5	$(-59.93, 0.53, 49.19)$	$(74.10, 114.61, 3.42)$	$(-85.77, 12.46)$	$0.7187; p < 0.0001$
S6	$(-58.56, -3.27, 38.30)$	$(26.07, 12.98, 8.66)$	$(-34.64, 16.97)$	$0.8659; p < 0.0001$
S7	$(-45.21, -1.08, 34.97)$	$(25.56, 47.12, 13.56)$	$(-20.48, 13.68)$	$0.9383; p < 0.0001$
S8	$(-58.7, 0.81, 60.31)$	$(154.9, 31.67, 145.89)$	$(-39.74, 40.97)$	$0.825; p < 0.0001$
S9	$(-57.42, -1.14, 59.22)$	$(91.19, 36.08, 71.71)$	$(-35.95, 33.90)$	$0.6004; p < 0.0001$
S10	$(-5.66, -1.85, 173.50)$	$(6.91, 74.12, 36.71)$	$(9.48, 89.99)$	$0.9552; p < 0.0001$
S11	$\varnothing$	$\varnothing$	$\varnothing$	$0.9889; p < 0.0001$

Table 4.2. Individual data. This table provides the estimated Von Mises distributions parameters and inferred approximate estimated thresholds, as well as Rayleigh test results. $\varnothing$ refer to subjects where the estimation algorithm could not yield a result for a mixture of three Von Mises functions.

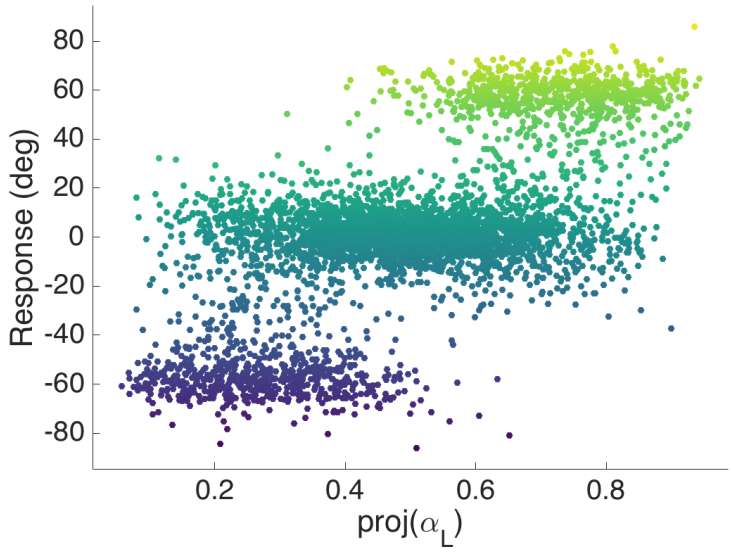
Perceived motion with respect to gratings' transparencies

The relationship to the manipulated variables, gratings' transparencies ($\boldsymbol{\alpha} = [\alpha_L, \alpha_R]$), is shown in Fig. 4.12a. The mouse button presses for each trial are represented as coloured dots—with the colour variation giving the angular direction of the perceived motion—in the transparencies (α_L, α_R) space. One can observe three clear *plateau* (i.e., clusters) corresponding to the three percepts. More interestingly, the response orientation data seems to vary little in the $\alpha_L = \alpha_R$ projection line but greatly in the $\alpha_L + \alpha_R = 1$ projection line. Therefore, the perceptual manipulation dynamics may be simplified to reduce this redundancy. Reaction times were homogeneous across the data set, with strictly positive tailed distribution.

In this data set, three clusters were identified for response orientations based on perceived motion directions. This data was projected on a sub-space of $\boldsymbol{\alpha} \in [0; 1]^2$ such that $\alpha_L + \alpha_R = 1$, and is presented in Fig. 4.12b (*read the box for more details*). The data in proj(α_L) is still organised in three clusters, suggesting that the reduced space can be investigated and manipulated to generate tri-stability. The distance between the clusters in the response orientation dimension shows that motion perception changes and stabilises in another state.



(a) Responses over entire space.



(b) Transparency sub-space projection.

Figure 4.12. Response orientation as a function of α.
(a) Direction of mouse responses' orientations, with respect to screen centre, depending on the randomly picked $\boldsymbol{\alpha} = [\alpha_L, \alpha_R]$ values during all trials. Data has been corrected so that 0° corresponds to the top of the screen, vertically. (b) Data projected on the line in the $\boldsymbol{\alpha}$ space that satisfies $\alpha_L + \alpha_R = 1$.

Data projection.
The projection is computed by finding the closest point on a line L on which a data point $\boldsymbol{b}$ will be projected. In this case, we consider L to be the diagonal in the $\boldsymbol{\alpha}$ space that satisfies $\alpha_L + \alpha_R = 1$. Its associated vector, considered is $\boldsymbol{a} = (-1, 1)$ and as such, a shift on $\boldsymbol{b}$ is operated such that $\boldsymbol{b_s} = \boldsymbol{b} + (-1, 0)$. To find the projected coordinates, one needs to find the point in $\operatorname{span} a$ closest to b. This is computationally achieved by finding the projected $\boldsymbol{p_s} = \frac{\boldsymbol{b_s} \cdot \boldsymbol{a}}{\boldsymbol{a} \cdot \boldsymbol{a}} \cdot \boldsymbol{a}$. Finally, the projected point is set back into the euclidean space to obtain $\boldsymbol{p} = \boldsymbol{p_s} - (-1, 0)$.

4.2.4 Discussion

To summarise, with this experiment:

- We confirmed that the tri-stability in the moving plaid can be linked to motion perception at 55.62° for left transparency (l), -0.72° for coherency (c) and -55.62° for right transparency (r)—Fig. 4.11—which could be fit to a mixture of three Von Mises density functions, thus providing inferred relevant boundaries at 25.56° and -22.50°.
- We showed that the transparency parameters $\boldsymbol{\alpha}$ can generate the three percepts, as the almost entire (α_L, α_R) space was uniformly explored (Fig. 4.12a).

Coherency, the more stable percept

This experiment shows that three percepts exist for motion direction perception in the moving plaid. Hence it is tri-stable when manipulating gratings' transparencies. However, individual data showed that two participants (S10 and S11) perceived motion in conditions that explained by a simpler uni-modal fit. Furthermore, estimated threshold were, for some participants (S5, S6, 10) less symmetrical along the vertical axis than for the others. This suggests that the direction of the transparency percepts (r and l) may vary more than for the coherency (c) percept. Indeed, we see similar results when looking at the standard deviations of estimate Von Mises means (μ_r, μ_c, μ_l) with little variance for the centre direction, corresponding to the coherency percept.

From two alphas to one alpha

Fig. 4.11 confirmed the presence of the three percepts and the physical basis. However, the interesting result here is the invariance described along the $\alpha_L = \alpha_R$

direction in Fig. 4.12. Indeed, the aim of this experiment was to confirm that the tri-stable nature of the percepts are linked to perceived motion. Also, it aimed to investigate whether visual manipulation of the stimulus could be controlled by the transparencies parameters $\boldsymbol{\alpha}$, in order to impact perception. The results showed that $\boldsymbol{\alpha}$ manipulation could be reduced to one of its component, e.g., α_L, while keeping the following relationship in order to remain in a reduced sub-space. The following equation can be used to reduce control parameter exploration:

$$\alpha_L + \alpha_R = 1 \tag{4.9}$$

This means that given a picked α_L, we obtain its counterpart with $\alpha_R = 1 - \alpha_L$. The practical consequences of these findings imply that it is possible to manipulate the first perception of the multi-stable moving plaid stimulus using $\boldsymbol{\alpha}$. Studies in the literature have shown that the duration of the first percept can give information about the dynamics on longer presentations (Hupé and Rubin, 2003; Hupé and Rubin, 2004; Rubin, Hupé, et al., 2005; Huguet et al., 2014). Therefore, one can expect the results of this experiment to be correlated to the evolution of perceptual dynamics, given changes to the transparencies of the gratings. The generalisation of first percept measures to continuous viewing paradigm and multi-stable perceptual dynamics was further investigated in the *Ambiguity* experiment, presented in Section 4.3.

Conclusion

> **To synthesise, this experiment aimed at, (i) finding the physical basis of the perceived motion, and (ii) investigate the manipulation of transparency and its effects on the first percept. Both of these goals were achieved and besides, a redundancy in the relationship between gratings' transparencies and percepts were found, showing a potential for simplification of the control variable in the following experiments.**

Mouse tracking was used to obtain data to answer the first point. The stimulus was implemented with luminance and contrast mechanisms to balance viewing conditions across trials, for the second point. The next step was therefore to investigate this simplified relationship between $\boldsymbol{\alpha}$ and perception. Furthermore, this experiment focused on the first percept but many questions remained about its generalisation to longer observation dynamics, which are the main interest in the phenomenon of perceptual multi-stability.

4.3 Ambiguity experiment: percept probabilities w.r.t. transparency

The *Percepts* experiment investigated the first percept but did not give answers about the dynamics over time, in a continuous viewing paradigm.

> **Hence the question: how are the perceptual dynamics affected by this stimulus manipulation over a continuous viewing of the moving plaid?**

The *Ambiguity* experiment also aimed at estimating the inter-individual differences regarding ambiguity in order to reduce potential variability that could be introduced by this manipulation. The experiment was organised in two phases[6]: a first one that reproduces the results of the *Percepts* experiment using key press, and a second one that uses the phase 1 data to model the observer in order to investigate the generalisation to continuous viewing paradigms. The section is presented following this structure for readability purposes.

Publication.
This section's work and results were published in a *GRETSI* conference paper in 2019, in french: *Modélisation de l'ambiguïté d'une multi-stabilité visuelle.*, Kevin Parisot, Alan Chauvin, Ronald Phlypo, & Steeve Zozor, *GRETSI, 2019* (Parisot, Chauvin, Phlypo, et al., 2019). The results presented here, however, expand on this publication.

Investigating the reduced parameter space

In this second experiment, the focus was on the manipulation and modulation of the moving plaid's ambiguity. Though percepts were defined and identified in the *Percepts* experiment—presented in Section 4.2—the reduction of control parameters was key such that the number of trials and the amount of time needed to complete the experiment could be kept to bearable levels for participants. If the entire gratings' transparency parameter space was to be explored, the experiment would require at least double the time. In the *Percepts* experiment, three clusters were identified for response orientations based on perceived motion directions.

Perceived state changes have been studied by psycho-physics research (Green, Swets, et al., 1966; S. Palmer, 1999). **Logistic** functions, also known as **Sigmoid** functions,

[6] In fact, the experiment was composed of three phases, though the third one will not be presented in details as it did not yield any significant or interesting results.

have been used for decades to model this type of phenomenon in perceptual studies. Their general formal expression is as follows:

$$f(x) = \frac{L}{1 + e^{-k(x-x_0)}} \tag{4.10}$$

where e is the natural logarithm base, x_0 the x-value of the Sigmoid's midpoint, L the curve's maximum value and k the logistic growth rate (Verhulst, 1838). However, for the moving plaid, there are not two, but three categories. With some modifications to the Equation (4.10), it is possible to account for that aspect of the problem—this is described in more details in this section's methods. The logistic function models classification of visual input into perceptual objects, and here as possible percepts of tri-stability. In fact, these categories can be recorded using the keys of a computer keyboard to simplify perceptual report and measuring.

Key press report for multi-stable perception.

In the literature (Hupé and Rubin, 2003), participants are typically asked to report their perception by using keyboard keys, thus applying hard discretisation on the measure[7]. However, this enables participants to report perceptual dynamics over continuous viewing of the stimulus, hence enabling multi-stable perception (see Chapter 1 for a detailed review). Methods may vary on the instruction given to participants for key press use: some research teams use the key press as an impulse variation to report perceptual change (Kornmeier, Ehm, et al., 2007) others require the observers to keep the key pressed as long as percepts are visible (Hupé and Rubin, 2004). The latter method is typically used in longer procedures as it keeps participants' motor commands active, thus this approach was considered over the former, which tends to be used in discontinuous viewing paradigms.

The results in the *Percepts* experiment concerned only the first percept, as a result of observation time constraints imposed by the protocol itself. The first percept is known to showcase special properties and be subject to strong biases, making it more predictable and reliable than the rest of the perceptual discourse (Hupé and Rubin, 2003; Hupé and Rubin, 2004; Rubin, Hupé, et al., 2005; Huguet et al., 2014; Moreno-Bote, Shpiro, et al., 2008; Moreno-Bote, Shpiro, et al., 2010; Mamassian and Goutcher, 2005). In particular, Hupé and Rubin (2003) showed that the duration of the first percept is a predictor of the empirical probabilities in continuous viewing paradigms (by a logarithmic relationship) and that it is an

[7] As a reminder, the orientation responses measured with the mouse key press in the *Percepts* experiment was also discrete, though its resolution was multiple orders of magnitude higher.

indicator of idiosyncratic differences and biases. In this experiment, an aim was to validate that the data obtained in short trials in the *Percepts* experiment, and more specifically the apparent thresholds, could be generalised to the longer viewing dynamics. This meant that, given a set of $\boldsymbol{\alpha}$ values, it is possible to estimate the empirical probability of the percepts $\hat{p}(X = x|\boldsymbol{\alpha})$. So even if the percepts change in continuous viewing, on the long run, the perceptual durations relative to the observation time reflect the psycho-physical probabilities estimated. This implied developing a **psycho-physical observer model**, based on the Bayesian framework (*see box for more details*) and using a set of rules to infer the probabilities of each percept (l, c, r) given a chosen α_L. The parameters $\boldsymbol{\Theta}$ of a model must then be estimated by a data-driven procedure, and the computation of $p(\boldsymbol{\alpha}|X, \boldsymbol{\Theta})$. That procedure should also be subject to a constraint of small duration (a few minutes), as it may be merely a step to estimate the profile of a participant in order to manipulate ambiguity in a larger protocol (presented in Section 4.1).

Building a model of the tri-stable perception relative to the transparency parameter is a powerful tool to gain quantitative insight on the characteristics of a participant. Indeed, the model can then be used to reduce uncontrolled perceptual dynamics variability in the data set that could be introduced by the use of $\boldsymbol{\alpha}$ values to bias for a percept, without accounting for potential individual internal biases. This type of approach is analogous to step wise procedures that estimate an adapted level of difficulty in psycho-physics task (Green, Swets, et al., 1966; S. Palmer, 1999). The aim is therefore to develop a custom procedure to adapt stimulus parameters to inter-individual differences, since the essential variable that requires manipulation is the level of ambiguity. This type of work has been done in psycho-physics also with Bayesian methods, though they often require a large number trials, are not tested on their generalisation to longer viewing times and tasks, and consider two category problems (Watson and Pelli, 1983; Watson, 2017).

Aims and hypotheses

> **To synthesise, the *Ambiguity* experiment aimed to (i) verify that tri-stability could be obtained using a reduced α space by using key press report, (ii) model the level of ambiguity through probabilistic methods and (iii) investigate the models' extensions and validity in longer observation durations.**

The hypothesis driving this work was the following: ambiguity when observing the moving plaid stimulus over continuous viewing paradigm can be estimated and

manipulated through probabilities. To investigate the ideas described above, a two phase experimental protocol was designed and carried out, as well as a probabilistic model. First the experimental protocol will be described, followed by the model construction and the methods for parameters' estimation.

4.3.1 Methods

The experiment was composed of two phases that contrasted the continuous and discontinuous observation paradigms. While **phase 1** was built upon the *Percepts* experiment's protocol, with the introduction of key press reporting, **phase 2** featured long trials. Furthermore, a probabilistic model was studied to infer maximal ambiguity in the control varibale space.

Psycho-physical observer model

The model is constructed on the measures of percept duration in a continuous viewing trial. For a given α_L, constant across a trial, one can measure the durations (sometimes also known as perceptual residence) that participants spent reporting the perception of each percept: t_l, t_c, t_r. These durations can be normalised so that the empirical probability of a percept (w.r.t. total reported perceptual time[8]), as a marginalisation of t:

$$\hat{p}(X = l|\alpha_L) = \frac{t_l}{t_l + t_c + t_r} \tag{4.11}$$

This equation can be used for all percepts' empirical probability measurement in a given trial. The model construction will be illustrated with the left transparency percept (l) as a basis as it can be used as an growing function with respect to α_L. Thus, the model aims to characterise $p(X = l|\alpha_L)$, the probability of observing the grating moving to the left as the closest object, given a set of grating transparency values. A first hypothesis of **symmetry** was introduced in the model: given the stimulus is symmetric along the vertical axis ($\alpha_L = 0.5$), and given the method for selection $\boldsymbol{\alpha}$ such that it follows $\alpha_L + \alpha_R = 1$. Therefore, the opposite left transparency percept can be linked to the right transparency as follows:

$$p(X = r|\alpha_L) = p(X = l|1 - \alpha_L) \tag{4.12}$$

The remaining coherency percept can then be inferred by doing the second hypothesis of the percepts being **complementary**, due to the tri-stability reducing the

[8] Note that one can also consider total trial time, which would include key press transition latencies.

problem to only three states and them being mutually exclusive (Chopin, 2012). The following expression shows the complete inter-dependency of the percepts:

$$p(X = l|\alpha_L) + p(X = c|\alpha_L) + p(X = r|\alpha_L) = 1 \tag{4.13}$$

Hence, it is possible to obtain all probabilities starting with the right transparency percept's one and by following the hypotheses that the transparency percepts are *symmetric* and all are *complementary*. The left transparency can be modelled using the Sigmoid function, a classic in psychometrics (further justified by the data in the results section when showing $p(\alpha\|X)$):

$$p(X = l|\alpha_L, \mathbf{\Theta}) = \frac{1}{1 + e^{-\Theta_1(\alpha_L - \frac{1}{2}) + \Theta_2}} \tag{4.14}$$

where $\mathbf{\Theta} = (\Theta_1, \Theta_2)$, with Θ_1 can be interpreted as the slope of the Sigmoid curve and Θ_2 the intercept-like parameter. Thus, we can derive the probability functions for the other percepts, with right transparency (r):

$$p(X = r|\alpha_L, \mathbf{\Theta}) = \frac{1}{1 + e^{\Theta_1(\alpha_L - \frac{1}{2}) + \Theta_2}} \tag{4.15}$$

and for coherency (c), by combining Equations (4.13) to (4.15):

$$p(X = c|\alpha_L, \mathbf{\Theta}) = \frac{e^{2\Theta_2} - 1}{1 + 2e^{\Theta_2} \cosh\left[\Theta_1(\alpha_L - \frac{1}{2})\right] + e^{2\Theta_2}} \tag{4.16}$$

For simpler reading, we will use the following notation in the paragraphs below: $p_l \equiv p(X = l|\alpha_L, \mathbf{\Theta})$, $p_c \equiv p(X = c|\alpha_L, \mathbf{\Theta})$ and $p_r \equiv p(X = r|\alpha_L, \mathbf{\Theta})$.

The following constraints apply for the probabilities $p(X|\alpha_L, \mathbf{\Theta}) \in [0; 1]$ and $\alpha_L \in [0; 1]$. Because Equation (4.14) provides a relationship between a percept's probability and transparency, which ultimately affects luminance, we can assume that the Sigmoid function of the probability of left transparency (l) must increase with α_L and vice-versa for the right transparency (r). Thus, on can say that, for physical reasons, $\Theta_1 \geq 0$, since Θ_1 dictates the slope of the psycho-physics function.

Also, we need $p_c \geq 0$. That is achieved when $e^{2\Theta_2} \geq 1$, i.e., $\Theta_2 \geq 0$.

Furthermore, $p_c \leq 1$, and therefore, $e^{2\Theta_2} - 1 \leq 1 + 2e^{\Theta_2} \cosh\left[\Theta_1(\alpha_L - \frac{1}{2})\right] + e^{2\Theta_2}$, i.e., $e^{\Theta_2} \cosh\left[\Theta_1(\alpha_L - \frac{1}{2})\right] \geq -1$, which is always true.

The model construction is schematised in Fig. 4.13.

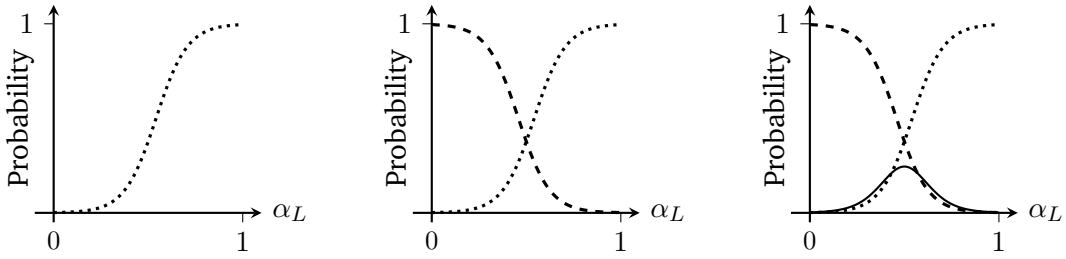


Figure 4.13. Model construction. Diagram showing the composition of the psycho-observer model. On the left, the Sigmoid function for the $X = l$ percept, see Equation (4.14). At the centre, the model when adding the $X = r$ percept using the hypothesis of symmetry between transparency percepts. On the right, the full model when adding the $X = c$ percept using the hypothesis of complementarity of the 3 percepts.

Maximal ambiguity

If we look for the α_L such that $p_l(\alpha_L) = p_c(\alpha_L) = p_r(\alpha_L) = \frac{1}{3}$, then, using the *symmetry* property, having $p_l(\alpha_L) = p_r(\alpha_L)$ requires that $\alpha_L = \frac{1}{2}$. And hence, $e^{\Theta_2} = 2$, which is possible if and only if $\Theta_2 = \ln 2$.

However, having one point of equi-probability is a very restrictive case and, in practice, never truly occurs. Therefore, a solution is to impose only that $p_c(\alpha_L) = \frac{1}{3}$, and by *symmetry*, we will find $\boldsymbol{\alpha}_\pm = (\alpha_+, \alpha_-)$ the symmetric points around $\frac{1}{2}$. If we randomly choose between (α_+, α_-) with probabilities $\Pi_+ = \frac{1}{2}$ and $\Pi_- = \frac{1}{2}$, then we obtain $\tilde{p}_c = \frac{1}{2}p_c(\alpha_+) + \frac{1}{2}p_c(\alpha_-) = \frac{1}{3}$. And also, $\tilde{p}_l = \frac{1}{2}p_l(\alpha_+) + \frac{1}{2}p_l(\alpha_-) = \frac{1}{2}p_l(\alpha_+) + \frac{1}{2}p_l(1-\alpha_+) = \frac{1}{2}p_r(1-\alpha_+) + \frac{1}{2}p_r(\alpha_+) = \tilde{p}_r$, which means that $\tilde{p}_l = \tilde{p}_r = \frac{1}{3}$. In this more realistic case, solving for coherency $p_c(\alpha_L) = \frac{1}{3}$ (Equation (4.16)), we obtain the two points:

$$\boldsymbol{\alpha}_\pm = \frac{1}{2} \pm \frac{\cosh^{-1}(e^{\Theta_2} - 2e^{-\Theta_2})}{\Theta_1} \tag{4.17}$$

Which exist if, and only if, $e^{\Theta_2} - 2e^{-\Theta_2} \geq 1$, i.e., $\Theta_2 \geq \ln 2$. Moreover, due to the constraint $\alpha_\pm \in [0; 1]$, we must have $\frac{\cosh^{-1}(e^{\Theta_2} - 2e^{-\Theta_2})}{\Theta_1} \leq \frac{1}{2}$, which imposes: $\Theta_2 \leq \ln\left(\frac{\cosh\left(\frac{\Theta_1}{2}\right) + \sqrt{\cosh\left(\frac{\Theta_1}{2}\right)^2 + 8}}{2}\right)$. For a visualisation, see Fig. 4.14.

The $\boldsymbol{\alpha}_\pm$ are thus considered as the **maximal ambiguity** points $\boldsymbol{\alpha}_{amb}$ such that $\boldsymbol{\alpha}_\pm = \boldsymbol{\alpha}_{amb} = (\alpha_+, \alpha_-)$. At such points, we expect to observe tri-stable equi-probability when the (α_+, α_-) are picked randomly with probabilities $\Pi_+ = \Pi_- = \frac{1}{2}$, respectively.

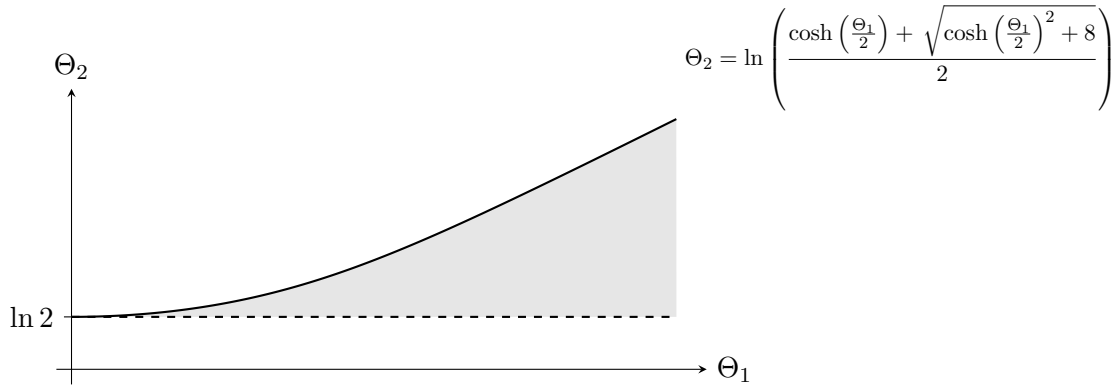


Figure 4.14. Maximal ambiguity space. Domain (Θ_1, Θ_2) allowing to derive $\alpha_\pm$ as specified in the text.

Participants

The experiment was carried out on 16 participants (12 naive, 11 females, mean age of 27.4 years old with standard deviation of 9.6 years) and participants were given a resting break between the two phases. All the naive participants signed a written form that follow the guidelines of the Declaration of Helsinki.

Stimulus

The stimulus used was the moving plaid, presented without any global orientation offset manipulation, unlike the first experiment (see Section 4.2), i.e., $\theta_{\text{offset}} = 0$ over all trails. Thus, the coherency percept was aligned in the vertical axis—here considered as 0°, to simplify data interpretation–, with motion of the percept towards the top of the screen. The percepts were associated to the arrow keys on a keyboard, with the left, up and right arrows corresponding to the left transparency (l), coherency (c) and right transparency (r) percepts, respectively. X is the random variable with its space of realisation $\{c, l, r\}$, measured by the key presses. The transparency parameters were manipulated with $\alpha_L \in [0; 1]$ and its counterpart being defined by the relation of Equation (4.9): $\alpha_L + \alpha_R = 1$. All the other moving plaid parameters had the same values as in the *Percepts* experiment (described in Section 4.1 and Section 4.2).

Protocol

Phase 1: short trials. Participants were asked to report their first percept in trials similar to those of the *Percepts* experiment (see Fig. 4.15**A**). A jittered fixation interval was displayed before the moving plaid's appearance and participants were asked to keep their eyes fixed on the dot at the screen's centre. Trials ended when participants pressed a response key press using one of the three arrows of the keyboard designated. The transparency parameters of the gratings ($\boldsymbol{\alpha}$) were randomly selected, using a uniform distribution ranging from 0 to 1:

$$\alpha_L \sim \mathcal{U}[0;1]$$

and the symmetric grating's (G_R) transparency was set by $\alpha_R = 1 - \alpha_L$. The phase was composed of a 200 trials block, with participants being free to rest their eyes between trials, if needed, but remained seated on a chin rest.

Phase 2: long trials. The second phase was composed of 10 trials that lasted 120 seconds, in which participants were asked to continuously report their perception using the same arrows of the key board as in phase 1, and by maintaining the key pressed while the percepts lasted (see Fig. 4.15**B**). Participants chose when to launch the trials, giving them the possibility to rest in between, if needed. A jittered fixation interval remained, as in phase 1, and participants were asked to keep their eyes fixed on the dot at the screen's centre. $\boldsymbol{\alpha}$ values were constant over the trial duration and were still chosen using the $\alpha_L + \alpha_R = 1$ relationship, but α_L was picked using ambiguity parameters inferred using the model. This is detailed further down in the model section.

Parameter estimation

In **phase 1**, we considered Bayes law (*see the box below, Equation* (4.24)) to justify the choice of using a Sigmoid function (Equation (4.14)), given the empirical observation can be linked to the theoretical probability law of the percept by $f(\alpha_L|X) \propto p(X|\alpha_L)$, since here, the uniform sorting meant $p(\alpha_L) = 1$. $f(\alpha_L|X)$ is estimated by a mixture of truncated Gaussian distributions described further down. Therefore, at the end of the phase, we estimated the $\boldsymbol{\Theta}$ parameters using a maximum likelihood estimator.

To estimate the model's parameters given a data set $\boldsymbol{x} = (\boldsymbol{x_1}, ..., \boldsymbol{x_n})$ of $n = 200$ independent observations from a participant, we used a maximum log-likelihood

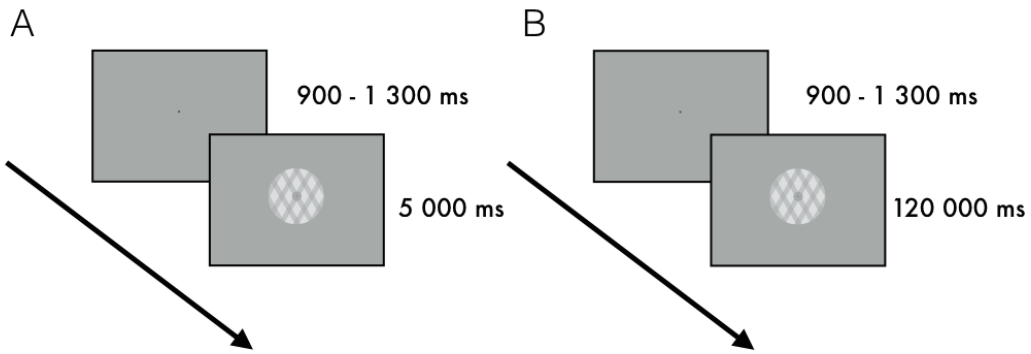


Figure 4.15. Protocol.
A: Phase 1 short trial timeline of stimulus presentation.
B: Phase 2 long trial timeline of stimulus presentation.

estimator. The method consists in finding the set of parameters $\boldsymbol{\Theta}$ that yield the maximum log-likelihood, given the first percept observations $\boldsymbol{x}$ with their associated picked $\alpha_{L,i}$ for first percept response trials (phase 1: $n = 200$). Here, we fed 200 observations of $(\alpha_{L,i}, x_i)$ into:

$$L(\boldsymbol{x}; \boldsymbol{\Theta}) = \sum_{i=1}^{n} \ln \left[p(X = x_i | \alpha_{L,i}, \boldsymbol{\Theta}) \right] \tag{4.18}$$

Finding the log-likelihood function's (L) maximum then becomes an optimisation problem with:

$$\boldsymbol{\Theta_{ML}} = \arg\max_{\Theta} L(\boldsymbol{x}; \boldsymbol{\Theta}) \tag{4.19}$$

In **phase 2**, however, as participants had continuous viewing of the stimulus with a constant α_L control variable, but provided key press responses reporting the duration of perception, we extracted fractions of dominance as defined by Equation (4.11). Thus, for each trial, three pairs associating $(\alpha_{L,i}, \hat{p}(x_i|\alpha_{L,i}))$ were fed to the maximum likelihood estimator to update the model such that $\alpha_{L,i+1} \approx \alpha_{amb}$. We chose the parametric family by the functions for empirical observations to be defined by a multinomial[9] function.

$$\hat{p}(\boldsymbol{x}|\boldsymbol{\alpha}, \boldsymbol{\Theta}) = \frac{T!}{t_l! t_c! t_r!} \left[p(\boldsymbol{x_i} = \boldsymbol{l} | \alpha_{L,i}, \boldsymbol{\Theta}) \right]^{t_l} \left[p(\boldsymbol{x_i} = \boldsymbol{c} | \alpha_{L,i}, \boldsymbol{\Theta}) \right]^{t_c} \left[p(\boldsymbol{x_i} = \boldsymbol{r} | \alpha_{L,i}, \boldsymbol{\Theta}) \right]^{t_r} \tag{4.20}$$

t_l, t_c, t_r correspond to the fraction of dominance (i.e., the number of realisation, if we assume independence) for each first percept, in phase 2 trials of duration T.

[9] We assume percepts are independent, even though this is unlikely.

Thus, a psycho-physical observer model and an associated method for estimating its parameters were implemented in the protocol. This required that the procedure be divided in a sampling phase for inference, and a testing phase to verify the generalisation.

Maximum likelihood estimation.
The maximum likelihood estimation aims at inferring the parameters that best explain a set of samples, more precisely the joint probability distribution of random variables $\{y_1, y_2, ...\}$, which are not necessarily independent and identically distributed—in our case, we have sample independence, but they are not identically distributed. A unique vector $\vartheta = [\vartheta_1, \vartheta_2, ..., \vartheta_k]^T$ of parameters that index the probability distribution within a parametric family $\{f(.;\vartheta)|\vartheta \in O\}$—where O is the parameter space, a finite-dimensional subset of Euclidean space—is associated with each distribution. One can evaluate the joint density for an observed set of data samples: $\boldsymbol{y} = (y_1, y_2, ..., y_n)$. A real-valued function, called the likelihood function, can then be obtained as:

$$L_n(\hat{\vartheta}) = L_n(\vartheta; \boldsymbol{y}) = f_n(\boldsymbol{y}; \vartheta) \tag{4.21}$$

For the maximum likelihood estimation, the aim is to find the values of the model parameters that maximise this likelihood function over the parameter space. Formally, this can be written as:

$$L_n(\hat{\vartheta}; \boldsymbol{y}) = \sup_{\vartheta \in O} L_n(\vartheta; \boldsymbol{y}) \tag{4.22}$$

Online use of the model

Here, using the Bayesian framework (*read the box below for a reminder*), parameter estimation was done at the end of phase 1, with the entire data set of first percept observations $\boldsymbol{x}$. Each data entry had a unitary weight in the estimation procedure input. But for phase 2, the estimation was carried out using $\hat{p}(X = x|\alpha_L)$, and for one trial, one probability per percept was fed, making the observation weighted by their observed probabilities[10]. The model updating was implemented by adding the likelihood with currently estimated parameters at a given trial in phase 2, which provided a prior. The Bayesian update can be expressed as:

$$p(X = x|\alpha_L)p(\alpha_L) = f(X = x, \alpha_L) = f(\alpha_L|X = x)p(X = x) \tag{4.23}$$

[10]Note that this is a bias as phase 1 data will have much more weight than phase 2 data. However, to counter act this, one would need to define $p(X|\alpha_L, t)$ with t a time duration of observation. Here, we did not carry out such generalisation.

where in phase 1, we read the equation from right to left, and in phase 2, from left to right. We can say that $f(\alpha_L|X = x) \propto p(X = x|\alpha_L)$ provides a justification for choosing the Sigmoid function family in our model.

Phase 1. First, $f(\alpha_L|X = x)$ is estimated using a truncated Gaussian kernel, as the probability function for $f(\alpha_L)$ is uniform over $[0; 1]$. Given that $p(X = x)$ is a normalisation term, we can derive the probability $\hat{p}(X = x|\alpha_L)$.

Phase 2. Now if we take $\Pi(\alpha_L)$ as a mixture of two Gaussian functions ($\mu = \alpha_\pm; \sigma = 0.05$; approximating a Dirac on α_+ or α_- chosen randomly with $\Pi_+ = \Pi_- = 1/2$) centred each on α_+ and α_-, and chose α_L according to this probability function, we can observe the observers' response over time $X_t = x$. Thus, by *repeating* the process, and by considering the law of large numbers, we obtain an estimation of $p(X = x|\alpha_L)$, here by using the faction of dominance of percepts (Equation (4.11)).

Therefore, if we also setup the family of laws as $p(X = x|\alpha_L, \boldsymbol{\Theta})$ and $f(\alpha_L|X = x, \boldsymbol{\Theta})$, as well as $\Pi(\alpha_L|\boldsymbol{\Theta})$, we obtain a model that can be updated with data from first percept (phase 1) and continuous viewing (phase 2) responses.

Bayesian models.
As a reminder of Chapter 1, Bayes' rule stipulates that

$$p(\boldsymbol{\Theta}|\boldsymbol{X}) = \frac{p(\boldsymbol{X}|\boldsymbol{\Theta})p(\boldsymbol{\Theta})}{p(\boldsymbol{X})} \tag{4.24}$$

where $\boldsymbol{X}$ is a set of data (i.e., sensory information) and $\boldsymbol{\Theta}$ is a set of parameters, here considered as random variables. $p(\boldsymbol{\Theta})$ is the *prior* which corresponds to the probability that the brain have such a state, independent of the sensory information. $p(\boldsymbol{X}|\boldsymbol{\Theta})$ is the conditional probability, i.e., the sensory evidence, of observing the sensory inputs given the current state of the system, also referred to as the *likelihood* or sampling distribution. $p(\boldsymbol{X})$ is the marginal probability that normalises $p(\boldsymbol{\Theta}|\boldsymbol{X})$, the *posterior* distribution which corresponds to the probability of the brain being in a state, defined by a set of combinations of $\boldsymbol{\Theta}$, given the sensory input.

Analyses

For the analyses listed below, the parameter estimations for the models were done offline, with a nonlinear programming solver from *MATLAB* which can address multidimensional unconstrained minimisation problems more efficiently than the optimisation process we applied to our original, experiment maximum likelihood

estimator. It is based on the simplex method for function minimisation proposed by Nelder and Mead (1965). In the results, when the data was analysed using the model estimation with this optimisation method, it will be referred to as **post-hoc**. However, in the experiment, for the online estimation in phase 2, a Newton-Raphson method was used.

Empirical probability density estimation. We represent the histogram of obtained responses for each percepts by estimating a probability density using a density kernel. Because the data is in a bounded interval $[0; 1]$, it is not possible to do a convolution of Gaussian distributions. However, to obtain these probability densities, we used a mixture of truncated Gaussian laws centred on the samples with weights $1/n$ for each of the n samples. The only free parameters is the standard deviation of the Gaussian laws σ. Thus, we applied a truncated Gaussian law centred on $\alpha_{L,i}$ with standard deviation $\sigma = 0.2$, also defined as $f_{\text{Gauss}}(X|\alpha_{L,i}, \sigma = 0.2)/p(X \in [0; 1]|X) \approx f_{\text{Gauss}}(;|\alpha_{L,i}, \sigma = 0.2))$. The resulting visualisation is analogous to a box histogram.

First percept model cross-validation. The models inferred at the end of phase 1 (short trials) are evaluated using a cross-validation with 50% of the phase 1 data being randomly picked and used for training, and the rest for test. The log-likelihood—Equation (4.18)—was used on both training, testing, and a null hypothesis set (in which percept labels were randomly sorted). Approximate Wilcoxon-Mann-Whitney tests using a bootstrap method with 10000 permutations was carried out using the *coin* package on *R* (Hothorn, Hornik, van de Wiel, et al., 2008). This test was chosen as it is non-parametric (the log-likelihood cannot be positive) and it allows to compare two samples' medians for nominal categories with non-normal quantitative measures. Thus, the estimation of the model's parameters can be validated if the median log-likelihood value for the train and test data sets are closer to 0 than the null hypothesis (H0).

First percept model to continuous viewing model cross-validation. For the analysis of continuous viewing perceptual durations, all epochs with no key pressed, two or more keys pressed were removed. Thus, empirical probabilities for percepts were computed based on Equation (4.11).

To validate whether the model was relevant in phase 2 trials, when stimulus observation was long and continuous, the **Kullback-Leibler divergence** (D_{KL}) between

theoretical probabilities issued by the model $p(X|\boldsymbol{\alpha} = [\hat{\alpha}_L, \hat{\alpha}_R]; \boldsymbol{\Theta})$ and the observed empirical probabilities $\hat{p}(X|\boldsymbol{\alpha} = [\hat{\alpha}_L, \hat{\alpha}_R]; \boldsymbol{\Theta})$ was computed. The divergence between the two distributions can be expressed as:

$$D_{KL} = -\sum_{x=1}^{3} \left[\hat{p}(x_i|\boldsymbol{\alpha}) \ln \left(\frac{p(x_i|\boldsymbol{\alpha}; \boldsymbol{\Theta})}{\hat{p}(x_i|\boldsymbol{\alpha})} \right) \right] \tag{4.25}$$

The divergence was computed for the each subject's phase 2 trial values, and averaged for group comparisons. For the null hypothesis (H0), all 6 combinations of percept label permutation for $p(x_i|\boldsymbol{\alpha}; \boldsymbol{\Theta})$ were used for each trial. Approximate Wilcoxon-Mann-Whitney tests using a bootstrap method with 10^5 permutations was carried out using the *coin* package on *R*. This test was chosen as it is non-parametric (the divergence cannot be negative) and it allows to compare two samples' medians for nominal categories with non-normal quantitative measures. Hence, to validate the model, the computed Kullback-Leibler divergence would need be lower than for the null hypoethsis (H0).

Continuous viewing trials. We used the empirical probabilities to compute the models' entropy over trials, in phase 2, to investigate the stability and convergence of the method, towards a maximally ambiguous perceptual phenomenon. Shannon entropy was computed using the following classic form:

$$H = -\sum_{i=1}^{3} \hat{p}(X = x_i|\boldsymbol{\alpha}) \ln \left(\hat{p}(X = x_i|\boldsymbol{\alpha}) \right) \tag{4.26}$$

with H the computed entropy, x_i the percept's observations for the i^{th} percept out of 3. If equi-probability over the 10 long trials was to be achieved, participants would showcase behaviours that would generate empirical probabilities such that the entropy $H = -\ln(3) = 1.10$, the maximal entropy for a three state system such as in tri-stable perception. Mean entropy values and standard deviations are reported in a table in the results section and provide a description of individual behaviours and their models' relative ambiguity control.

4.3.2 Results

Short trial data description

Empirical probability were estimated based on the continuous histogram method using Gaussian kernels above. The data was fed into the model to infer a set

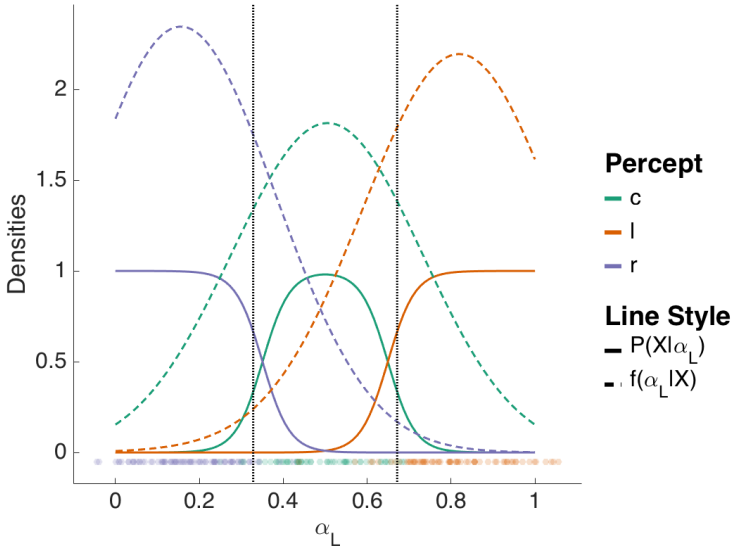


Figure 4.16. Empirical data & model example. Density functions are displayed in this figure. It shows an example resulting empirical $f(\alpha_L|X)$ density estimation using a bounded convoluted Gaussian kernel and the theoretical probability densities $p(X|\alpha_L)$ computed, based on the parameters estimated from phase 1 data using maximum log-likelihood, for participant S15. Dashed lines represent the empirical density functions, while the continuous lines represent the estimated model's probability densities. And the dotted black lines represent the inferred α_+ and α_- points of maximal ambiguity $\boldsymbol{\alpha_{amb}}$. Data points collected in phase 1 are displayed below.

of $\boldsymbol{\Theta}$ parameter for an observer, and probability density function are displayed. Fig. 4.16 shows an example of such computation for participant S1, with dashed line representing the empirical density function. Data points collected in phase 1 are displayed below. Qualitatively, the sigmoidal model densities, issued by the simplex method maximum log-likelihood estimation, proposes an interesting fit to the empirical data—note that the Gaussian mixture, as it is constructed is biased towards the centre of the α_L space. Here, the empirical density function suggested that $p(\alpha_L|X) = p(X|\alpha_L)p(\alpha_L)$ and thus $p(\alpha_L|X) \propto p(X|\alpha_L)$. It further provides clear threshold values, with rapid drops. In fact, this differed from the Newton-Raphson optimisation algorithm used and report in Parisot, Chauvin, Phlypo, et al. (2019)—a comparison is given below in Tab. 4.3. In this older work, the optimisation algorithm was more susceptible to remain stuck in a local log-likelihood maximum and thus, the parameters estimated showed less steep Sigmoid slopes (Θ_1).

Fig. 4.17 shows empirical and theoretical densities for all participants and Tab. 4.3 provides maximal ambiguity points $\boldsymbol{\alpha_{amb}}$ associated to the parameters estimated using the *post-hoc* and the original methods. S1 only had complex solutions for

$N = 16$	Post-hoc	Post-hoc	Parisot et al., 2019	Parisot et al., 2019
Participant	$\boldsymbol{\alpha_{amb}} = (\alpha_+, \alpha_-)$	$\boldsymbol{\Theta} = (\Theta_1, \Theta_2)$	$\boldsymbol{\alpha_{amb}} = (\alpha_+, \alpha_-)$	$\boldsymbol{\Theta} = (\Theta_1, \Theta_2)$
S1	$\varnothing$	$(28.11, 3.23)$	$\varnothing$	$(7.04, 0.90)$
S3	$(0.73, 0.27)$	$(38.46, 8.33)$	$(0.76, 0.24)$	$(8.34, 1.57)$
S9	$(0.73, 0.27)$	$(19.32, 3.72)$	$(0.71, 0.29)$	$(8.86, 1.31)$
S10	$(0.69, 0.31)$	$(11.53, 1.64)$	$(0.70, 0.30)$	$(7.92, 1.15)$
S11	$(0.90, 0.10)$	$(27.72, 10.30)$	$(0.99, 0.01)$	$(10.59, 4.47)$
S12	$(0.82, 0.18)$	$(25.56, 7.51)$	$(0.85, 0.15)$	$(10.20, 2.92)$
S13	$(0.63, 0.37)$	$(21.56, 2.13)$	$(0.50, 0.50)$	$(7.07, 0.65)$
S14	$(0.73, 0.27)$	$(24.92, 4.98)$	$(0.76, 0.24)$	$(8.28, 1.59)$
S15	$(0.67, 0.33)$	$(31.01, 4.63)$	$(0.66, 0.34)$	$(7.45, 0.93)$
S16	$(0.72, 0.28)$	$(19.46, 3.55)$	$(0.71, 0.29)$	$(8.37, 1.24)$
S17	$(0.63, 0.37)$	$(26.32, 2.78)$	$(0.55, 0.45)$	$(7.17, 0.71)$
S18	$(0.75, 0.25)$	$(38.04, 8.73)$	$(0.75, 0.25)$	$(9.05, 1.68)$
S19	$(0.50, 0.50)$	$(0.03, 0.36)$	$(0.50, 0.50)$	$(6.51, 0.67)$
S20	$(0.78, 0.22)$	$(18.96, 4.55)$	$(0.81, 0.19)$	$(9.15, 2.19)$
S21	$(0.58, 0.42)$	$(22.56, 1.31)$	$(0.50, 0.50)$	$(6.17, 0.44)$
S22	$(0.70, 0.30)$	$(16.23, 2.54)$	$(0.63, 0.37)$	$(8.02, 0.88)$

Table 4.3. Model estimation over short trials. The resulting maximal ambiguity points given the estimated $\boldsymbol{\Theta}$ parameters for each participant using the *post-hoc* method and the original method from Parisot, Chauvin, Phlypo, et al. (2019). $\varnothing$ corresponds to data sets with no $\boldsymbol{\alpha_{amb}}$ solutions in $\mathbb{R}$.

$\boldsymbol{\alpha_{amb}}$, while S19's parameters do not respect the constraints described above in the methods section. Overall, the *post-hoc* method finds solutions with much higher Θ_1 values, meaning that the slopes in the probabilistic models are steeper, and that the optimisation algorithm finds higher log-likelihood maximums[11].

Cross-validation of first percept models

Cross-validation, for models based on phase 1 data, was done by splitting the 200 first percept responses in two equal parts, with data points being selected randomly in the set. Log-likelihood values were computed on training, testing and randomly mixed labels (H0) sets and are reported in Fig. 4.18a. Approximate Wilcoxon-Mann-Whitney tests showed no difference between train and test data sets over all participants ($Z = -1.4707$; $p = 0.1514$), while differences between train and H0 were significant ($Z = -3.5311$; $p < 0.0001$). Differences between test and H0 were also significant ($Z = -3.5932$; $p < 0.0001$). These results validate the stability of the models for phase 1 data sets.

[11] Though the implementation searches a minimum on a positive transformation the log-likelihood.

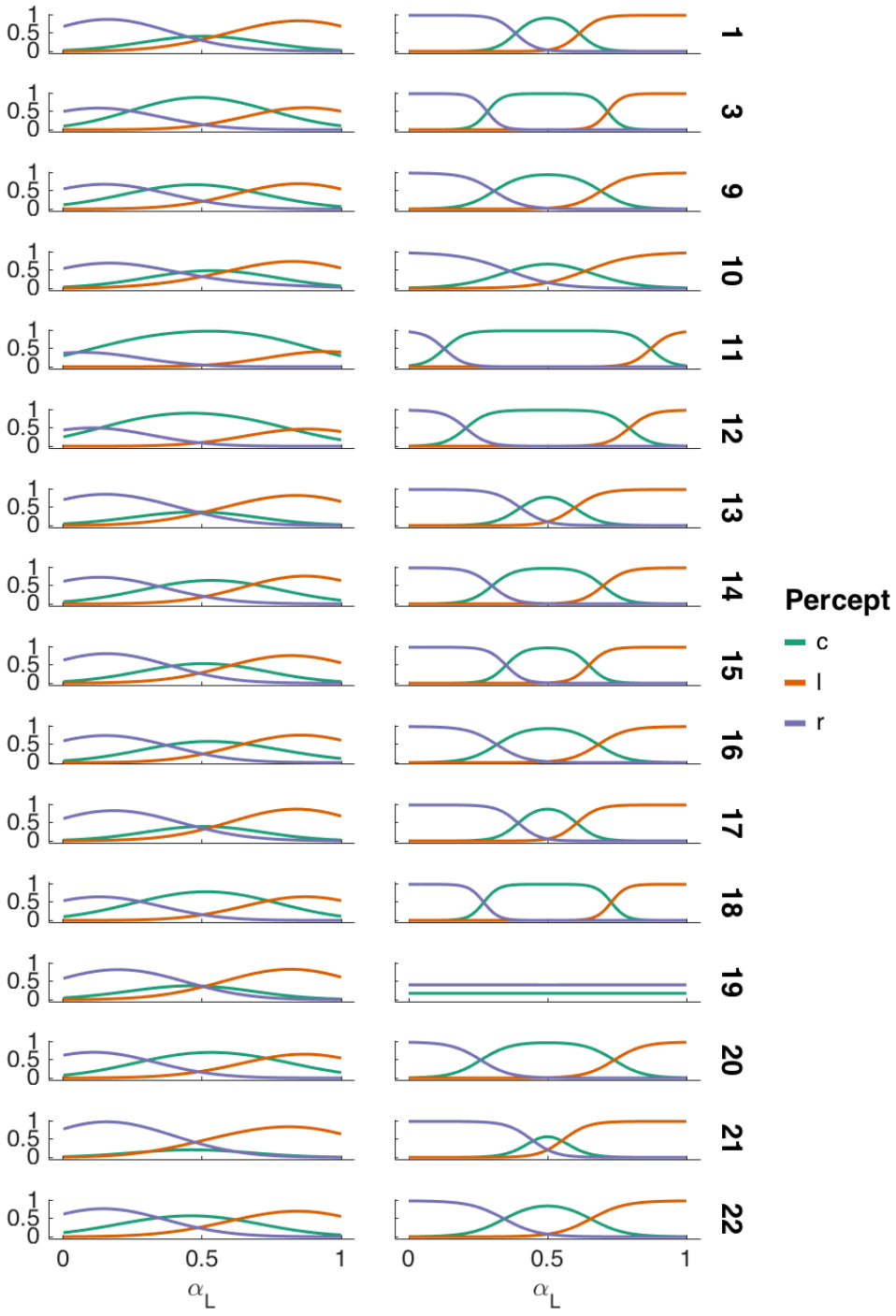


Figure 4.17. Empirical data & model. *Left column*, the $p(\alpha_L|X)$ (normalised) densities using the empirical probability density estimation method. *Right column*, the $p(X|\alpha_L)$ using the observer model after estimating $\boldsymbol{\Theta}$ for each participant over the 200 short trial data set, using the *post-hoc* approach. Subject numbers are listed on each row.

Cross-validation of model generalisation to continuous viewing

An aim of the experiment was to investigate whether first percept data, when used to infer the parameters of a probabilistic model, would generalise to longer viewing conditions. To verify whether this worked in our data set, we computed the Kullback-Leibler divergence on observed data in phase 2 and on model distribution inferred on phase 1 data, using the *post-hoc* optimisation method. We compared these values to divergence when the model probabilities were randomly shuffled; a null hypothesis condition. An approximate Wilcoxon-Mann-Whitney test was carried out to verify if group medians differed and revealed a significant difference between data and H0 ($Z = -3.6228$; $p < 0.0001$). The data is shown in Fig. 4.18b.

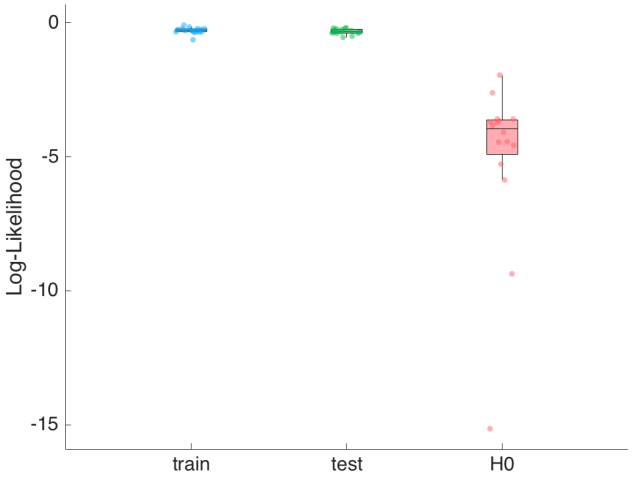
Continuous viewing data description

Entropy was computed on the empirical probabilities for percepts in the phase 2 data set. The $\boldsymbol{\alpha}$ values were selected based on the models' probabilities and the derived maximal ambiguity points $\boldsymbol{\alpha}_{amb}$, using a truncated Gaussian distribution centred on the latter. The computed entropy for each participants provides a quantitative measure of the distance to equi-probability, where $H = \ln(3) = 1.1$ can be expected, and are reported in Tab. 4.4. Group mean entropy was at 0.53 with mean standard deviation at 0.21. The data shows that the procedure led to very different perceptual dynamics across the observer population. S11 and S19 displayed bi-stable behaviour and remained stuck as far from equi-probable tri-stability as possible. Moreover, the empirical mean entropy derivative is reported in Tab. 4.4, giving information on the evolution of entropy over phase 2. Most participants had positive slopes and finished with higher entropy than at the start of the phase.

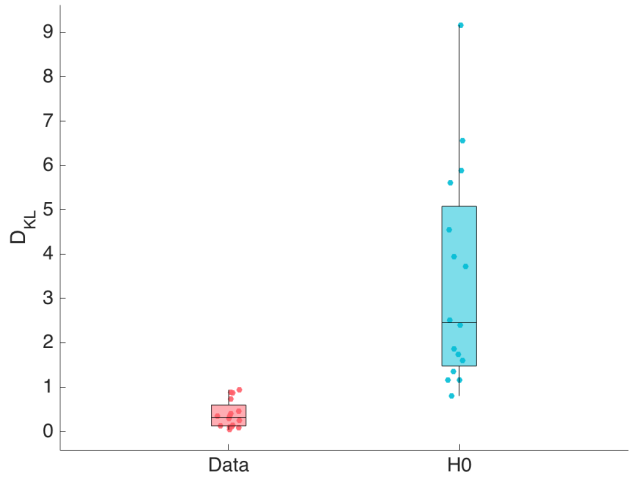
4.3.3 Discussion

To summarise, in this experiment:

- We showed that the manipulation of ambiguity and tri-stability can be achieved efficiently by using the $\alpha_L + \alpha_R = 1$ sub-space of the $\boldsymbol{\alpha}$ by replicating the results from the *Percepts* experiment with a key press method.
- We proposed a probabilistic model to account for individual differences and infer the points of maximal ambiguity in the α_L space. The model parameters were estimated using maximum likelihood and cross-validation tests for both phase 1 and phase 2 data sets provided evidence of stability and coherency.



(a) Log-likelihood cross-validation.



(b) Kullback-Leibler divergence cross-validation.

Figure 4.18. Cross-validation.
(a) Box-plot of log-likelihood computed over train, test, and H0 data sets show the stability of the model over phase 1 data. Individual data points are shown as points.
(b) Box-plot of Kullback-Leibler divergences (D_{KL}) computed over the data and H0 show the stability of the model over phase 2 data and phase 1 estimation. Individual data points are shown as points.

Participant	$\mu(H)$	$\sigma(H)$	$\mu(\Delta H)$	$\mu(\hat{p}_l)$	$\mu(\hat{p}_c)$	$\mu(\hat{p}_r)$
S1	0.82	0.27	0.0552	0.24	0.31	0.45
S3	0.43	0.37	−0.0040	0.24	0.19	0.57
S9	0.80	0.37	0.1119	0.42	0.22	0.37
S10	0.68	0.34	0.0387	0.2	0.25	0.55
S11	0	0	0	0.5	0	0.5
S12	0.62	0.05	−0.0128	0.47	0.34	0.19
S13	0.98	0.17	0.0626	0.3	0.38	0.33
S14	0.32	0.26	0	0.46	0.16	0.38
S15	0.50	0.22	0.0733	0.5	0.26	0.25
S16	0.43	0.24	−0.0011	0.47	0.33	0.2
S17	0.48	0.34	0.0719	0.19	0.4	0.41
S18	0.75	0.26	0.0597	0.36	0.33	0.3
S19	0.01	0.04	0	0.4	0	0.6
S20	0.45	0.28	0.0544	0.24	0.45	0.31
S21	0.58	0.07	0.0208	0.37	0.26	0.36
S22	0.67	0.15	0.0050	0.31	0.34	0.35

Table 4.4. Empirical probability entropy. Phase 2 data observed entropy for all participants, averaged ($\mu(H)$) over all 10 trials, and with standard deviation ($\sigma(H)$) reported. For reference, the maximal entropy expected for a tri-stable system is $\ln(3) = 1.1$. The mean entropy derivative over phase 2 trials is also reported under $\mu(\Delta H)$ and mean perceptual fraction of dominance for l, c and r are reported under $\mu(\hat{p}_l)$, $\mu(\hat{p}_c)$ and $\mu(\hat{p}_r)$.

- We investigated the relationship between first percept (short trials) and continuous viewing (long trials) dynamics, and we provided quantified results of observers' and models' stability in the continuous multi-stable perception.

Results interpretation

In this *Ambiguity* experiment, the manipulation of the moving plaid's transparencies $\boldsymbol{\alpha}$ to generate a tri-stable perceptual phenomenon, as described in the *Percepts* experiment, was replicated with the phase 1 short trial protocol, using key press as a reporting method. In fact, the manipulation of the $\boldsymbol{\alpha}$ parameters was simplified to the manipulation one of the gratings' transparency, and deriving the other grating's one by symmetry using: $\alpha_L + \alpha_R = 1$. Moreover, a psycho-physical observer model was proposed to account for and quantify observers' perception in a Bayesian framework, and to estimate the point of maximal ambiguity in the transparency space. More precisely, this work shows that when using first percept responses (phase 1), which can be obtained in a quick procedure with short trials, one can infer and estimate observers' parameters that are consistent with the continuous viewing of a multi-stable stimulus (phase 2). This was validated by two cross-validation analyses. This is a complementary result to previous reports that the first percept's duration was a predictor of the tri-stable percepts' empirical probabilities in a continuous viewing trial (Hupé and Rubin, 2003). This result opens perspectives for the calibration of ambiguity in a tri-stable stimulus like the moving plaid. It also raises questions on how to choose the next control variables' values when the model is running online and what are its limitations. These points are discussed below.

Stimulus calibration

The diversity in the individuals' profiles in the collected data set solidify the need for quantitative tools to adapt the stimulus in order to reduce inter-individual variance. Finding a set of stimulus parameters generating true ambiguity for all participants is not trivial as inter-individual biases exist, especially in a tri-stable example like the moving plaid used here. In a bi-stable example, one can envision taking central values ($\boldsymbol{\alpha} = 0.5$), however, here, given the histograms presented in Fig. 4.17, one can see that this choice would heavily bias the likelihood of observing coherency percept reports. Therefore, given that the estimation is relatively stable (see Tab. 4.4) and that the 200 short trials of phase 1 took approximately 10 minutes to be completed, it is possible to use phase 1 as a calibration procedure. This calibration can optimise the parameters associated with ambiguity manipulation in

order to generate ambiguous and non-ambiguous trials (as presented in Section 4.1). The *post-hoc* method also shows much better performances than the method used for this work in Parisot, Chauvin, Phlypo, et al. (2019) with higher log-likelihood values, and efficient exploration of the parameter space, given the data. It is noteworthy that results remained consistent across the use of both methods.

Limitations

This calibration is dependent on the model proposed here, which, as shown by Tab. 4.4 is not systematically adaptable to all observers, though it fits a large majority of profiles. However, the procedure could be used to select participants that could be confidently manipulated with regards to visual ambiguity in the moving plaid stimulus. Indeed, participants showing contradictory responses would disqualify in the calibration procedure, in an analogous manner to those that cannot be calibrated with the eye-tracker, as their ambiguity manipulation would not be guaranteed. For such a procedure, it is however important to make the procedure short and efficient to estimate a participant's profile. The model proposed here is specific to the phenomenon and characteristics of the multi-stable moving plaid. Further work could be carried out on varying the formal description of the probability function with less constraints and a higher potential for generalisation. It is also dependent on the optimisation algorithm used to estimate the maximum likelihood and infer the parameters for an observer.

A part of this work which has not received sufficient attention and needs addressing is around the definition of the model with respect to time. This would provide a function that can be used to link in a balanced fashion the data from first percept and continuous viewing trials. Here, however, the weights between the two data sets is highly unequal as the former has 200 data points of unitary weight while the latter is only composed of 30 data points, which are weighted by the empirical probabilities. A key perspective for improvement would consist in finding a relationship that can set a weight on first percept observation, for instance, by taking into account reaction time. However, it is unsure how reaction may relate to perceptual time and further studies need to investigate this relationship. An alternative path would be to consider a method to weight empirical probabilities based on a sliding window over time. But here, a potential obstacle is the risk of turning first percept responses as impulse samples while the rest becomes large masses over time. Indeed, this is due to the difference of use of the key press for perceptual report sampling, and solving such an issue may offer a bridge between continuous and discontinuous viewing paradigms.

Adaptive experimental designs

The online use of the model was done here to verify whether observers would tend towards equi-probability if phase 2 trial α were sampled near the α_{amb} values inferred. However, the arbitrary choice of using a truncated Gaussian distribution centred on one of the randomly chosen α_{amb} could be further investigated. If the model is fed live data, and that an ambiguous and non-ambiguous contrast exist in a protocol, it may be useful to sample the non-ambiguous α couple in a way such that it maximises the information fed into the model by the Bayesian procedure. For instance, looking at the Sigmoid's inflexion points might provide more information to maximum likelihood estimator and reduce sampling redundancy.

Adaptive experimental approaches have been developed over the past decades, notably in tasks where the difficulty level needs to be adapted for each participants. Classical approaches use stair case procedures in which the controlled parameter is gradually increased with regards to a performance measure. However, this has some inconveniences; stair case procedures are, for instance, sensitive to hysteresis (Green, Swets, et al., 1966; Treutwein, 1995; García-Pérez, 1998). Other methods (QUEST or ZEST) have been developed, using the Bayesian framework for instance, to estimate the internal threshold of a participants (Watson and Pelli, 1983; Watson, 2017; Bak and Pillow, 2018). Though they are mostly conceived for 2-alternatives forced choice (2AFC), they show versatility to many psychology problems.

The more recent works by Bak and Pillow (2018) is, to the author's knowledge, a first attempt to extend the problem to more than two categories. The fact that this work was published after the *Ambiguity* experiment was designed and carried out is unfortunate. Their approach uses an information theory criterion computed by Markov-chain Monte Carlo (MCMC) and combined with a Bayesian inference of a psycho-metric observer model. Though the simulated tasks are based on the 2AFC paradigm, they report and show that considering omissions as a third type of response is important to estimate the model properly. This work shows similarities to the methods presented here. But the approach reported in Bak and Pillow (2018) does not relate to specific stimulus parameters, but rather to performance or detection tasks from psycho-physics. Adaptive experiments still need to address limitations around the speed at which it adapts to an observer leaving a stationary behaviour, in which case, the model has over fitted the data and may be stuck in a local minimum of its parameter space, unable to account for unexpected new behaviour. In Bayesian framework, this is manifested by the weight of prior distribution on the learning and model updating process.

Conclusion

In this experiment, a method for adapting the ambiguity control parameter α to individual participants was presented. It relies on implementing a probabilistic pscyho-physical observer model for tri-stability. Inferring its parameters was carried out using the maximum likelihood approach. The methods were tested in an experimental protocol that featured two phases; the first with many short trials for an initial estimation, the second with long trials to test and update online the estimation. Results showed that the estimation was effective within a short time, and that it remained robust to longer trial generalisation. However, many features could be enhanced as discussed, allowing to account for participants with incoherent behaviours, optimising the procedure and the possibility to generalise the methods. Overall, the experiment shows satisfactory results to select participants and calibrate the stimulus to reduce inter-individual variability.

This work provides results on the moving plaids' multi-stability: it is possible to manipulate ambiguity, in a quantitative and individual manner, based on observers' first percept's choice responses.

4.4 Conclusion

The work presented in this chapter provide further empirical insight on the moving plaid stimulus' multi-stable dynamics. In particular, the role of the gratings' transparency parameters was investigated. The *Percepts* experiment described the ground truth of a tri-stable perception being linked to motion perception. The *Ambiguity* experiment showed that this tri-stability can be manipulated and controlled in a sub-space of the transparency parameters. Furthermore, the development of a probabilistic model provides quantified inferred information on the maximal ambiguity level for each observer, making ambiguity control possible with the moving plaid. These results can be used to build up more complex experiments, in order to investigate the relationship between oculomotor and perceptual systems when facing ambiguous stimulation.

5

Multi-stability as a probe of synergy between action and perception?

> *Perception is never passive. We are not only receivers of the world; we also actively produce it. There is a hallucinatory quality to all perception, and illusions are easy to create.*
>
> **— Siri Hustvedt**
> **"The Summer Without Men"**
> **2011.**

This chapter reflects on the works presented in this thesis. Its aim is to synthesise and provide perspectives. Data and analyses from partially completed experiments, will be presented to illustrate and provide insights on this thesis' conclusions. In Chapter 2, we reported the systematic measurement of FEM that we propose to call micro-pursuits. The task in which micro-pursuit behaviour emerged was linked to the bi-stable perception of an ambiguous Necker cube. This prompted the investigation of oculomotor and perceptual multi-stable models, presented in Chapter 3. Hypotheses were derived from this theoretical framework in order to investigate multi-stability in perception and eye movement in experimental work in the moving plaid stimulus. Results on the manipulation of ambiguity for the moving plaid were reported in Chapter 4. However, the investigation of oculomotor manipulation, in the context of moving plaid multi-stability, remains incomplete and is discussed in the following pages. Finally, new paths for investigation are proposed, inviting the reader to think beyond multi-stability and consider it as a complex system's regime of behaviours.

Contents

5.1 Synthesis of contributions

The following paragraphs summarise the results presented in the previous chapters.

5.1.1 Micro-pursuits reveal bi-stability in the oculomotor dynamics

Human vision and eye movements are intrinsically linked as the latter change the visual input projected on the retina. Though our visual representation is stable, the eyes never truly stay still and generate small amplitude FEM that can be interesting markers of cognitive states. Research in the field of FEM has been extensive on micro-saccades, but less is known about drift and slow movements. When searching for micro-saccades, our data showed the presence of a secondary sequence, contrasting with the well-known main sequence exhibited by micro-saccades. Detected micro-saccades that belonged to this secondary sequence showed lower peak velocities as well as higher similarity with the target, which has led us to classify these movements as micro-pursuits.

Micro-pursuits are proposed as a type of FEM occurring at small amplitude, within a fixation, as the gaze follows a target while being constrained to a fixation, for instance. Drift and slow movements tend to be considered as independent from visual stimulation, since larger eye movements are typically used to explore the visual field. In addition, the gaze showed high similarity with the target trajectory, measured through maximally projected correlation. Individual and group analyses gave significant results both in an implicit (Necker) and an explicit (Cross) pursuit task experiment, but not in a secondary implicit (Square) pursuit task experiment. The inter-experiment analysis results suggest that the manipulation of task, stimulus target motion, and the complexity of the stimulus may play a role in the generation of micro-pursuits.

Micro-pursuit here is presented as a class of fixation, but further research is needed to identify the physical properties and distinguish it from other FEM. Moreover, this work calls for further investigation on the functional role of micro-pursuits, and how the oculomotor and perceptual systems interact during such movements.

> **Indeed, the data suggested a link between perceptual and oculomotor multi-stability as micro-pursuits were observed in an ambiguous bi-stable perception condition (Necker) and an explicit pursuit task (Cross), but not in an implicit pursuit task (Square). We interpreted**

this result as a sign of competition between parts of the stimulus to attend to: the fixation cross or the target object (Necker cube or square).

Micro-pursuits.
The work is presented in detail in Chapter 2 and has been accepted for publication to the *Journal of Vision*, as an article under the title *Micro-pursuit: a class of fixational eye movements correlating with smooth, predictable, small-scale target trajectories*, Kevin Parisot, Steeve Zozor, Anne Guérin-Dugué, Ronald Phlypo, & Alan Chauvin, and is excepted to be published shortly after the time of writing this manuscript.

5.1.2 Energy field attractors reproduce oculomotor and perceptual dynamics

Eye movements and multi-stable perception have been further understood and deciphered with the study of predictive models capable of producing analogous behaviours. Models, however, offer a key advantage: they can be manipulated and understood in fine details through theoretical studies and numerical simulations. We proposed a model based on gravitational energy potentials to generate eye movements. Perceptual tri-stability was studied in order to investigate the generalisation results found for bi-stability and how they fit multi-stability. The proposed framework provides tools and results towards the construction of formal models that bind perception and action.

The models gave explanations on how the oculomotor system might produce micro-pursuits and other eye movements. It also provided a framework in which active vision processes can be generated, for the moving plaid stimulus for instance, and thus, predictions could be made to drive the experimental investigation.

Models were based on the same framework, gravitational energy potential fields, and were used to generate and study both eye movement and perceptual dynamics. Differences in the perceptual durations generated by bi-stable and tri-stable models were described, suggesting that noise plays a crucial role in model stability, but the number of attractors may also explain rapidly changing behaviours. An interpretation worth researching lays in considering the context of multi-stability and how this regime might interact with other stability regimes (e.g., mono-stability, meta-stability, instability).

Gravitational potential energy models.
The predictive models are presented in detail in Chapter 3 and served to design the experiments in Chapter 4. However, this work could be extended by inferring parameters given data for both models, and combining both oculomotor and perceptual models into a fused model of active perception.

5.1.3 Ambiguity manipulation of a tri-stable moving plaid

The ideas and design were driven from the models proposed in Chapter 3, and aimed to investigate the multi-stable dynamics of perception with the moving plaid stimulus on one hand, and the dynamics of the oculomotor system on the other. The emergence of multi-stability is the result of combining a stimulus with specific signal properties and contradictory inferences for the visual system. It is possible to stabilise or bias these perceptions by changing the stimulus through expectations or task manipulation or also by modifying the oculomotor dynamics. The perceptual and oculomotor systems being interlinked, acting on one of them has impact on the other.

We provided further empirical insight on the moving plaid stimulus' multi-stable dynamics. In particular, the role of the gratings' transparency parameters was investigated.

> **The *Percepts* experiment described the observer's perceived direction of a tri-stable perception being linked to motion perception. The *Ambiguity* experiment showed that this tri-stability can be manipulated and controlled in a sub-space of the transparency parameters. Furthermore, the development of a probabilistic model provides quantified inferred information on the maximal ambiguity level for each observer by controlling gratings' transparencies simultaneously.**

These results can be used to build up more complex experiments, in order to investigate the relationship between oculomotor and perceptual systems when facing ambiguous stimulation. The aim was to reach a situation where it is possible to show that motor control is a physiological marker of perceptual content, in a no-report paradigm.

However methods to control oculomotor dynamics need further investigation and validation. Some have been explored and are presented in the following section.

Multi-stable moving plaid experiments.
This work is presented in detail in Chapter 4 and some of the results were published in a *GRETSI* conference paper in 2019, in french: *Modélisation de l'ambiguïté d'une multi-stabilité visuelle.*, Kevin Parisot, Alan Chauvin, Ronald Phlypo, & Steeve Zozor, *GRETSI, 2019* (Parisot, Chauvin, Phlypo, et al., 2019). The results presented there, however, expand on this publication.

5.2 Influencing gaze control with random dot kinematograms

***Eye Movements* experiment.**
In this section, we cover our attempts to manipulate the oculomotor behaviour of the observers by manipulating the moving plaid stimulus. Some results are presented in Appendix B.2, though they remain preliminary. They may still provide some insights for the reader. Here, only a summary is provided.

The *Eye Movements* experiment focused on the central disk of the stimulus that is used to generate the gaze fixation point. Given our understanding of the GraFEM model (see Chapter 3) and if the attractors in the perceptual and oculomotor spaces are merged as hypothesised in Section 4.1, one can expect the attractors to move in direction of the motion perceived, as identified in the *Percepts* experiment (Section 4.2). We know from our simulations on GraFEM that the motion of an attractor can lead to the generation of spatio-temporal patterns for the gaze signals that can be interpreted and classified as pursuits or saccades. It is however possible that a perceptual attractor is not strong enough, i.e., its parameters make it relatively shallow and wide enough to detach the gaze particle from the attractor linked to the fixation point. Competition between attractors is key in driving the dynamics of the model, therefore the predictions will depend on their relative strength. The aim in this experiment was to test whether a visual stimulus can be added to the moving plaid, so that the strength of attraction of the motion perceived are manipulated in an excitatory or inhibitory way.

One method to induce oculomotor biases was to use a RDK at the centre of the stimulus. Our aim was to establish a coherence ratio that would generate implicit micro-pursuit movements in the same direction as the signal and corrective micro-saccades in the opposite direction.

> **Hypothesis: there exist some interaction between the direction of induced eye movements by a RDK, and the moving plaid percepts' durations, i.e., one can influence percept durations by manipulating the RDK. If true, one could then deduce that a bottom-up process explains the perceptual dynamic, partially. If no interactions are measured, one can only conclude that eye movements induced by the RDK cannot influence percepts' durations.**

The relationship was not expected to be necessarily symmetrical; indeed inhibition should be less powerful as excitation.

Random-dot kinematogram.
A RDK is a set of points that have random movements in a defined area. Different implementations exist (Scase et al., 1996) with the three main noise methods combined with two dot signal selection methods (see Fig. 5.1).

- **Random position**—when a dot is selected as noise, its next position will be chosen using a random distribution—mostly a uniform distribution over an area, but not necessarily—and it will be relocated at that position at the next iteration.
- **Random walk**—when a dot is selected as noise, it will follow a random walk, also known as Brownian motion (Einstein, 1956), in which the dot's direction and amplitude are randomly picked over each iteration.
- **Random direction**—when a dot is selected as noise, it will have a fixed motion direction which is drawn once for its entire life time, at birth.
- **Same**—when a dot is selected as noise, it remains so for its entire life time.
- **Different**—dots are selected as noise at each iteration.

These methods generate different spatio-temporal dynamics, visually and perceptively, and the coherence ratio—e.g., the percentage of dots that are associated to the signal—affects visual decoding with varying degrees of efficacy (Scase et al., 1996; Pilly and Seitz, 2009). In other words, the threshold for the correct detection of the signal's direction varies across RDK methods. Moreover, it has been reported that the smooth pursuit latency and early eye acceleration are not affected by the type of RDK, but late eye acceleration, pursuit gain and perceived velocity were dependent on RDK type (Schütz, D. I. Braun, Movshon, et al., 2010). Perception and pursuit performance also showed correlated dynamics. The authors interpreted their results as the pursuit system showing a capacity to integrate across directions of the RDK's signal, but not velocity.

The *Eye Movements* experiment used a RDK to verify whether oculomotor dynamics could be manipulated over a continuous observation trial of a multi-stable moving plaid stimulus. The experiment explored the combination of RDK with and without the moving plaid, a variety of orientations for the RDK signal direction, different proportions of RDK dots being attributed to the manipulated signal, and control trials without RDK and full noise RDK. The experiment, for now, does not find empirical evidence that a RDK based on *same random direction*, 10% or 20% of dots moving in a coherent direction and presented in the gaze fixation circle area at the centre of screen, and in the fovea, will influence oculomotor dynamics. In fact, when presented with the moving plaid, these dynamics seemed to be dominated by the multi-stable perceptual changes. In other words, the use of RDK was not effective enough for oculomotor modulation, such that excitatory and inhibitory behaviours' impact on perceptual dynamics may be tested in our *Gaze-EEG* experiment (more details on the design in Appendix A.8). However, the algorithm developed to analyse

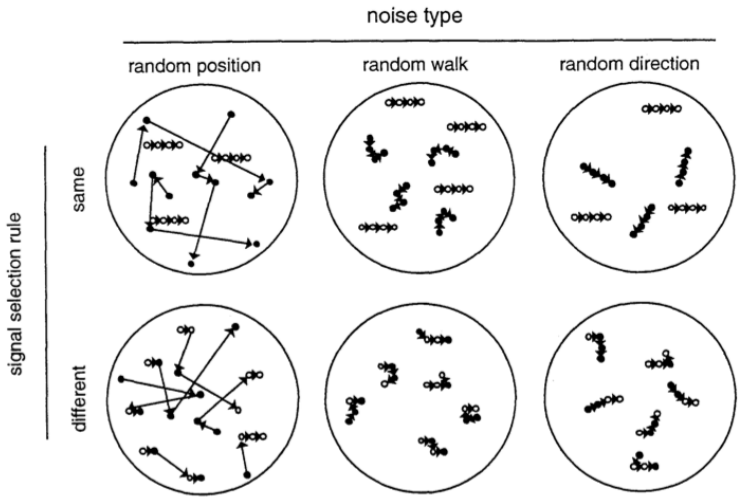


Figure 5.1. Random-dot kinematogram. Schematic illustration of the six types of signal/noise display generated by the rules described in the text. The figure shows examples with 50% coherence and rightward signal motion in each case. Dots designated as signal dots for the following displacement are shown as open circles, those designated as noise dots are solid circles. In the random-position case, the displacement vectors shown join each noise dot to its new position selected by the plotting algorithm; for the visual system, these are not necessarily the most effective pairings for generating motion signals. Figure taken from Scase et al. (1996).

gaze data, in the context of RDK or moving plaid influences, can be used for further investigation of the impact of eye movements in multi-stable perception.

To provide answers to the hypothesis given above, we would need to carry another version of this experiment with less RDK orientation conditions (e.g., a minimum of three directions: -60°, 0°, and 60°) and with at least two conditions for RDK amplitude values (e.g., presence vs absence of RDK). Based on this initial exploration, the size, or spatial location, of the RDK should also be investigated: pursuit movement, OKN, or OFR eye movements are best generated when affecting peripheral vision, rather than central vision, as is the case in the presented setup. Disentangling eye movements produced by the moving plaids from an additional oculomotor manipulation is not a trivial task, from an experimental perspective. It is however crucial to strike a controlled balance between the percepts and a oculomotor manipulation, so as to investigate the influence of the latter on the former.

5.3 Eye movements as objective markers in ambiguous perception

***Noisy Motor Events* experiment.**
In this section, we cover a first attempt at implementing a no-report paradigm for the experimental investigation of the moving plaid. This work was carried out in the context of Eva April's internship, in which she carried out most of the experimental work, which had to be finished in a reduced time frame. This is why, some of the methods developed in Chapter 4 (e.g., the probabilistic observer model) were not applied in the protocol. The presented analyses were carried out after the internship. The results presented here remain preliminary but may provide insights for the reader.

An experiment to test the blind condition, proposed in Section 4.1, was carried out to verify that perceptual dynamics could be inferred based on indirect, objective, and physiological markers. Few attempts have managed to achieve this feature, and dependency on subjective report remains a large scientific gap in the study of multi-stable perception. The aim of this experiment was to show that perceptual dynamics could be inferred from gaze signals. We contrasted trials in which participants were asked to report their perception, and those in which they simply had to observe the moving plaid stimulus. Furthermore, we evaluated whether removing the gaze fixation task constraint would enable to have more powerful effects in the oculomotor signals, or not. An algorithm to detect micro-pursuits and pursuits tracking moving percepts is presented and was used to estimate perceptual epochs in our data.

5.3.1 Motivation

Bi-stability, eye movements and no-reporting literature review

Neural correlates of consciousness.
The no-reporting research field is related to research for NCC, in which the former is often a methodological basis for the investigation of the latter. A short review of this growing area of cognitive research is provided in Appendix A.7.

Attempts on inferring perceptual dynamics based on pupil dynamics have been reported (Einhäuser, Stout, et al., 2008) though they have also been contested (Hupé, Lamirel, et al., 2009). A more successful approach has been based on exploiting the OKN associated with motion perceived on a large part of the visual field in a binocular rivalry setting (Naber et al., 2011; Frässle et al., 2014; Aleshin et al., 2019). The motion of gaze is directly associated to the illusory percept (Madelain

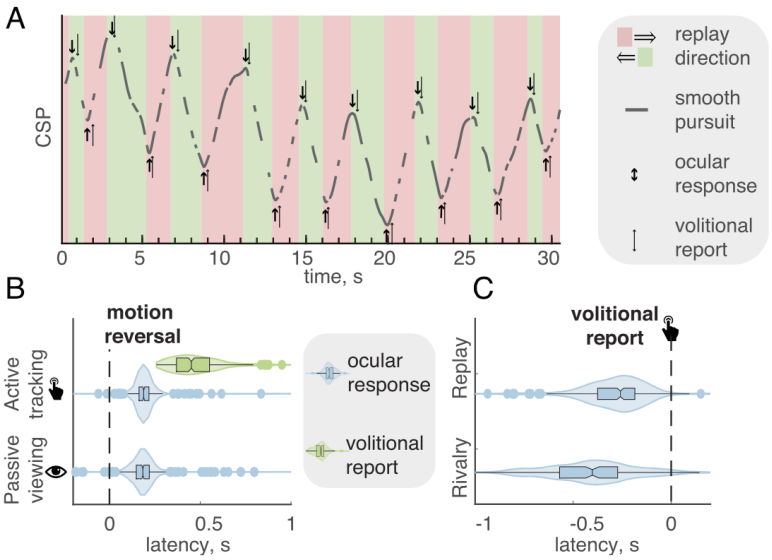


Figure 5.2. No-report perceptual inference example.
A: Cumulative smooth pursuit (CSP) over time in an example trial showing the temporal dynamics of OKN eye movements for a binocular rivalry. The arrows show that ocular response is faster than key press report.
B: Violin plots showing the distribution of latencies for ocular responses and key press report, thus confirming the tendency for faster perceptual change reports using eye movements.
C: Violin plots of the latencies relative to key press, showing yet again, that ocular response is systematically faster than key press volitional report.
Figures taken from Aleshin et al. (2019).

and Krauzlis, 2003) and is used to interpret the oculomotor data to estimate the perceptual temporal series (see Fig. 5.2, taken from Aleshin et al. (2019)).

Key press and motor programming

As presented briefly in Chapter 1, multi-stable perception relies almost entirely on key press subjective report to obtain perceptual dynamics. This is specially true for the moving plaid stimulus (Hupé and Rubin, 2003; Hupé and Rubin, 2004; Rubin, Hupé, et al., 2005; Moreno-Bote, Shpiro, et al., 2008; Moreno-Bote, Shpiro, et al., 2010; Huguet et al., 2014; Hupé, Signorelli, et al., 2019). Key press report require the observer to engage in motor programming and in cognitive processes that transfer the perceptual representation experienced into the mapping on the keyboard and the necessary associated action. This process can be cognitively expensive, depending on the participant's habituation (Ballanger and Boulinguez, 2009), and is subject to variability in learning performances across individuals (Veltman and Gaillard, 1998).

Given the results provided in Kornmeier and Bach (2012), we know that the motor response is highly variable in the case of disambiguated bi-stable stimuli such as the Necker Cube. This variability affects the precision of the measurement method of key press for perceptual time series. In turn, if other phenomena are studied jointly with multi-stability, correlations are hard to estimate. To the author's knowledge, though there have been attempts to measure multi-stable perception by other means than key press (Naber et al., 2011; Frässle et al., 2014; Aleshin et al., 2019)—and these have focused on using eye movements, but see Rees (2007) and Sterzer et al. (2009) for attempts at decoding bi-stability with neural signals—these studies have not tried to estimate the reliability of key press report. However, some studies have looked at the issue from a neuroscience methodology point of view, in order to reduce issues with signal synchronisation (Kornmeier, Ehm, et al., 2007), but questions remain as to whether their solutions give rise to a phenomenon equivalent to multi-stable perception (VanRullen and Koch, 2003; VanRullen, Busch, et al., 2011). Another interesting approach has also been oriented on taking advantage of multi-sensory binding to manipulate perception across modalities (Schwartz et al., 2012).

The aim of replacing key press and motor programming is to reduce the attentional shifts that are linked to such actions, as explained in Section 4.1. Attention is known to affect the dynamics of multi-stable perception (Kohler et al., 2008; Li et al., 2017), though its effects differ between binocular rivalry and ambiguous figures (Dieter et al., 2016). Attention is also known to have an impact on motor action and learning (Song, 2019).

Other ways to report perception: eye movement

Recent results and methods using a combination of moving gratings in a binocular rivalry setup has shown that a no-report experimental setup is possible to study multi-stability using eye movements, and more specifically OKN, a reflexive type of pursuit (Naber et al., 2011; Frässle et al., 2014; Aleshin et al., 2019), but other types of eye movements seem to provide information on perception (Ee, Van Dam, et al., 2005; Laubrock et al., 2008). In fact, in Aleshin et al. (2019), the oculomotor analysis to detect pursuits for perceptual dynamics inference is improved using an algorithm (see Fig. 5.3) that (1) removes unwanted oculomotor events—e.g., blinks, saccades and fixations—and retain pursuit epochs, (2) shifts the signal's position to compensate the introduced offsets due to these events and obtain a CSP and (3) obtain velocities to estimate changes of direction of the CSP. This method allows to obtain epochs that can then be classified as linked to one of the competing percepts, based on velocity's direction and velocity.

The similarities between the moving plaid stimulus, used in our experiments, and the binocular rivalry stimulus used in Aleshin et al. (2019) are advantageous. Indeed, the moving plaid stimulus has perceived movements with same directional properties in the case of the transparent percept, as established in Section 4.2. Only a class for the coherent upward moving percept needs to be added. Therefore, this approach shows great potential to be applied in our experimental setup, in order to have a technique to infer perceptual dynamics based on the oculomotor signals. Questions remained however on whether the oculomotor fixation task given to reduce retinal image variations across the trials is not a factor that can reduce the pursuit effect expected to be exploited. Interestingly, the *micro-pursuits* defined in Chapter 2 may be a key type of oculomotor events to interpret perceptual dynamics for the moving plaid, when the oculomotor fixation task applies.

For this experiment, our hypotheses were the following:

- **Oculomotor markers of perception exist and can be detected in order to investigate perceptual dynamics in a no-report paradigm;**
- **The oculomotor task plays a role in amplifying the oculomotor markers;**
- **Key press and exogenous (non-ambiguous) changes show latencies (> 200 ms) larger than those observed for oculomotor markers (Aleshin et al., 2019).**

Noisy Motor Events experiment
The *Noisy Motor Events* experiment's methods and preliminary results are presented in Appendix B.3.

5.3.2 Expected results

Based on the preliminary results and exploration of the data collected in the *Noisy Motor Events* experiment, we will present some of the expected, and necessary, results required to achieve a reliable no-report paradigm for the study of multi-stable perception with the moving plaid.

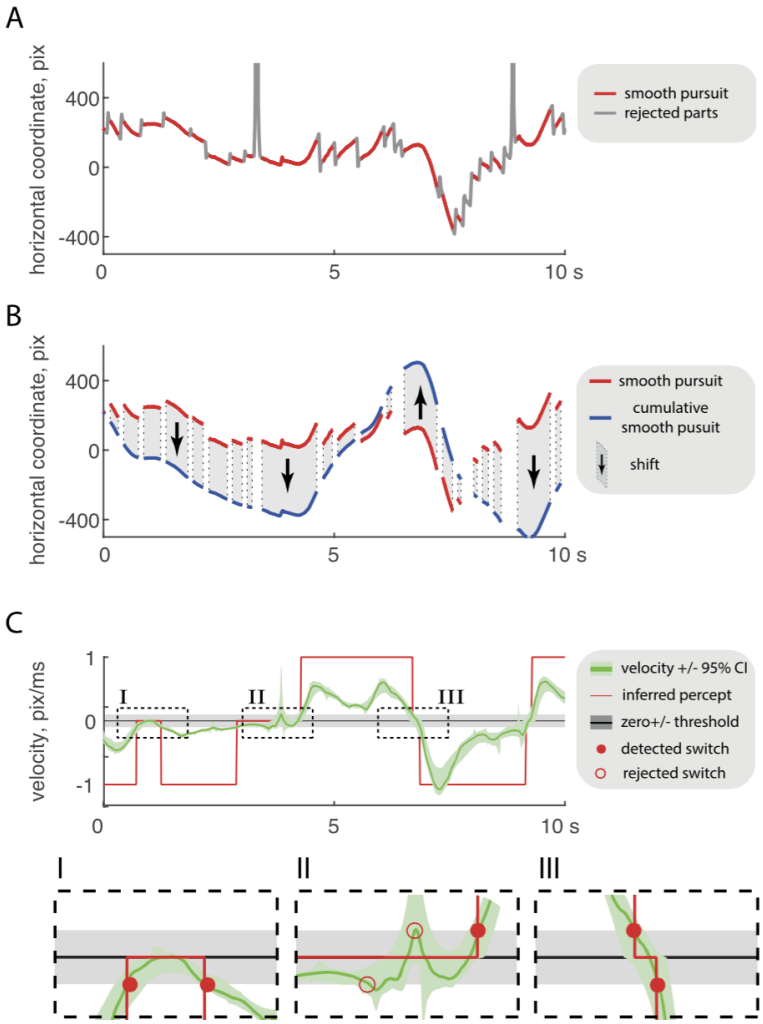


Figure 5.3. Cumulative smooth pursuit algorithm. Diagrams explaining the different steps of gaze analysis presented in Aleshin et al. (2019) by using an example temporal series.
A: Plot showing the original data (grey) and the detection of smooth pursuit epochs (red) and rejected parts of the signal.
B: Plot showing the effect of the cumulative smooth pursuit shift (blue) such that spatial offsets introduced by saccades and blinks are compensated.
C: Plot showing the first order derivative, velocity, of the gaze signal in green with shades for $\pm$95% confidence intervals, with the inferred percepts in red, the threshold area used in the algorithm for perceptual inference in grey, detected and rejected perceptual switches marked by dots on the time series. Below, snap shots of **C** showcasing **I** a remain in percept scenario, **II** an extended and prolonged transition from one percept to the other with rejected threshold crossings within, and **III** a clear transition from one percept to another.

Perceptual inference based on eye movements

To achieve a reliable no-report paradigm, we need to have signal processing and percept inference algorithms, based on eye movement data, that provide results highly similar to the reported perceptual changes, in the report condition. For that purpose, the metric, used to compute the similarity between reported subjective perceptual timelines and the timelines inferred from physiological markers, needs to be chosen adequately. For instance, we can expect eye movement percept reversal to precede key press, as the former is associated to lower latencies, around 100 ms, while the latter tends to be closer to 400 ms. In fact, this was observed in some of our data, in qualitative manner, and is coherent with the results reported by Aleshin et al. (2019).

Given good perceptual inference results could be achieved, we would be able to contrast perceptual dynamics, based on objective markers, and compare latencies, percept duration distributions and investigate how the key press motor effort may affect the perceptual (and oculomotor) dynamics. For instance, our theoretical prediction predicts that pressing a keyboard key requires some level of attentional shift (see Section 1.3 in Chapter 1), which should in turn, introduce noise in the perceptual system, thus increasing the probability of perceptual change. This could be analysed by looking at how often rapid consecutive changes, i.e., short consecutive percept durations, occur in report versus no-report conditions. Such a result would provide further evidence for the Premotor theory of attention, and the necessity to consider its consequences in continuous viewing paradigms with subject motor report.

A last non negligible point is the following: not all three moving plaid's percepts are equal. Our preliminary analysis and the results presented in Section 4.2 in Chapter 4 show that transparency percepts are harder to discriminate, based on perceived motion direction. Indeed, unlike for coherency where the entire stimulus moves in one direction, the transparency percepts have two direction of motion, and the competition relies on depth ordering. This may lead to observer reporting one transparency percept, e.g., left transparency, but tracking with the eyes the grating perceived at the back moving rightwards. Furthermore, as explained in Section 4.1 in Chapter 4, the ambiguity rising from the aperture problem leads to slower perceived motion in the direction that is orthogonal to the bars' square wave front. This may pose problems if pursuit is selected as a physiological marker, since pursuit operates best in a restraint target velocity interval. For instance, participants may track the transparency percepts' motion, while moving gaze along the bar, thus creating elliptical motion in space and over time.

Choosing an oculomotor task

Intention, or top-down processes, also play a role in trials lasting over 20 seconds. In the *Noisy Motor Events* experiment (presented in Appendix B.3), we tried to manipulate the oculomotor task given to participants by having blocks where participants had to fixate a static dot at the centre of the stimulus, and blocks in which, participants could explore the stimulus freely. We expected to observe more percept pursuits and micro-pursuits in the latter condition, which would provide clearer perceptual signature to decipher perception blindly. However, manipulating the oculomotor task makes controlling the visual input on the retinal projection much harder. Even though fixation tasks are hardly immune to FEM, they greatly reduce the variations on the retinal projection, hence giving more certainty on the interpretation of the perceptual behaviours observed.

A third possible oculomotor task, which was not investigated here, would consist in asking the participants to explicitly track, with their eyes, their percepts. With such a task, decoding perception would be easier, as pursuit suppression would be expected to be less present; for instance, the OKN is a powerful oculomotor phenomenon hardly suppressed. Another argument for this approach can be derived from the results observed in the explicit micro-pursuit task's results, reported in Chapter 2. These results suggested that intentional pursuit can be detected more efficiently, when the task is to explicitly follow a target, than for a distractor target. In future works, this third oculomotor task should be explored, and we expect to observe better pursuits, and thus, better perceptual inference.

Ambiguity manipulation

As explored in Section 4.3 in Chapter 4, manipulating the ambiguity of the moving plaid can be achieved by controlling the gratings' transparencies, and improved when a probabilistic model allows to compute the points of maximal ambiguity. In the *Noisy Motor Events* experiment, unfortunately, we could not implement this method, and brutal exogenous changes to the stimulus were carried out. These trials provide a third time series to compare with the key press and the eye movement time series. This part of the data can be useful to characterise the differences between report and no-report conditions; for instance, for latency, i.e., reaction time, with exogenous changes, one can access a precise onset. Finally, the non-ambiguous data is key to infer neural correlates and signatures, using neuro-imaging techniques, such as presented and designed for the Gaze-EEG experiment, presented in Appendix A.8.

Synthesis & perspectives

This last experiment shows that moving towards no-report paradigms in on a tri-stable moving plaid stimulus is a complex procedure. Unlike the bi-stable stimulus used in (Aleshin et al., 2019) which generated OKN reflex eye movements, here the role of attention and top-down processes may be more important, making eye movement behaviour less systematic. A change in the oculomotor task phrasing might be an interesting path to improve such an approach and have more confidence in the data processing, such that data in the no-report condition can be analysed and interpreted. Finally, the data processing methods may be improved, as in Appendix B.2, by the use of model library (such as the GraFEM model from Chapter 3) in order to make links between formally understood model parameters and observed data. This may be achieved by comparing measured gaze trajectories to a library of simulated gaze trajectories, using, for instance, the similarity metric of MPC presented in Chapter 2, or also by computing the probability that the data has been generated by a model given fixed parameters such as the one presented in Chapter 3. Another interesting perspective would consist in finding characteristics in the Fourier domain of the gaze signals. Quaternion Fourier analysis has been recently applied to different bi-variate signal problems (Flamant, 2018) and exploring the gaze's elliptic properties through such analyses could offer new insights.

5.4 What does stability mean for perception?

In this thesis, we have looked at multi-stability, a regime of stability where multiple stable exist for a system. We have looked at it in eye movements, in perception and in models. Here we reflect on the larger context, and how this regime might relate to other regimes of stability (e.g., instability).

5.4.1 Oculomotor multi-stability

The work presented in Chapter 2 on micro-pursuit, combined with the GraFEM model simulations, presented in Chapter 3, provides a view of eye movements as a multi-stable process, in the experimental conditions where a fixation target and a perceived target co-exist. This approach provides a way to read oculomotor dynamics, such as the ones presented by Yarbus (1967) (see Chapter 1), as driven by a system that has multiple stable attractors, in the visual field. These attractors may be due to bottom-up processes (e.g., salience, attention) or top-down signals (e.g., task, intention). The measured and observed gaze dynamics thus live in multi-stable regime, where multiple varying and dynamical attractors co-exist and compete for eye movement programming.

The framework proposed allows to reproduce all FEM (and possibly macro eye movements) as shown in Chapter 3, and provides a theoretical tool that can be interpreted intuitively—the gravitational basis of the model is understood by most who experience gravity on earth—and formally, as it has been a extensively researched and documented family of models in Physics. Furthermore, this type of approach could be adapted to more ecological contexts as the number of attractors may increase, and attractors may represent aggregated features, in a similar way as perceptual objects are aggregation of visual features, interpreted by the brain.

Going beyond eye movements, one may look at integrating them with neural correlates, by making joint neuro-imaging measurements. Indeed, this is something that motivated the experiments presented in Chapter 4, Appendix B.2 and Section 5.3. Though a design is proposed in Appendix A.8, based on the results presented in the thesis and the theoretical analysis conducted in Section 4.1, many challenges remain to be solved, before having sufficiently robust methods to investigate the NCC of multi-stable perception. To summarise them, one should (i) further develop an oculomotor manipulation method on the stimulus, (ii) enhance perceptual inference and detection algorithms, based on eye movements, (iii) anticipate issues with temporal synchronisation using EEG (or MEG), and (iv) find practical solution to

solve attention variation in continuous viewing paradigms, especially with multiple measurement instruments interacting on participants.

5.4.2 Model extension

In Chapter 3, an oculomotor model and a perceptual model were presented, and briefly studied for the goals of the experiments presented in Section 4.1 (Chapter 4). But a lot of work remains to be done on such models. The GraFEM, oculomotor, model has been used to reproduce only FEM, but simulating other oculomotor phenomena (e.g., saccade deviation, anti-saccade, scene exploration, reading, visual search, etc) should be feasible. Interpreting the model's parameter for each phenomenon may provide a holistic vision of oculomotricity and how these observations relate to one another. Each phenomenon would require to inverse the model, a considerable theoretical work, which far exceeded the scope of this thesis, but would probably provide a lot of predictions that could be tested experimentally.

As such, the perceptual model studied in Chapter 3, based on GraFEM, was also studied in a very limited way here. More theoretical work on the role of the deterministic force, namely adaptation, is needed to understand its relationship perceptual dynamics, and provide predictions for experimental work. Indeed, as mentioned in Chapter 1, this point remains unclear to the research community, in particular on how it may be implemented in the brain. Though this model may not provide direct answers, it could be a tool to model multi-stable perception beyond bi-stability. By investigating the phenomenon with more states, the properties of adaptation, or mutual inhibition between percepts, may be exposed further since it is the force that *chooses* the percept.

One of the proposed framework's strength is its flexibility to provide models for perception *and* eye movements. Although these have been shown separately in Chapter 3, model fusion was not investigated, unfortunately. Indeed, an interesting perspective lays in the possibility of looking at both models simultaneously, and proposing different architectures, i.e., hierarchies, to study an enacted multi-stable perception, and to test and identify the best fitting models that explain the data. This approach could provide insights on how the oculomotor and perceptual system are coupled together when facing motion ambiguity. This research question further extends to understanding whether multi-stability is a regime of stability linked to system coupling, and what this regime means for the studied system; the brain. Similar approaches exist for data analysis of neural activity. For instance, T. Watanabe et al. (2014) investigated macro dynamics of fMRI activity using methods based on

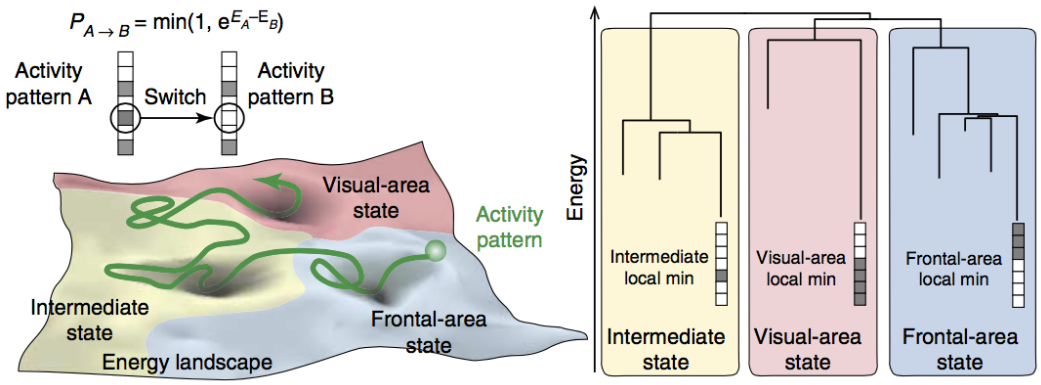


Figure 5.4. fMRI activity in multi-stable energy landscapes. Schematic illustration of the methods used to analyse fMRI data during a bi-stable perception. As different regions of the brain display changing fMRI activity pattern, the activation is considered as a system in its parametric space, where attractor states exist. The energy landscape provides probabilistic information on the system's behaviour patterns. Figure taken from T. Watanabe et al. (2014).

energy potential modelling, where attractors provide stable states for the system. Hence they were able to interpret activation over various anatomical structure in the brain to individual's bi-stable behaviour (see Fig. 5.4 for a schematic illustration).

5.4.3 Meta-, mono-, & multi-stability

A higher level of abstraction and investigation consists of placing multi-stability in its context of behaviour regimes. As mentioned in Chapter 1 and Chapter 3, theoretical research are investing questions underlying how, in the brain, system interact and couple their processes. In this context, multi-stability can be interpreted as a behavioural marker of system's synergy and degeneracy[1] state (Kelso, 2012). It provides information on how complementary sub-systems of the visual one are, and invites researchers in the field to learn new tools of analysis. One of these new tools is the *bifurcation* analysis proposed by Henri Poincaré in 1885 (Poincaré, 1885), where the changes on the topological structure of the integrated curves of a family of vector fields is studied by finding the solutions of a family of differential equations. For instance, bifurcation theory has been applied to bi-stable models to isolate the role of the system's internal noise on the observed dynamics (Pisarchik et al., 2014; Magallón-Garcıa et al., 2017). This method was also applied to investigate the impact of attention on binocular rivalry by combining, and comparing, model and experimental data (Li et al., 2017). This study provided a complete demonstration

[1] Degeneracy is value of component independence to generate a function in a complex system.

of how attention is necessary for binocular rivalry to occur, and what regimes drive the changes; namely, equal activity, oscillation or winner-take-all. These regimes depended on percepts' mutual inhibition and attentional modulation.

Further research is still needed to compare meta-stable, mono-stable and multi-stable phenomena. Investigating such different regimes experimentally, while keeping some level of control to compare data set may be challenging. However, multi-modal investigation of multi-stable perception research have shown promising potential (Schwartz et al., 2012). Studies coupling bi-stable auditory and visual stimuli provide evidence that sensory modality binding occurs such that the brain infers the most coherent interpretation of the stimulation. If methods to quantify and control maximal ambiguity in a modality, such as the ones presented in the *Ambiguity* experiment in Chapter 4, while a mono-stable or meta-stable stimulus is presented in another modality, one could contrast stability regimes and ambiguity interactions. Hence, such an approach may provide a first step towards empirical and experimental characterisation of stability regimes with regards to multi-stable perception.

5.5 Conclusion

The works presented in this thesis show results from a trans-disciplinary approach to visual multi-stable perception and eye movement research. By combining methods from the signal processing, psychology and physics communities, we showed that the oculomotor system can provide micro-pursuit eye movements in an explicit and implicit context, with the latter being linked to bi-stable perception (Chapter 2). We proposed an energy field particle model for eye movements and multi-stable perception, thus providing a theory for how the two systems might be coupled, and by using a formalism that allows the study of more than two state multi-stability (Chapter 3). In a series of experiments, we laid ground work for the design of an experimental design that can investigate the link between oculomotor and perceptual system in the context of multi-stability, e.g., the moving plaid stimulus (Chapters 4 and 5). This approach featured the use of Bayesian methods on an observers' probabilistic model of ambiguity, investigation of stimulus manipulation for the inhibition or excitation of eye movements correlated with perceived motion, and a first approach at a no-report protocol for the moving plaid. Overall, this thesis aimed to provide a fresh look at a fascinating phenomenon that has been studied over centuries. One of the most promising perspectives is to start looking beyond multi-stability to study it. Understanding the context in which such a phenomenon emerges, to identify contrasts with other complex dynamic systems' regimes should provide fundamental, and radically, new observations and understanding.

Bibliography

Abramov, Israel and Christopher M Harris (1984). “Artificial eye for assessing corneal-reflection eye trackers”. In: *Behavior Research Methods, Instruments, & Computers* 16.5, pp. 437–438 (cit. on p. 283).

Adami, Christoph, Charles Ofria, and Travis C Collier (2000). “Evolution of biological complexity”. In: *Proceedings of the National Academy of Sciences* 97.9, pp. 4463–4468 (cit. on p. 263).

Agostinelli, C. and U. Lund (2017). *R package `circular`: Circular Statistics (version 0.4-93)*. CA: Department of Environmental Sciences, Informatics and Statistics, Ca’ Foscari University, Venice, Italy. UL: Department of Statistics, California Polytechnic State University, San Luis Obispo, California, USA (cit. on p. 163).

Alais, David, Jeroen J van Boxtel, Amanda Parker, and Raymond van Ee (2010). “Attending to auditory signals slows visual alternations in binocular rivalry”. In: *Vision research* 50.10, pp. 929–935 (cit. on p. 288).

Aleshin, Stepan, Gergő Ziman, Ilona Kovács, and Jochen Braun (2019). “Perceptual reversals in binocular rivalry: Improved detection from OKN”. In: *Journal of vision* 19.3, pp. 5–5 (cit. on pp. 44, 65, 142, 206–211, 213, 295, 296, 298, 303, 312, 319, 324).

Aloimonos, John, Isaac Weiss, and Amit Bandyopadhyay (1988). “Active vision”. In: *International journal of computer vision* 1.4, pp. 333–356 (cit. on p. 25).

Anderson, Michael L (2003). “Embodied cognition: A field guide”. In: *Artificial intelligence* 149.1, pp. 91–130 (cit. on p. 260).

Anderson, Nicola, Fraser Anderson, Alan Kingstone, and Walter Bischof (2015). “A comparison of scanpath comparison methods”. In: *Behavior research methods* 47.4, pp. 1377–1392 (cit. on p. 22).

Anderson, T.W. (2003). *An Introduction to Multivariate Statistical Analysis*. Hoboken, New-Jersey, USA: John Wiley & Sons (cit. on p. 82).

Andler, Daniel (1990). “Connexionnisme et cognition: à la recherche des bonnes questions”. In: *Revue de synthèse* 111.1-2, pp. 95–127 (cit. on p. 259).

Anstis, Stuart (2013). “Contour adaptation”. In: *Journal of Vision* 13.2, pp. 25–25 (cit. on p. 25).

– (1980). “The perception of apparent movement”. In: *Philosophical Transactions of the Royal Society of London. B, Biological Sciences* 290.1038, pp. 153–168 (cit. on p. 26).

Anstis, Stuart, Frans Verstraten, and George Mather (1998). “The motion aftereffect”. In: *Trends in cognitive sciences* 2.3, pp. 111–117 (cit. on pp. 25, 48, 51, 63).

Astrand, Elaine, Guilhem Ibos, Jean-René Duhamel, and Suliann Ben Hamed (2015). "Differential dynamics of spatial attention, position, and color coding within the parietofrontal network". In: *Journal of Neuroscience* 35.7, pp. 3174–3189 (cit. on pp. 13, 55, 106, 124, 141).

Baå-Eroglu, C, D Strüber, Michael Stadler, P Kruse, and E Baå (1993). "Multistable visual perception induces a slow positive EEG wave". In: *International Journal of Neuroscience* 73.1-2, pp. 139–151 (cit. on p. 299).

Baars, Bernard J (1993). *A cognitive theory of consciousness*. Cambridge University Press (cit. on p. 265).

Baars, Bernard J and Stan Franklin (2007). "An architectural model of conscious and unconscious brain functions: Global Workspace Theory and IDA". In: *Neural Networks* 20.9, pp. 955–961 (cit. on p. 299).

Bach, Michael and Ch M Poloschek (2006). "Optical illusions". In: *Adv Clin Neurosci Rehabil* 6.2, pp. 20–21 (cit. on p. 25).

Bahill, A Terry, Michael R Clark, and Lawrence Stark (1975). "The main sequence, a tool for studying human eye movements". In: *Mathematical Biosciences* 24.3-4, pp. 191–204 (cit. on pp. 16–18, 89, 104).

Bak, Ji Hyun and Jonathan W Pillow (2018). "Adaptive stimulus selection for multi-alternative psychometric functions with lapses". In: *Journal of vision* 18.12, pp. 4–4 (cit. on pp. 193, 301).

Ballanger, Benedicte and Philippe Boulinguez (2009). "EMG as a key tool to assess motor lateralization and hand reaction time asymmetries". In: *Journal of neuroscience methods* 179.1, pp. 85–89 (cit. on pp. 38, 39, 207).

Barbot, Antoine and Marisa Carrasco (2017). "Attention modifies spatial resolution according to task demands". In: *Psychological science* 28.3, pp. 285–296 (cit. on p. 41).

Barbot, Antoine, Sirui Liu, Ruth Kimchi, and Marisa Carrasco (2018). "Attention enhances apparent perceptual organization". In: *Psychonomic bulletin & review* 25.5, pp. 1824–1832 (cit. on p. 41).

Başar-Eroglu, Canan, Daniel Strüber, Martin Schürmann, Michael Stadler, and Erol Başar (1996). "Gamma-band responses in the brain: a short review of psychophysiological correlates and functional significance". In: *International journal of psychophysiology* 24.1-2, pp. 101–112 (cit. on pp. 301, 303).

Bee, Clair (2020). *EM.vonmises: Expectation-maximization algorithm for mixture of von Mises*. R package version 0.0.0.9000 (cit. on p. 163).

Behrens, Frank, Manfred MacKeben, and Wolfgang Schröder-Preikschat (2010). "An improved algorithm for automatic detection of saccades in eye movement data and for calculating saccade parameters". In: *Behavior research methods* 42.3, pp. 701–708 (cit. on pp. 16, 71).

Bentivoglio, Anna Rita, Susan B Bressman, Emanuele Cassetta, et al. (1997). "Analysis of blink rate patterns in normal subjects". In: *Movement disorders* 12.6, pp. 1028–1034 (cit. on p. 282).

Berens, Philipp et al. (2009). “CircStat: a MATLAB toolbox for circular statistics”. In: *J Stat Softw* 31.10, pp. 1–21 (cit. on pp. 313, 319).

Beutter, Brent and Leland Stone (1998). “Human motion perception and smooth eye movements slow similar directional biases for elongated apertures”. In: *Vision research* 38.9, pp. 1273–1286 (cit. on p. 295).

BIPM, Le (2006). *Système international d’unités (The International System of Units)* (cit. on p. 266).

Blake, Randolph, Robert P O’Shea, and TJ Mueller (1992). “Spatial zones of binocular rivalry in central and peripheral vision”. In: *Visual Neuroscience* 8.5, pp. 469–478 (cit. on p. 33).

Blake, Randolph, Kenith V Sobel, and Lee A Gilroy (2003). “Visual motion retards alternations between conflicting perceptual interpretations”. In: *Neuron* 39.5, pp. 869–878 (cit. on pp. 44, 288).

Blake, Randolph and Frank Tong (2008). “Binocular rivalry”. In: *Scholarpedia* 3.12, p. 1578 (cit. on p. 33).

Blakemore, Colin, JAMES PJ MUNCEY, and Rosalind M Ridley (1971). “Perceptual fading of a stabilized cortical image”. In: *Nature* 233.5316, pp. 204–205 (cit. on p. 264).

Blythe, Hazel I, Simon P Liversedge, Holly SSL Joseph, Sarah J White, and Keith Rayner (2009). “Visual information capture during fixations in reading for children and adults”. In: *Vision research* 49.12, pp. 1583–1591 (cit. on p. 263).

Boccignone, Giuseppe (2019). “Advanced statistical methods for eye movement analysis and modelling: a gentle introduction”. In: *Eye Movement Research*. Springer, pp. 309–405 (cit. on pp. 14, 23).

Boi, Marco, Martina Poletti, Jonathan D Victor, and Michele Rucci (2017). “Consequences of the oculomotor cycle for the dynamics of perception”. In: *Current Biology* 27.9, pp. 1268–1277 (cit. on p. 12).

Bonneh, Yoram and Tobias Donner (2011). “Motion induced blindness”. In: *Scholarpedia* 6.6, p. 3321 (cit. on p. 36).

Born, Richard T, Jennifer M Groh, R Zhao, and SJ Lukasewycz (2000). “Segregation of object and background motion in visual area MT: effects of microstimulation on eye movements”. In: *Neuron* 26.3, pp. 725–734 (cit. on p. 300).

Brainard, David H (1997). “The psychophysics toolbox”. In: *Spatial vision* 10.4, pp. 433–436 (cit. on pp. 86, 157, 159).

Brascamp, Jan, Raymond Van Ee, Wiebe Pestman, and Albert Van Den Berg (2005). “Distributions of alternation rates in various forms of bistable perception”. In: *Journal of Vision* 5.4, pp. 1–1 (cit. on pp. 36, 285).

Brascamp, Jan W, Raymond Van Ee, Andre J Noest, Richard HAH Jacobs, and Albert V van den Berg (2006). “The time course of binocular rivalry reveals a fundamental role of noise”. In: *Journal of vision* 6.11, pp. 8–8 (cit. on p. 31).

Brascamp, JW, J Pearson, R Blake, and AV Van Den Berg (2009). “Intermittent ambiguous stimuli: Implicit memory causes periodic perceptual alternations”. In: *Journal of Vision* 9.3, pp. 3–3 (cit. on pp. 40, 45).

Braun, Doris, Lars Pracejus, and Karl Gegenfurtner (2006). “Motion aftereffect elicits smooth pursuit eye movements”. In: *Journal of Vision* 6.7, pp. 1–1 (cit. on pp. 17, 18, 72).

Braun, Jochen and Maurizio Mattia (2010). “Attractors and noise: twin drivers of decisions and multistability”. In: *Neuroimage* 52.3, pp. 740–751 (cit. on pp. 107, 125).

Breese, B (1899). “Inhibition”. In: *Psychological Review* 6.2, pp. 202–203 (cit. on p. 31).

Breitmeyer, Bruno, Haluk Ogmen, and Haluk Öğmen (2006). *Visual masking: Time slices through conscious and unconscious vision*. 41. Oxford University Press (cit. on p. 280).

Breuil, Camille (2018). “Étude du rôle de la couleur dans la perception visuelle des scènes naturelles”. PhD thesis (cit. on p. 277).

Brouwer, Gijs Joost and Raymond van Ee (2006). “Endogenous influences on perceptual bistability depend on exogenous stimulus characteristics”. In: *Vision research* 46.20, pp. 3393–3402 (cit. on p. 35).

– (2007). “Visual cortex allows prediction of perceptual states during ambiguous structure-from-motion”. In: *Journal of Neuroscience* 27.5, pp. 1015–1023 (cit. on pp. 280, 301).

Brown, Richard E and Peter M Milner (2003). “The legacy of Donald O. Hebb: more than the Hebb synapse”. In: *Nature Reviews Neuroscience* 4.12, p. 1013 (cit. on p. 262).

Burr, David C and John Ross (1982). “Contrast sensitivity at high velocities”. In: *Vision research* 22.4, pp. 479–484 (cit. on p. 16).

Bylinskii, Zoya, Tilke Judd, Aude Oliva, Antonio Torralba, and Frédo Durand (2018). “What do different evaluation metrics tell us about saliency models?” In: *IEEE transactions on pattern analysis and machine intelligence* 41.3, pp. 740–757 (cit. on p. 22).

Carl, JR and RS Gellman (1987). “Human smooth pursuit: stimulus-dependent responses”. In: *Journal of Neurophysiology* 57.5, pp. 1446–1463 (cit. on p. 18).

Carmel, David, Michael Arcaro, Sabine Kastner, and Uri Hasson (2010). “How to create and use binocular rivalry”. In: *JoVE (Journal of Visualized Experiments)* 45, e2030 (cit. on p. 34).

Casco, Clara, Patrizio E Tressoldi, and Annamaria Dellantonio (1998). “Visual selective attention and reading efficiency are related in children”. In: *Cortex* 34.4, pp. 531–546 (cit. on p. 263).

Castelluccio, Marco, Giovanni Poggi, Carlo Sansone, and Luisa Verdoliva (2015). “Land use classification in remote sensing images by convolutional neural networks”. In: *arXiv preprint arXiv:1508.00092* (cit. on p. 23).

Castelo-Branco, Miguel, Rainer Goebel, Sergio Neuenschwander, and Wolf Singer (2000). “Neural synchrony correlates with surface segregation rules”. In: *Nature* 405.6787, p. 685 (cit. on pp. 152, 300).

Castet, Eric and Guillaume S Masson (2000). “Motion perception during saccadic eye movements”. In: *Nature neuroscience* 3.2, p. 177 (cit. on p. 16).

Cavanagh, Patrick and George A Alvarez (2005). “Tracking multiple targets with multifocal attention”. In: *Trends in cognitive sciences* 9.7, pp. 349–354 (cit. on p. 107).

Cecotti, Hubert, Bertrand Rivet, Marco Congedo, et al. (2011). “A robust sensor-selection method for P300 brain–computer interfaces”. In: *Journal of neural engineering* 8.1, p. 016001 (cit. on pp. 300, 303).

Cerf, Moran, Jonathan Harel, Wolfgang Einhäuser, and Christof Koch (2008). “Predicting human gaze using low-level saliency combined with face detection”. In: *Advances in neural information processing systems*, pp. 241–248 (cit. on p. 54).

Chalmers, David (2007). “The hard problem of consciousness”. In: *The Blackwell companion to consciousness*, pp. 225–235 (cit. on pp. 12, 264, 279).

Chater, Nick, Joshua B Tenenbaum, and Alan Yuille (2006). *Probabilistic models of cognition: Conceptual foundations* (cit. on p. 58).

Chen, Chih-Yang and Ziad M Hafed (2013). “Postmicrosaccadic enhancement of slow eye movements”. In: *Journal of Neuroscience* 33.12, pp. 5375–5386 (cit. on pp. 74, 79, 105).

Chen, Yi-Chuan, Su-Ling Yeh, and Charles Spence (2011). “Crossmodal constraints on human perceptual awareness: auditory semantic modulation of binocular rivalry”. In: *Frontiers in Psychology* 2, p. 212 (cit. on p. 148).

Cherici, Claudia, Xutao Kuang, Martina Poletti, and Michele Rucci (2012). “Precision of sustained fixation in trained and untrained observers”. In: *Journal of Vision* 12.6, pp. 31–31 (cit. on p. 124).

Choe, Kyoung Whan, Randolph Blake, and Sang-Hun Lee (2016). “Pupil size dynamics during fixation impact the accuracy and precision of video-based gaze estimation”. In: *Vision research* 118, pp. 48–59 (cit. on pp. 74, 87, 103).

Chopin, Adrien (2012). “Traitements probabilistes implicites de la perception ambiguë en vision humaine”. PhD thesis. Atelier national de Reproduction des Thèses (cit. on pp. 31, 153, 176, 265, 284, 285, 287).

Coey, Charles A, Sebastian Wallot, Michael J Richardson, and Guy Van Orden (2012). “On the structure of measurement noise in eye-tracking”. In: (cit. on p. 283).

Cohen, Bernard, Victor Matsuo, and Theodore Raphan (1977). “Quantitative analysis of the velocity characteristics of optokinetic nystagmus and optokinetic after-nystagmus”. In: *The Journal of physiology* 270.2, pp. 321–344 (cit. on p. 264).

Conel, Jesse LeRoy (1959). *The Cortex of the Twenty-four-month Child*. Harvard University Press (cit. on p. 262).

Corbett, J.J. and J. Chen (2018). “Chapter 20 - The Visual System”. In: *Fundamental Neuroscience for Basic and Clinical Applications (Fifth Edition)*. Ed. by Duane E. Haines and Gregory A. Mihailoff. Fifth Edition. Elsevier, 286–305.e1 (cit. on p. 272).

Corbetta, Maurizio, Erbil Akbudak, Thomas E Conturo, et al. (1998). “A common network of functional areas for attention and eye movements”. In: *Neuron* 21.4, pp. 761–773 (cit. on pp. 55, 107).

Cornsweet, Tom N (1956). “Determination of the stimuli for involuntary drifts and saccadic eye movements”. In: *JOSA* 46.11, pp. 987–993 (cit. on pp. 20, 74).

Cornsweet, Tom N and Hewitt D Crane (1973). “Accurate two-dimensional eye tracker using first and fourth Purkinje images”. In: *JOSA* 63.8, pp. 921–928 (cit. on p. 283).

Cosmelli, Diego, Olivier David, Jean-Philippe Lachaux, et al. (2004). “Waves of consciousness: ongoing cortical patterns during binocular rivalry”. In: *Neuroimage* 23.1, pp. 128–140 (cit. on pp. 289, 291, 298, 301).

Coutrot, Antoine, Janet H Hsiao, and Antoni B Chan (2018). “Scanpath modeling and classification with hidden Markov models”. In: *Behavior research methods* 50.1, pp. 362–379 (cit. on pp. 22, 24).

Craighero, Laila, Arturo Carta, and Luciano Fadiga (2001). “Peripheral oculomotor palsy affects orienting of visuospatial attention”. In: *NeuroReport* 12.15, pp. 3283–3286 (cit. on p. 54).

Craighero, Laila, Mauro Nascimben, and Luciano Fadiga (2004). “Eye position affects orienting of visuospatial attention”. In: *Current Biology* 14.4, pp. 331–333 (cit. on p. 54).

Crane, Tim and Craig French (2017). “The Problem of Perception”. In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Spring 2017. Metaphysics Research Lab, Stanford University (cit. on p. 256).

Crick, Francis and Christof Koch (1990). “Towards a neurobiological theory of consciousness”. In: *Seminars in the Neurosciences*. Vol. 2. Saunders Scientific Publications, pp. 263–275 (cit. on p. 299).

Crouzet, Sébastien M, Holle Kirchner, and Simon J Thorpe (2010). “Fast saccades toward faces: face detection in just 100 ms”. In: *Journal of vision* 10.4, pp. 16–16 (cit. on p. 41).

Cunitz, Robert Jesse (1970). “Relationship between slow drift and smooth pursuit eye movements”. In: (cit. on p. 74).

Curcio, Christine A and Kimberly A Allen (1990). “Topography of ganglion cells in human retina”. In: *Journal of comparative Neurology* 300.1, pp. 5–25 (cit. on p. 11).

St-Cyr, GJ and DH Fender (1969). “Nonlinearities of the human oculomotor system: gain”. In: *Vision research* 9.10, pp. 1235–1246 (cit. on p. 18).

Dam, Loes van and Raymond van Ee (2005). “The role of (micro) saccades and blinks in perceptual bi-stability from slant rivalry”. In: *Vision research* 45.18, pp. 2417–2435 (cit. on pp. 36, 40, 42, 43, 45, 142, 295, 298).

Dam, Loes CJ van and Raymond van Ee (2006a). “Retinal image shifts, but not eye movements per se, cause alternations in awareness during binocular rivalry”. In: *Journal of Vision* 6.11, pp. 3–3 (cit. on pp. 43, 44, 295).

– (2006b). “The role of saccades in exerting voluntary control in perceptual and binocular rivalry”. In: *Vision research* 46.6-7, pp. 787–799 (cit. on p. 142).

Damasio, Antonio R (2006). *Descartes’ error*. Random House (cit. on p. 263).

Dayan, Peter, Laurence F Abbott, and L Abbott (2001). “Theoretical neuroscience: computational and mathematical modeling of neural systems”. In: (cit. on pp. 48, 263).

De Brouwer, Sophie, Demet Yuksel, Gunnar Blohm, Marcus Missal, and Philippe Lefèvre (2002). “What triggers catch-up saccades during visual tracking?” In: *Journal of Neurophysiology* 87.3, pp. 1646–1650 (cit. on pp. 18, 71).

Dehaene, Stanislas and Jean-Pierre Changeux (2011). “Experimental and theoretical approaches to conscious processing”. In: *Neuron* 70.2, pp. 200–227 (cit. on pp. 281, 299).

Dehaene, Stanislas and Lionel Naccache (2001). “Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework”. In: *Cognition* 79.1-2, pp. 1–37 (cit. on pp. 265, 280).

Dehaene, Stanislas, Lionel Naccache, Laurent Cohen, et al. (2001). “Cerebral mechanisms of word masking and unconscious repetition priming”. In: *Nature neuroscience* 4.7, p. 752 (cit. on pp. 265, 280).

Delorme, André and Michelangelo Flückiger (2003). *Perception et réalité: Introduction à la psychologie des perceptions*. De Boeck Supérieur (cit. on pp. 256, 260).

Denison, Rachel N, Shlomit Yuval-Greenberg, and Marisa Carrasco (2019). “Directing voluntary temporal attention increases fixational stability”. In: *Journal of Neuroscience* 39.2, pp. 353–363 (cit. on p. 12).

Dennett, Daniel C (1993). *Consciousness explained*. Penguin uk (cit. on pp. 12, 263, 264).

Descartes, René et al. (1970). *Traité de l’homme*. Vol. 11. Gallimard (cit. on p. 263).

Desimone, Robert and John Duncan (1995). “Neural mechanisms of selective visual attention”. In: *Annual review of neuroscience* 18.1, pp. 193–222 (cit. on p. 54).

Devillez, Hélène, Emmanuelle Kristensen, Nathalie Guyader, Bertrand Rivet, and Anne Guérin-Dugué (2015). “The P300 potential for fixations onto target object when exploring natural scenes during a visual task after denoising overlapped EFRP”. In: *2015 7th international IEEE/EMBS conference on neural engineering (NER)*. IEEE, pp. 1024–1027 (cit. on pp. 301, 304).

Dieter, Kevin C, Jan Brascamp, Duje Tadin, and Randolph Blake (2016). “Does visual attention drive the dynamics of bistable perception?” In: *Attention, Perception, & Psychophysics* 78.7, pp. 1861–1873 (cit. on pp. 41, 208).

Ditchburn, RW and BL Ginsborg (1953). “Involuntary eye movements during fixation”. In: *The Journal of physiology* 119.1, pp. 1–17 (cit. on pp. 73, 74).

Dodge, Raymond (1907). “An experimental study of visual fixation.” In: *The Psychological Review: Monograph Supplements* 8.4, p. i (cit. on p. 74).

Dodge, Raymond and Thomas Sparks Cline (1901). “The angular velocity of eye movements”. In: *Psychological Review* 8, pp. 145–157 (cit. on p. 283).

Donchin, Emanuel (1981). “Surprise!... surprise?” In: *Psychophysiology* 18.5, pp. 493–513 (cit. on pp. 300, 303).

Drugowitsch, Jan, Rubén Moreno-Bote, Anne K Churchland, Michael N Shadlen, and Alexandre Pouget (2012). “The cost of accumulating evidence in perceptual decision making”. In: *Journal of Neuroscience* 32.11, pp. 3612–3628 (cit. on p. 57).

Dubois, Julien and Rufin VanRullen (2011). “Visual trails: do the doors of perception open periodically?” In: *PLoS biology* 9.5, e1001056 (cit. on pp. 281, 300).

Eagleman, David M (2001). “Visual illusions and neurobiology”. In: *Nature Reviews Neuroscience* 2.12, p. 920 (cit. on pp. 25, 26).

Eccles, John C (1965). “Cerebral synaptic mechanisms”. In: *Brain and conscious experience*. Springer, pp. 24–58 (cit. on p. 262).

Eckstein, Maria K, Belén Guerra-Carrillo, Alison T Miller Singley, and Silvia A Bunge (2017). “Beyond eye gaze: What else can eyetracking reveal about cognition and cognitive development?” In: *Developmental cognitive neuroscience* 25, pp. 69–91 (cit. on p. 13).

Ee, Raymond van, Wendy J Adams, and Pascal Mamassian (2003). “Bayesian modeling of cue interaction: bistability in stereoscopic slant perception”. In: *JOSA A* 20.7, pp. 1398–1406 (cit. on p. 287).

Ee, Raymond van, Loes Van Dam, and Gijs Brouwer (2005). “Voluntary control and the dynamics of perceptual bi-stability”. In: *Vision research* 45.1, pp. 41–55 (cit. on pp. 37, 43, 208).

Ehm, Werner, Michael Bach, and Jürgen Kornmeier (2011). “Ambiguous figures and binding: EEG frequency modulations during multistable perception”. In: *Psychophysiology* 48.4, pp. 547–558 (cit. on p. 152).

Einhäuser, Wolfgang, Kevan AC Martin, and Peter König (2004). “Are switches in perception of the Necker cube related to eye position?” In: *European Journal of Neuroscience* 20.10, pp. 2811–2818 (cit. on p. 43).

Einhäuser, Wolfgang, James Stout, Christof Koch, and Olivia Carter (2008). “Pupil dilation reflects perceptual selection and predicts subsequent stability in perceptual rivalry”. In: *Proceedings of the National Academy of Sciences* 105.5, pp. 1704–1709 (cit. on pp. 45, 206).

Einhäuser, Wolfgang, Sabine Thomassen, and Alexandra Bendixen (2017). “Using binocular rivalry to tag foreground sounds: Towards an objective visual measure for auditory multistability”. In: *Journal of vision* 17.1, pp. 34–34 (cit. on pp. 142, 295, 298).

Einstein, Albert (1956). *Investigations on the Theory of the Brownian Movement*. Courier Corporation (cit. on pp. 48, 204, 309).

Ekstrom, Leeland B, Pieter R Roelfsema, John T Arsenault, Giorgio Bonmassar, and Wim Vanduffel (2008). “Bottom-up dependent gating of frontal signals in early visual cortex”. In: *Science* 321.5887, pp. 414–417 (cit. on p. 56).

Engbert, Ralf and Reinhold Kliegl (2004). “Microsaccades keep the eyes’ balance during fixation”. In: *Psychological science* 15.6, pp. 431–431 (cit. on pp. 20, 74, 107).

– (2003). “Microsaccades uncover the orientation of covert attention”. In: *Vision research* 43.9, pp. 1035–1045 (cit. on pp. 20, 87–89, 104, 118, 142, 312, 319).

Engbert, Ralf and Konstantin Mergenthaler (2006). “Microsaccades are triggered by low retinal image slip”. In: *Proceedings of the National Academy of Sciences* 103.18, pp. 7192–7197 (cit. on pp. 75, 79, 82).

Engbert, Ralf, Konstantin Mergenthaler, Petra Sinn, and Arkady Pikovsky (2011). “An integrated model of fixational eye movements and microsaccades”. In: *Proceedings of the National Academy of Sciences* 108.39, E765–E770 (cit. on pp. 20, 21, 67, 107, 114, 115, 122, 135, 141).

Engbert, Ralf, Hans A Trukenbrod, Simon Barthelmé, and Felix A Wichmann (2015). “Spatial statistics and attentional dynamics in scene viewing”. In: *Journal of vision* 15.1, pp. 14–14 (cit. on pp. 54, 56, 107).

Engelmann, Felix, Shravan Vasishth, Ralf Engbert, and Reinhold Kliegl (2013). “A framework for modeling the interaction of syntactic processing and eye movement control”. In: *Topics in cognitive science* 5.3, pp. 452–474 (cit. on p. 13).

Epelboim, Julie and Eileen Kowler (1993). “Slow control with eccentric targets: evidence against a position-corrective model”. In: *Vision research* 33.3, pp. 361–380 (cit. on pp. 76, 79, 82, 105).

Esterman, Michael and David Rothlein (2019). “Models of sustained attention”. In: *Current Opinion in Psychology* 29. Attention Perception, pp. 174–180 (cit. on pp. 56, 107).

Faivre, Nathan, Elisa Filevich, Guillermo Solovey, Simone Kühn, and Olaf Blanke (2018). “Behavioral, modeling, and electrophysiological evidence for supramodality in human metacognition”. In: *Journal of Neuroscience* 38.2, pp. 263–277 (cit. on p. 299).

Faivre, Nathan, Laurène Vuillaume, Fosco Bernasconi, et al. (2020). “Sensorimotor conflicts alter metacognitive and action monitoring”. In: *Cortex* 124, pp. 224–234 (cit. on p. 299).

Fang, Fang and Sheng He (2004). “Stabilized structure from motion without disparity induces disparity adaptation”. In: *Current Biology* 14.3, pp. 247–251 (cit. on pp. 35, 36).

Faul, Franz, Edgar Erdfelder, Axel Buchner, and Albert-Georg Lang (2009). “Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses”. In: *Behavior research methods* 41.4, pp. 1149–1160 (cit. on p. 96).

Feinberg, Todd E and Jon Mallatt (2013). “The evolutionary and genetic origins of consciousness in the Cambrian Period over 500 million years ago”. In: *Frontiers in psychology* 4, p. 667 (cit. on p. 258).

Feldman, Anatol G (2016). “Active sensing without efference copy: referent control of perception”. In: *Journal of neurophysiology* 116.3, pp. 960–976 (cit. on p. 13).

Fentress, John C (1999). “The organization of behaviour revisited.” In: *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale* 53.1, p. 8 (cit. on p. 262).

Ferrera, Vincent P and Stephen G Lisberger (1997). “Neuronal responses in visual areas MT and MST during smooth pursuit target selection”. In: *Journal of Neurophysiology* 78.3, pp. 1433–1446 (cit. on p. 300).

Firestone, Chaz and Brian J Scholl (2016). “Cognition does not affect perception: Evaluating the evidence for “top-down” effects”. In: *Behavioral and brain sciences* 39 (cit. on pp. 56, 308).

Flamant, Julien (2018). “A general approach for the analysis and filtering of bivariate signals”. PhD thesis. Thèse de doctorat, Centrale Lille (cit. on pp. 213, 315).

Fodor, Jerry A and Zenon W Pylyshyn (1988). “Connectionism and cognitive architecture: A critical analysis”. In: *Cognition* 28.1-2, pp. 3–71 (cit. on p. 259).

Frässle, Stefan, Jens Sommer, Andreas Jansen, Marnix Naber, and Wolfgang Einhäuser (2014). “Binocular rivalry: frontal activity relates to introspection and action but not to perception”. In: *Journal of Neuroscience* 34.5, pp. 1738–1747 (cit. on pp. 12, 26, 44, 65, 142, 206, 208, 280, 296, 298, 301, 303).

Freund, Harald and Peter Grassberger (1992). “The red queen’s walk”. In: *Physica A: Statistical Mechanics and its Applications* 190.3-4, pp. 218–237 (cit. on p. 114).

Friston, Karl (2010). “The free-energy principle: a unified brain theory?” In: *Nature reviews neuroscience* 11.2, p. 127 (cit. on pp. 58, 59, 260).

Friston, Karl and Stefan Kiebel (2009). “Predictive coding under the free-energy principle”. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 364.1521, pp. 1211–1221 (cit. on pp. 12, 26, 58, 146).

Funke, Klaus, Nicolas J Kerscher, and Florentin Wörgötter (2007). “Noise-improved signal detection in cat primary visual cortex via a well-balanced stochastic resonance-like procedure”. In: *European Journal of Neuroscience* 26.5, pp. 1322–1332 (cit. on p. 52).

Gabbiani, Fabrizio and Christof Koch (1998). “Principles of spike train analysis”. In: *Methods in neuronal modeling* 12.4, pp. 313–360 (cit. on p. 262).

Galdi, Chiara, Michele Nappi, Daniel Riccio, and Harry Wechsler (2016). “Eye movement analysis for human authentication: a critical survey”. In: *Pattern Recognition Letters* 84, pp. 272–283 (cit. on p. 23).

Gammaitoni, Luca, Peter Hänggi, Peter Jung, and Fabio Marchesoni (1998). “Stochastic resonance”. In: *Reviews of modern physics* 70.1, p. 223 (cit. on pp. 52, 63, 107, 125, 126).

García-Pérez, Miguel A (1998). “Forced-choice staircases with fixed step sizes: asymptotic and small-sample properties”. In: *Vision research* 38.12, pp. 1861–1881 (cit. on p. 193).

– (1989). “Visual inhomogeneity and eye movements in multistable perception”. In: *Attention, Perception, & Psychophysics* 46.4, pp. 397–400 (cit. on p. 295).

Garg, Arun, Daniel Schwartz, and Alexander A Stevens (2007). “Orienting auditory spatial attention engages frontal eye fields and medial occipital cortex in congenitally blind humans”. In: *Neuropsychologia* 45.10, pp. 2307–2321 (cit. on p. 55).

Gerstner, Wulfram, Werner M Kistler, Richard Naud, and Liam Paninski (2014). *Neuronal dynamics: From single neurons to networks and models of cognition*. Cambridge University Press (cit. on p. 114).

Gervain, Judit and Jacques Mehler (2010). "Speech perception and language acquisition in the first year of life". In: *Annual review of psychology* 61, pp. 191–218 (cit. on p. 259).

Gibson, James J (2014). *The ecological approach to visual perception: classic edition*. Psychology Press (cit. on p. 268).

Gide, Milind S, Lina J Karam, et al. (2017). "Computational visual attention models". In: *Foundations and Trends® in Signal Processing* 10.4, pp. 347–427 (cit. on pp. 23, 53, 54, 107, 113).

Glasgow, Rupert (2018). *Minimal Selfhood and the Origins of Consciousness*. BoD–Books on Demand (cit. on p. 260).

Glenberg, Arthur M and Michael P Kaschak (2003). "The body's contribution to language". In: *Psychology of learning and motivation* 43, pp. 93–126 (cit. on pp. 260, 264).

Glickstein, Mitchell (1988). "The discovery of the visual cortex". In: *Scientific American* 259.3, pp. 118–127 (cit. on p. 275).

Godfrey-Smith, Peter (2002). "Environmental complexity and the evolution of cognition". In: *The evolution of intelligence*, pp. 233–249 (cit. on p. 263).

Gold, Joshua I and Michael N Shadlen (2003). "The influence of behavioral context on the representation of a perceptual decision in developing oculomotor commands". In: *Journal of Neuroscience* 23.2, pp. 632–651 (cit. on p. 12).

Goldberg, ME, HM Eggers, and P Gouras (1991). "The ocular motor system. Principles of Neural Science". In: *Kandel ER, Schwartz JH, Jessell TM* (cit. on p. 16).

Gonze, Didier, Leo Lahti, Jeroen Raes, and Karoline Faust (2017). "Multi-stability and the origin of microbial community types". In: *The ISME journal* 11.10, p. 2159 (cit. on p. 29).

Goodale, Melvyn A and A David Milner (1992). "Separate visual pathways for perception and action". In: *Trends in neurosciences* 15.1, pp. 20–25 (cit. on p. 278).

Gorea, Andrei and Jean Lorenceau (1991). "Directional performances with moving plaids: component-related and plaid-related processing modes coexist". In: *Spatial vision* 5.4, pp. 231–252 (cit. on p. 145).

Green, David Marvin, John A Swets, et al. (1966). *Signal detection theory and psychophysics*. Vol. 1. Wiley New York (cit. on pp. 172, 174, 193, 298).

Guyader, Nathalie, Alan Chauvin, Muriel Boucart, and Carole Peyrin (2017). "Do low spatial frequencies explain the extremely fast saccades towards human faces?" In: *Vision research* 133, pp. 100–111 (cit. on p. 41).

Hafed, Ziad M, Laurent Goffart, and Richard J Krauzlis (2009). "A neural mechanism for microsaccade generation in the primate superior colliculus". In: *science* 323.5916, pp. 940–943 (cit. on pp. 13, 124, 141).

Hafed, Ziad M and Richard J Krauzlis (2006). “Ongoing eye movements constrain visual perception”. In: *Nature Neuroscience* 9.11, p. 1449 (cit. on p. 12).

Hamedi, HR, SH Asadpour, M Sahrai, B Arzhang, and D Taherkhani (2013). “Optical bistability and multi-stability in a four-level atomic scheme”. In: *Optical and Quantum Electronics* 45.3, pp. 295–306 (cit. on p. 29).

Hannan, Edward James (2009). *Multiple time series*. Vol. 38. John Wiley & Sons (cit. on p. 116).

Hardin, Clyde L (1988). *Color for philosophers: Unweaving the rainbow*. Hackett Publishing (cit. on p. 25).

Harris, Christopher M and Daniel M Wolpert (2006). “The main sequence of saccades optimizes speed-accuracy trade-off”. In: *Biological cybernetics* 95.1, pp. 21–29 (cit. on p. 16).

Hatfield, Gary (2001). “Perception: History of the concept”. In: (cit. on p. 256).

Haynes, John-Dylan and Geraint Rees (2005). “Predicting the stream of consciousness from activity in human visual cortex”. In: *Current Biology* 15.14, pp. 1301–1307 (cit. on pp. 277, 280, 289, 301).

Hebb, Donald Olding and DO Hebb (1949). *The organization of behavior*. Vol. 65. Wiley New York (cit. on p. 280).

Heinen, Stephen J and Scott NJ Watamaniuk (1998). “Spatial integration in human smooth pursuit”. In: *Vision research* 38.23, pp. 3785–3794 (cit. on pp. 17, 18, 72).

Hendry, Stewart H and Takashi Yoshioka (1994). “A neurochemically distinct third channel in the macaque dorsal lateral geniculate nucleus”. In: *Science* 264.5158, pp. 575–577 (cit. on p. 277).

Hérault, Jeanny (2010). *Vision: Images, signals and neural networks: Models of neural processing in visual perception*. World Scientific (cit. on p. 262).

Hérault, Jeanny and Christian Jutten (1994). *Réseaux neuronaux et traitement du signal*. Hermes (cit. on pp. 57, 262).

Herrmann, Carl, Ralf Metzler, and Ralf Engbert (2017). “A self-avoiding walk with neural delays as a model of fixational eye movements”. In: *Scientific Reports* 7.1, p. 12958 (cit. on pp. 107, 122).

Herrmann, Christoph S, Daniel Strüber, Randolph F Helfrich, and Andreas K Engel (2016). “EEG oscillations: from correlation to causality”. In: *International Journal of Psychophysiology* 103, pp. 12–21 (cit. on pp. 301, 303).

Hess, Eckhard H and James M Polt (1960). “Pupil size as related to interest value of visual stimuli”. In: *Science* 132.3423, pp. 349–350 (cit. on p. 270).

– (1964). “Pupil size in relation to mental activity during simple problem-solving”. In: *Science* 143.3611, pp. 1190–1192 (cit. on p. 270).

Hessels, Roy S, Diederick C Niehorster, Marcus Nyström, Richard Andersson, and Ignace TC Hooge (2018). "Is the eye-movement field confused about fixations and saccades? A survey among 124 researchers". In: *Royal Society open science* 5.8, p. 180502 (cit. on pp. 19, 22).

Heywood, Simon and John Churcher (1972). "Eye movements and the after-image—II the effect of foveal and non-foveal after-images on saccadic behaviour". In: *Vision research* 12.5, pp. 1033–1043 (cit. on pp. 18, 77, 105, 124, 283).

– (1971). "Eye movements and the afterimage—I. Tracking the afterimage". In: *Vision Research* 11.10, pp. 1163–1168 (cit. on pp. 18, 77, 105, 283).

Hicheur, Halim, Steeve Zozor, Aurelie Campagne, and Alan Chauvin (2013). "Microsaccades are modulated by both attentional demands of a visual discrimination task and background noise". In: *Journal of vision* 13.13, pp. 18–18 (cit. on pp. 12, 19, 20, 104, 142).

Hillis, James M, Marc O Ernst, Martin S Banks, and Michael S Landy (2002). "Combining sensory information: mandatory fusion within, but not between, senses". In: *Science* 298.5598, pp. 1627–1630 (cit. on p. 25).

Hoffman, James E and Baskaran Subramaniam (1995). "The role of visual attention in saccadic eye movements". In: *Perception & psychophysics* 57.6, pp. 787–795 (cit. on pp. 54, 56).

Hopfield, John J (1982). "Neural networks and physical systems with emergent collective computational abilities". In: *Proceedings of the national academy of sciences* 79.8, pp. 2554–2558 (cit. on p. 262).

Hoppe, Sabrina and Andreas Bulling (2016). "End-to-end eye movement detection using convolutional neural networks". In: *arXiv preprint arXiv:1609.02452* (cit. on p. 19).

Hothorn, Torsten, Kurt Hornik, Mark A Van De Wiel, and Achim Zeileis (2006). "A lego system for conditional inference". In: *The American Statistician* 60.3, pp. 257–263 (cit. on p. 88).

Hothorn, Torsten, Kurt Hornik, Mark A. van de Wiel, and Achim Zeileis (2008). "Implementing a class of permutation tests: The coin package". In: *Journal of Statistical Software* 28.8, pp. 1–23 (cit. on p. 183).

Hou, Xiaodi and Liqing Zhang (2007). "Saliency detection: A spectral residual approach". In: *2007 IEEE Conference on Computer Vision and Pattern Recognition*. Ieee, pp. 1–8 (cit. on p. 53).

Hubel, David H and Torsten N Wiesel (1968). "Receptive fields and functional architecture of monkey striate cortex". In: *The Journal of physiology* 195.1, pp. 215–243 (cit. on pp. 262, 275).

– (1959). "Receptive fields of single neurones in the cat's striate cortex". In: *The Journal of physiology* 148.3, pp. 574–591 (cit. on pp. 275, 276).

– (1962). "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex". In: *The Journal of physiology* 160.1, pp. 106–154 (cit. on pp. 262, 275).

Huey, Edmund Burke (1908). *The psychology and pedagogy of reading*. The Macmillan Company (cit. on p. 283).

Huguet, Gemma, John Rinzel, and Jean-Michel Hupé (2014). "Noise and adaptation in multistable perception: Noise drives when to switch, adaptation determines percept choice". In: *Journal of vision* 14.3, pp. 19–19 (cit. on pp. 48, 51, 62, 67, 146, 148, 171, 173, 207).

Hupé, Jean-Michel, Lu-Ming Joffo, and Daniel Pressnitzer (2008). "Bistability for audiovisual stimuli: Perceptual decision is modality specific". In: *Journal of Vision* 8.7, pp. 1–1 (cit. on p. 46).

Hupé, Jean-Michel, Cédric Lamirel, and Jean Lorenceau (2009). "Pupil dynamics during bistable motion perception". In: *Journal of vision* 9.7, pp. 10–10 (cit. on pp. 45, 206, 282, 295, 296).

Hupé, Jean-Michel and Nava Rubin (2003). "The dynamics of bi-stable alternation in ambiguous motion displays: a fresh look at plaids". In: *Vision research* 43.5, pp. 531–548 (cit. on pp. 29, 36, 39, 143, 144, 148, 150, 156, 157, 171, 173, 191, 207, 285, 287, 288, 308).

– (2004). "The oblique plaid effect". In: *Vision Research* 44.5, pp. 489–500 (cit. on pp. 39, 144, 149, 151, 156, 157, 171, 173, 207).

Hupé, Jean-Michel, Camilo Miguel Signorelli, and David Alais (2019). "Two paradigms of bistable plaid motion reveal independent mutual inhibition processes". In: *Journal of vision* 19.4, pp. 5–5 (cit. on pp. 48, 51, 63, 146, 207).

Ilg, Uwe J (1997). "Slow eye movements". In: *Progress in neurobiology* 53.3, pp. 293–329 (cit. on p. 73).

Intaitė, Monika, Mika Koivisto, and Miguel Castelo-Branco (2014). "Event-related potential responses to perceptual reversals are modulated by working memory load". In: *Neuropsychologia* 56, pp. 428–438 (cit. on p. 291).

Ishizu, Tomohiro and Semir Zeki (2014). "Varieties of perceptual instability and their neural correlates". In: *NeuroImage* 91, pp. 203–209 (cit. on pp. 36, 293).

Itti, Laurent, Christof Koch, and Ernst Niebur (1998). "A model of saliency-based visual attention for rapid scene analysis". In: *IEEE Transactions on Pattern Analysis & Machine Intelligence* 11, pp. 1254–1259 (cit. on p. 53).

James, William, Frederick Burkhardt, Fredson Bowers, and Ignas K Skrupskelis (1890). *The principles of psychology*. Vol. 1. 2. Macmillan London (cit. on p. 53).

Jaynes, Julian (2000). *The origin of consciousness in the breakdown of the bicameral mind*. Houghton Mifflin Harcourt (cit. on p. 264).

Jeannerod, M and M Arbib (2003). "Action monitoring and forward control of movements". In: *The Handbook of Brain Theory and Neural Networks,* pp. 83–85 (cit. on p. 13).

Jones, E.G. (2004). "Cerebral Cortex". In: *Encyclopedia of Neuroscience*. Ed. by Larry R. Squire. Oxford: Academic Press, pp. 769–773 (cit. on p. 276).

Jones, Judson P and Larry A Palmer (1987a). “An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex”. In: *Journal of neurophysiology* 58.6, pp. 1233–1258 (cit. on p. 277).

– (1987b). “The two-dimensional spatial structure of simple receptive fields in cat striate cortex”. In: *Journal of neurophysiology* 58.6, pp. 1187–1211 (cit. on p. 277).

Judd, Tilke, Frédo Durand, and Antonio Torralba (2012). “A benchmark of computational models of saliency to predict human fixations”. In: (cit. on p. 22).

Just, Marcel Adam and Patricia A Carpenter (1976). “Eye fixations and cognitive processes”. In: *Cognitive psychology* 8.4, pp. 441–480 (cit. on p. 14).

Kagan, Igor and David C Burr (2017). “Active Vision: Dynamic Reformatting of Visual Information by the Saccade-Drift Cycle”. In: *Current Biology* 27.9, R341–R344 (cit. on p. 12).

Kagan, Igor and Ziad M Hafed (2013). “Active vision: microsaccades direct the eye to where it matters most”. In: *Current Biology* 23.17, R712–R714 (cit. on pp. 12, 25).

Kaiser, Peter K (2004). *The Joy of visual perception [: a Web book*. York University (cit. on p. 268).

Kalogeropoulou, Zampeta and Martin Rolfs (2017). “Saccadic eye movements do not disrupt the deployment of feature-based attention”. In: *Journal of vision* 17.8, pp. 4–4 (cit. on pp. 54, 56).

Kanai, Ryota, Bahador Bahrami, and Geraint Rees (2010). “Human parietal cortex structure predicts individual differences in perceptual rivalry”. In: *Current biology* 20.18, pp. 1626–1630 (cit. on pp. 280, 301, 303).

Kassambara, Alboukadel (2020). *rstatix: Pipe-Friendly Framework for Basic Statistical Tests*. R package version 0.5.0 (cit. on p. 88).

Kawamoto, Alan H and James A Anderson (1985). “A neural network model of multistable perception”. In: *Acta psychologica* 59.1, pp. 35–65 (cit. on p. 299).

Keating, EG (1991). “Frontal eye field lesions impair predictive and visually-guided pursuit eye movements”. In: *Experimental Brain Research* 86.2, pp. 311–323 (cit. on p. 141).

Kelley, Carl T (1999). *Iterative methods for optimization*. SIAM (cit. on p. 48).

Kelso, JA Scott (2012). “Multistability and metastability: understanding dynamic coordination in the brain”. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 367.1591, pp. 906–918 (cit. on pp. 13, 29, 59, 61, 64, 108, 121, 125, 135, 216).

Kersten, Daniel, Pascal Mamassian, and Alan Yuille (2004). “Object perception as Bayesian inference”. In: *Annu. Rev. Psychol.* 55, pp. 271–304 (cit. on p. 25).

Kersten, Daniel and Alan Yuille (2003). “Bayesian models of object perception”. In: *Current opinion in neurobiology* 13.2, pp. 150–158 (cit. on p. 25).

Kim, Yee-Joon, Marcia Grabowecky, and Satoru Suzuki (2006). “Stochastic resonance in binocular rivalry”. In: *Vision research* 46.3, pp. 392–406 (cit. on pp. 52, 63, 107, 125).

King, Jean-Rémi and Stanislas Dehaene (2014). “A model of subjective report and objective discrimination as categorical decisions in a vast representational space”. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 369.1641, p. 20130204 (cit. on p. 298).

Klein, Stanley B, Leda Cosmides, John Tooby, and Sarah Chance (2002). “Decisions and the evolution of memory: multiple systems, multiple functions.” In: *Psychological review* 109.2, p. 306 (cit. on p. 258).

Kliegl, Reinhold, Ellen Grabner, Martin Rolfs, and Ralf Engbert (2004). “Length, frequency, and predictability effects of words on eye movements in reading”. In: *European journal of cognitive psychology* 16.1-2, pp. 262–284 (cit. on p. 13).

Kloeden, Peter E and Eckhard Platen (2013). *Numerical solution of stochastic differential equations*. Vol. 23. Springer Science & Business Media (cit. on p. 117).

Knapen, Tomas, Jan Willem de Gee, Jan Brascamp, et al. (2016). “Cognitive and ocular factors jointly determine pupil responses under equiluminance”. In: *PLoS One* 11.5, e0155574 (cit. on p. 282).

Knill, David C and Whitman Richards (1996). *Perception as Bayesian inference*. Cambridge University Press (cit. on p. 25).

Ko, Hee-kyoung, Martina Poletti, and Michele Rucci (2010). “Microsaccades precisely relocate gaze in a high visual acuity task”. In: *Nature neuroscience* 13.12, p. 1549 (cit. on pp. 71, 73, 104).

Ko, Hee-kyoung, D Max Snodderly, and Martina Poletti (2016). “Eye movements between saccades: Measuring ocular drift and tremor”. In: *Vision research* 122, pp. 93–104 (cit. on p. 73).

Koch, Christof, Marcello Massimini, Melanie Boly, and Giulio Tononi (2016). “Neural correlates of consciousness: progress and problems”. In: *Nature Reviews Neuroscience* 17.5, p. 307 (cit. on p. 299).

Kohler, Axel, Leila Haddad, Wolf Singer, and Lars Muckli (2008). “Deciding what to see: The role of intention and attention in the perception of apparent motion”. In: *Vision Research* 48.8, pp. 1096–1106 (cit. on pp. 56, 106, 208).

Köhler, Wolfgang (1929). “Gestalt psychology.” In: (cit. on p. 25).

Komogortsev, Oleg V and Alex Karpov (2013). “Automated classification and scoring of smooth pursuit eye movements in the presence of fixations and saccades”. In: *Behavior research methods* 45.1, pp. 203–215 (cit. on pp. 17, 71).

Kornmeier, Jürgen and Michael Bach (2012). “Ambiguous figures–what happens in the brain when perception changes but not the stimulus”. In: *Frontiers in human neuroscience* 6, p. 51 (cit. on pp. 12, 34, 38–42, 66, 152, 208, 289–292, 300, 301, 303, 324).

– (2014). “EEG correlates of perceptual reversals in Boring’s ambiguous old/young woman stimulus”. In: *Perception* 43.9, pp. 950–962 (cit. on pp. 27, 152, 290).

Kornmeier, Jürgen, Werner Ehm, Heiko Bigalke, and Michael Bach (2007). “Discontinuous presentation of ambiguous figures: How interstimulus-interval durations affect reversal dynamics and ERPs”. In: *Psychophysiology* 44.4, pp. 552–560 (cit. on pp. 38, 40, 66, 173, 208, 290).

Kowler, Eileen (2011). “Eye movements: The past 25 years”. In: *Vision research* 51.13, pp. 1457–1483 (cit. on pp. 14, 20, 73).

Kowler, Eileen, Albert J Martins, and M Pavel (1984). “The effect of expectations on slow oculomotor control—IV. Anticipatory smooth eye movements depend on prior target motions”. In: *Vision research* 24.3, pp. 197–210 (cit. on pp. 76, 105, 114).

Kowler, Eileen and Robert M Steinman (1979a). “The effect of expectations on slow oculomotor control—II. Single target displacements”. In: *Vision research* 19.6, pp. 633–646 (cit. on pp. 76, 105).

– (1979b). “The effect of expectations on slow oculomotor control—II. Single target displacements”. In: *Vision research* 19.6, pp. 633–646 (cit. on pp. 76, 105).

– (1981). “The effect of expectations on slow oculomotor control—III. Guessing unpredictable target displacements”. In: *Vision research* 21.2, pp. 191–203 (cit. on pp. 76, 105).

Krauzlis, Richard J (2004). “Recasting the smooth pursuit eye movement system”. In: *Journal of neurophysiology* 91.2, pp. 591–603 (cit. on pp. 13, 17, 18, 71, 124, 141).

– (2005). “The control of voluntary eye movements: new perspectives”. In: *The Neuroscientist* 11.2, pp. 124–137 (cit. on pp. 124, 141).

Krauzlis, Richard J, Laurent Goffart, and Ziad M Hafed (2017). “Neuronal control of fixation and fixational eye movements”. In: *Phil. Trans. R. Soc. B* 372.1718, p. 20160205 (cit. on pp. 13, 19, 21, 71, 103, 124, 141).

Kristensen, Emmanuelle (2017). “Méthodologie de traitement conjoint des signaux EEG et oculométriques: applications aux tâches d’exploration visuelle libre”. PhD thesis (cit. on pp. 301, 304).

Kristensen, Emmanuelle, Bertrand Rivet, and Anne Guérin-Dugué (2017). “Estimation of overlapped Eye Fixation Related Potentials: The General Linear Model, a more flexible framework than the ADJAR algorithm”. In: *Journal of Eye Movement Research* 10.1, pp. 1–27 (cit. on p. 304).

Kuang, Xutao, Martina Poletti, Jonathan D Victor, and Michele Rucci (2012). “Temporal encoding of spatial information during active visual fixation”. In: *Current Biology* 22.6, pp. 510–514 (cit. on p. 122).

Kuhn, Gustav, Benjamin W Tatler, and Geoff G Cole (2009). “You look where I look! Effect of gaze cues on overt and covert attention in misdirection”. In: *Visual Cognition* 17.6-7, pp. 925–944 (cit. on pp. 12, 54, 56).

Kustov, Alexander and David Robinson (1996). “Shared neural control of attentional shifts and eye movements”. In: *Nature* 384.6604, p. 74 (cit. on p. 55).

Lakoff, George and Mark Johnson (2008). *Metaphors we live by*. University of Chicago press (cit. on p. 264).

Lamme, Victor AF (2006). "Towards a true neural stance on consciousness". In: *Trends in cognitive sciences* 10.11, pp. 494–501 (cit. on pp. 279, 280, 299).

Langevin, Paul (1908). "Sur la théorie du mouvement brownien". In: *Compt. Rendus* 146, pp. 530–533 (cit. on p. 116).

Lau, Hakwan and David Rosenthal (2011). "The higher-order view does not require consciously self-directed introspection: response to Malach". In: *Trends in cognitive sciences* 15.11, pp. 508–509 (cit. on p. 299).

Laubrock, Jochen, Ralf Engbert, and Reinhold Kliegl (2008). "Fixational eye movements predict the perceived direction of ambiguous apparent motion". In: *Journal of Vision* 8.14, pp. 13–13 (cit. on pp. 114, 142, 208, 295, 298).

Le Meur, Olivier and Thierry Baccino (2013). "Methods for comparing scanpaths and saliency maps: strengths and weaknesses". In: *Behavior research methods* 45.1, pp. 251–266 (cit. on p. 22).

Le Meur, Olivier, Patrick Le Callet, Dominique Barba, and Dominique Thoreau (2006). "A coherent computational approach to model bottom-up visual attention". In: *IEEE transactions on pattern analysis and machine intelligence* 28.5, pp. 802–817 (cit. on p. 22).

Lehky, Sidney R (1995). "Binocular rivalry is not chaotic". In: *Proceedings of the Royal Society of London. Series B: Biological Sciences* 259.1354, pp. 71–76 (cit. on pp. 29, 48, 63, 285).

Leopold, David and Nikos Logothetis (1999). "Multistable phenomena: changing views in perception". In: *Trends in cognitive sciences* 3.7, pp. 254–264 (cit. on pp. 26, 29, 33, 280, 289, 293, 301).

Leopold, David, Melanie Wilke, Alexander Maier, and Nikos Logothetis (2002). "Stable perception of visually ambiguous patterns". In: *Nature neuroscience* 5.6, p. 605 (cit. on pp. 40, 45, 284, 288, 300, 303).

Leopold, David Anthony (1997). "Brain mechanisms of visual awareness". PhD thesis. Baylor College of Medicine. Division of Neuroscience (cit. on pp. 265, 280).

Levelt, Willem JM (1967). "Note on the distribution of dominance times in binocular rivalry". In: *British Journal of Psychology* 58.1-2, pp. 143–145 (cit. on pp. 29, 49, 159, 285, 286).

– (1966). "The alternation process in binocular rivalry". In: *British Journal of Psychology* 57.3-4, pp. 225–238 (cit. on pp. 30, 49, 153, 159).

Lewicki, Michael S, Bruno A Olshausen, Annemarie Surlykke, and Cynthia F Moss (2014). "Scene analysis in the natural environment". In: *Frontiers in psychology* 5, p. 199 (cit. on p. 25).

Li, Hsin-Hung, James Rankin, John Rinzel, Marisa Carrasco, and David J Heeger (2017). "Attention model of binocular rivalry". In: *Proceedings of the National Academy of Sciences* 114.30, E6192–E6201 (cit. on pp. 41, 56, 208, 216, 303).

Lisberger, Stephen G, EJ Morris, and Lawrence Tychsen (1987). "Visual motion processing and sensory-motor integration for smooth pursuit eye movements". In: *Annual review of neuroscience* 10.1, pp. 97–129 (cit. on p. 16).

Liu, Zhi, Le Meur, and Shuhua Luo (2013). "Superpixel-based saliency detection". In: *2013 14th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS)*. IEEE, pp. 1–4 (cit. on p. 53).

Liversedge, Simon, Iain Gilchrist, and Stefan Everling (2011). *The Oxford handbook of eye movements*. Oxford University Press (cit. on pp. 12, 14, 15, 17, 71, 79, 123).

Long, Gerald M and Thomas C Toppino (2004). "Enduring interest in perceptual ambiguity: alternating views of reversible figures." In: *Psychological bulletin* 130.5, p. 748 (cit. on p. 43).

Long, Gerald M, Thomas C Toppino, and Gregory W Mondin (1992). "Prime time: Fatigue and set effects in the perception of reversible figures". In: *Perception & psychophysics* 52.6, pp. 609–616 (cit. on p. 288).

Lumer, Erik D, Karl J Friston, and Geraint Rees (1998). "Neural correlates of perceptual rivalry in the human brain". In: *Science* 280.5371, pp. 1930–1934 (cit. on p. 293).

Mach, Ernst (1901). *Die Mechanik in ihrer Entwickelung historisch-kritisch dargestellt*. Vol. 59. Brockhaus (cit. on p. 35).

Madelain, Laurent and Richard J Krauzlis (2003). "Pursuit of the ineffable: perceptual and motor reversals during the tracking of apparent motion". In: *Journal of Vision* 3.11, pp. 1–1 (cit. on pp. 17, 18, 72, 206, 295).

Magallón-Garcıa, DA, R Jaimes-Reátegui, G Huerta-Cuellar, et al. (2017). "Study of multi-stable visual perception by stochastic modulation using a synergetic model". In: *Cybernetics and Physics* 6.3, pp. 135–140 (cit. on p. 216).

Magnussen, Svein and Mark Greenlee (1999). "The psychophysics of perceptual memory". In: *Psychological research* 62.2-3, pp. 81–92 (cit. on p. 45).

Malhotra, Gaurav, David S Leslie, Casimir JH Ludwig, and Rafal Bogacz (2017). "Overcoming indecision by changing the decision boundary." In: *Journal of Experimental Psychology: General* 146.6, p. 776 (cit. on p. 58).

Mamassian, Pascal (2006). "Métamères perceptifs et perception bistable". In: *Intellectica* 43.1, pp. 73–77 (cit. on pp. 25, 26, 265).

Mamassian, Pascal and Ross Goutcher (2005). "Temporal dynamics in bistable perception". In: *Journal of Vision* 5.4, pp. 7–7 (cit. on pp. 29, 31, 40, 45, 66, 173, 287, 298).

Mao, Xiaochen (2012). "Stability switches, bifurcation, and multi-stability of coupled networks with time delays". In: *Applied Mathematics and Computation* 218.11, pp. 6263–6274 (cit. on p. 29).

Marat, Sophie, Tien Ho Phuoc, Lionel Granjon, et al. (2009). "Modelling spatio-temporal saliency to predict gaze direction for short videos". In: *International journal of computer vision* 82.3, p. 231 (cit. on p. 113).

Martinez-Conde, Susana and Stephen Macknik (2017). *Champions of Illusion: The Science Behind Mind-boggling Images and Mystifying Brain Puzzles*. Scientific American/Farrar, Straus and Giroux (cit. on p. 26).

Martinez-Conde, Susana, Stephen Macknik, and David Hubel (2004). "The role of fixational eye movements in visual perception". In: *Nature Reviews Neuroscience* 5.3, p. 229 (cit. on pp. 12, 19–21, 73, 118, 264).

Martins, Albert J, Eileen Kowler, and Caroline Palmer (1985). "Smooth pursuit of small-amplitude sinusoidal motion". In: *JOSA A* 2.2, pp. 234–242 (cit. on pp. 77, 81, 105, 106).

Mashour, George A and Michael T Alkire (2013). "Evolution of consciousness: Phylogeny, ontogeny, and emergence from general anesthesia". In: *Proceedings of the National Academy of Sciences* 110.Supplement 2, pp. 10357–10364 (cit. on p. 258).

Massey Jr, Frank J (1951). "The Kolmogorov-Smirnov test for goodness of fit". In: *Journal of the American statistical Association* 46.253, pp. 68–78 (cit. on p. 128).

Masson, Guillaume and Leland Stone (2002). "From following edges to pursuing objects". In: *Journal of neurophysiology* 88.5, pp. 2869–2873 (cit. on pp. 18, 72).

Mathes, Birgit, Daniel Strüber, Michael A Stadler, and Canan Basar-Eroglu (2006). "Voluntary control of Necker cube reversals modulates the EEG delta-and gamma-band response". In: *Neuroscience letters* 402.1-2, pp. 145–149 (cit. on p. 301).

McDonnell, Mark D and Derek Abbott (2009). "What is stochastic resonance? Definitions, misconceptions, debates, and its relevance to biology". In: *PLoS computational biology* 5.5, e1000348 (cit. on pp. 52, 63).

McManus, Ian C, Matthew Freegard, James Moore, and Richard Rawles (2010). "Science in the making: Right hand, left hand. II: The duck–rabbit figure". In: *Laterality* 15.1-2, pp. 166–185 (cit. on p. 30).

Meghanathan, Radha Nila, Cees van Leeuwen, and Andrey R Nikolaev (2015). "Fixation duration surpasses pupil size as a measure of memory load in free viewing". In: *Frontiers in human neuroscience* 8, p. 1063 (cit. on p. 45).

Megumi, Fukuda, Bahador Bahrami, Ryota Kanai, and Geraint Rees (2015). "Brain activity dynamics in human parietal regions during spontaneous switches in bistable perception". In: *NeuroImage* 107, pp. 190–197 (cit. on pp. 146, 301).

Mele, Maria Laura and Stefano Federici (2012). "A psychotechnological review on eye-tracking systems: towards user experience". In: *Disability and Rehabilitation: Assistive Technology* 7.4, pp. 261–281 (cit. on p. 283).

Meng, Ming and Frank Tong (2004). "Can attention selectively bias bistable perception? Differences between binocular rivalry and ambiguous figures". In: *Journal of vision* 4.7, pp. 2–2 (cit. on p. 37).

Meyberg, Susann, Petra Sinn, Ralf Engbert, and Werner Sommer (2017). "Revising the link between microsaccades and the spatial cueing of voluntary attention". In: *Vision Research* 133, pp. 47–60 (cit. on pp. 54, 56).

Michalski, A, M Kossut, and B Żernicki (1977). “The ocular following reflex elicited from the retinal periphery in the cat”. In: *Vision research* 17.6, pp. 731–736 (cit. on pp. 22, 72).

Michel, Matthias and Jorge Morales (2019). “Minority Reports: Consciousness and the Prefrontal Cortex”. In: (cit. on pp. 280, 298, 299).

Mihali, Andra, Bas van Opheusden, and Wei Ji Ma (2017). “Bayesian microsaccade detection”. In: *Journal of vision* 17.1, pp. 13–13 (cit. on pp. 16, 20, 71).

Milner, A David and Melvyn A Goodale (2008). “Two visual systems re-viewed”. In: *Neuropsychologia* 46.3, pp. 774–785 (cit. on p. 278).

Milner, AD and MA Goodale (1995). *Oxford psychology series, No. 27* (cit. on p. 278).

Mirza, M Berk, Rick A Adams, Karl Friston, and Thomas Parr (2019). “Introducing a Bayesian model of selective attention based on active inference”. In: *Scientific reports* 9.1, pp. 1–22 (cit. on p. 56).

Mishkin, Mortimer, Leslie G Ungerleider, and Kathleen A Macko (1983). “Object vision and spatial vision: two cortical pathways”. In: *Trends in neurosciences* 6, pp. 414–417 (cit. on p. 278).

Mitra, Chiranjit, Jürgen Kurths, and Reik V Donner (2015). “An integrative quantifier of multistability in complex systems based on ecological resilience”. In: *Scientific reports* 5, p. 16196 (cit. on p. 29).

Mitroff, Stephen R, David M Sobel, and Alison Gopnik (2006). “Reversing how to think about ambiguous figure reversals: Spontaneous alternating by uninformed observers”. In: *Perception* 35.5, pp. 709–715 (cit. on pp. 59, 284).

Moore, Tirin and Katherine M Armstrong (2003). “Selective gating of visual signals by microstimulation of frontal cortex”. In: *Nature* 421.6921, p. 370 (cit. on p. 55).

Moore, Tirin and Mazyar Fallah (2001). “Control of eye movements and spatial attention”. In: *Proceedings of the National Academy of Sciences* 98.3, pp. 1273–1276 (cit. on p. 55).

Moreno-Bote, Rubén and Jan Drugowitsch (2015). “Causal inference and explaining away in a spiking network”. In: *Scientific reports* 5, p. 17531 (cit. on pp. 25, 57, 107, 122).

Moreno-Bote, Rubén, David C Knill, and Alexandre Pouget (2011). “Bayesian sampling in visual perception”. In: *Proceedings of the National Academy of Sciences* 108.30, pp. 12491–12496 (cit. on pp. 13, 25, 48–51, 58, 67, 107, 121, 126, 136, 141, 151, 298).

Moreno-Bote, Rubén and Néstor Parga (2005). “Simple model neurons with AMPA and NMDA filters: role of synaptic time scales”. In: *Neurocomputing* 65, pp. 441–448 (cit. on p. 49).

Moreno-Bote, Rubén, John Rinzel, and Nava Rubin (2007). “Noise-induced alternations in an attractor network model of perceptual bistability”. In: *Journal of neurophysiology* 98.3, pp. 1125–1139 (cit. on pp. 46–49, 107, 121, 123, 141, 298).

Moreno-Bote, Rubén, Asya Shpiro, John Rinzel, and Nava Rubin (2010). “Alternation rate in perceptual bistability is maximal at and symmetric around equi-dominance”. In: *Journal of Vision* 10.11, pp. 1–1 (cit. on pp. 36, 39, 136, 144, 145, 151, 153, 173, 207).

Moreno-Bote, Rubén, Asya Shpiro, John Rinzel, and Nava Rubin (2008). “Bi-stable depth ordering of superimposed moving gratings”. In: *Journal of Vision* 8.7, pp. 20–20 (cit. on pp. 36, 37, 39, 136, 143–145, 151, 173, 207).

Muirhead, Robb J (2009). *Aspects of multivariate statistical theory*. Vol. 197. John Wiley & Sons (cit. on p. 82).

Murphy, Brian J, Eileen Kowler, and Robert M Steinman (1975). “Slow oculomotor control in the presence of moving backgrounds”. In: *Vision research* 15.11, pp. 1263–1268 (cit. on pp. 75, 77).

Naber, Marnix, Stefan Frässle, and Wolfgang Einhäuser (2011). “Perceptual rivalry: reflexes reveal the gradual nature of visual awareness”. In: *PLoS One* 6.6, e20910 (cit. on pp. 206, 208).

Nachmias, Jacob (1961). “Determiners of the drift of the eye during monocular fixation”. In: *Josa* 51.7, pp. 761–766 (cit. on pp. 73, 74).

Necker, Louis Albert (1832). “LXI. Observations on some remarkable optical phænomena seen in Switzerland; and on an optical phænomenon which occurs on viewing a figure of a crystal or geometrical solid”. In: *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 1.5, pp. 329–337 (cit. on pp. 28–30, 34, 43).

Nelder, John A and Roger Mead (1965). “A simplex method for function minimization”. In: *The computer journal* 7.4, pp. 308–313 (cit. on p. 183).

Newsome, William T and Edmond B Pare (1988). “A selective impairment of motion perception following lesions of the middle temporal visual area (MT)”. In: *Journal of Neuroscience* 8.6, pp. 2201–2211 (cit. on p. 300).

Nilsson, Dan-E and Susanne Pelger (1994). “A pessimistic estimate of the time required for an eye to evolve”. In: *Proceedings of the Royal Society of London. Series B: Biological Sciences* 256.1345, pp. 53–58 (cit. on p. 25).

Nishimoto, Shinji, Tsugitaka Ishida, and Izumi Ohzawa (2006). “Receptive field properties of neurons in the early visual cortex revealed by local spectral reverse correlation”. In: *Journal of Neuroscience* 26.12, pp. 3269–3280 (cit. on p. 277).

Nobre, Anna C, DR Gitelman, EC Dias, and Marek-Marsel Mesulam (2000). “Covert visual spatial orienting and saccades: overlapping neural systems”. In: *Neuroimage* 11.3, pp. 210–216 (cit. on p. 55).

Noton, David and Lawrence Stark (1971a). “Scanpaths in eye movements during pattern perception”. In: *Science* 171.3968, pp. 308–311 (cit. on p. 19).

– (1971b). “Scanpaths in saccadic eye movements while viewing and recognizing patterns”. In: *Vision research* 11.9, 929–IN8 (cit. on p. 19).

Núñez, Rafael, Michael Allen, Richard Gao, et al. (2019). “What happened to cognitive science?” In: *Nature human behaviour* 3.8, pp. 782–791 (cit. on pp. 260, 261).

Nyström, Marcus and Kenneth Holmqvist (2010). “An adaptive algorithm for fixation, saccade, and glissade detection in eyetracking data”. In: *Behavior research methods* 42.1, pp. 188–204 (cit. on pp. 16, 71).

Orban de Xivry, Jean-Jacques and Philippe Lefevre (2007). “Saccades and pursuit: two outcomes of a single sensorimotor process”. In: *The Journal of physiology* 584.1, pp. 11–23 (cit. on pp. 72, 79).

Orquin, Jacob L and Simone Mueller Loose (2013). “Attention and choice: A review on eye movements in decision making”. In: *Acta psychologica* 144.1, pp. 190–206 (cit. on pp. 12, 113).

Otero-Millan, Jorge, Stephen L Macknik, Rachel E Langston, and Susana Martinez-Conde (2013). “An oculomotor continuum from exploration to fixation”. In: *Proceedings of the National Academy of Sciences* 110.15, pp. 6175–6180 (cit. on p. 71).

Palmer, Stephen (1999). *Vision science: Photons to phenomenology*. MIT press (cit. on pp. 11, 16, 71, 172, 174, 266, 269, 274, 278, 279).

Parisot, Kevin, Alan Chauvin, Anne Guérin-Dugué, Ronald Phlypo, and Steeve Zozor (2017). “A multistable gravitational potential approach to fixational eye movements”. In: *European Conference on Visual Perception (ECVP 2017)* (cit. on p. 114).

Parisot, Kevin, Alan Chauvin, Ronald Phlypo, and Steeve Zozor (Aug. 2019). “Modélisation de l’ambiguïté d’une multi-stabilité visuelle”. In: *XXVIIème colloque GRETSI (GRETSI 2019)*. Lille, France (cit. on pp. 156, 172, 185, 186, 192, 202).

– (2018). “Multi-stability and fixationnal eye movements: an energy potential fields modeling approach”. In: (cit. on p. 115).

Park, Hyeong-Dong, Stéphanie Correia, Antoine Ducorps, and Catherine Tallon-Baudry (2014). “Spontaneous fluctuations in neural responses to heartbeats predict visual detection”. In: *Nature neuroscience* 17.4, p. 612 (cit. on p. 264).

Parkkonen, Lauri, Jesper Andersson, Matti Hämäläinen, and Riitta Hari (2008). “Early visual brain areas reflect the percept of an ambiguous scene”. In: *Proceedings of the National Academy of Sciences* 105.51, pp. 20500–20504 (cit. on pp. 27, 30, 39, 40, 277, 280, 289, 291, 292, 298, 301).

Parr, Thomas and Karl Friston (2019). “Attention or salience?” In: *Current opinion in psychology* 29, pp. 1–5 (cit. on p. 56).

Pastukhov, Alexander and Jochen Braun (2013). “Disparate time-courses of adaptation and facilitation in multi-stable perception”. In: *Learning & Perception* 5.Supplement 2, pp. 101–118 (cit. on p. 63).

Patel, Ashok and Bart Kosko (2005). “Stochastic resonance in noisy spiking retinal and sensory neuron models”. In: *Neural Networks* 18.5-6, pp. 467–478 (cit. on p. 125).

Pearson, Joel and Jan Brascamp (2008). “Sensory memory for ambiguous vision”. In: *Trends in cognitive sciences* 12.9, pp. 334–341 (cit. on p. 293).

Pearson, Joel, Duje Tadin, and Randolph Blake (2007). “The effects of transcranial magnetic stimulation on visual rivalry”. In: *Journal of vision* 7.7, pp. 2–2 (cit. on pp. 38, 289).

Peel, Tyler R, Ziad M Hafed, Suryadeep Dash, Stephen G Lomber, and Brian D Corneil (2016). “A causal role for the cortical frontal eye fields in microsaccade deployment”. In: *PLoS biology* 14.8, e1002531 (cit. on pp. 13, 124).

Penrose, Roger and Palle ET Jorgensen (2006). “The road to reality: A complete guide to the laws of the universe”. In: *The Mathematical Intelligencer* 28.3, pp. 59–61 (cit. on p. 267).

Penrose, Roger and N David Mermin (1990). *The emperor’s new mind: Concerning computers, minds, and the laws of physics* (cit. on p. 264).

Petersen, Steven E and Michael I Posner (2012). “The attention system of the human brain: 20 years after”. In: *Annual review of neuroscience* 35, pp. 73–89 (cit. on p. 53).

Ho-Phuoc, Tien, Anne Guérin-Dugué, and Nathalie Guyader (2009). “A Computational Saliency Model Integrating Saccade Programming.” In: *BIOSIGNALS*, pp. 57–64 (cit. on p. 113).

Piaget, Jean and Margaret Cook (1952). *The origins of intelligence in children*. Vol. 8. 5. International Universities Press New York (cit. on p. 259).

Pilly, Praveen K and Aaron R Seitz (2009). “What a difference a parameter makes: A psychophysical comparison of random dot motion algorithms”. In: *Vision research* 49.13, pp. 1599–1612 (cit. on pp. 204, 309).

Pinna, Baingio and Stephen Grossberg (2005). “The watercolor illusion and neon color spreading: a unified analysis of new cases and neural mechanisms”. In: *JOSA A* 22.10, pp. 2207–2221 (cit. on p. 25).

Pisarchik, Alexander N, Rider Jaimes-Reátegui, CD Alejandro Magallón-Garcıa, and C Obed Castillo-Morales (2014). “Critical slowing down and noise-induced intermittency in bistable perception: bifurcation analysis”. In: *Biological Cybernetics* 108.4, pp. 397–404 (cit. on p. 216).

Poincaré, Henri (1885). “Sur l’équilibre d’une masse fluide animée d’un mouvement de rotation”. In: *Acta mathematica* 7.1, pp. 259–380 (cit. on p. 216).

Poletti, Martina, Chiara Listorti, and Michele Rucci (2013). “Microscopic eye movements compensate for nonhomogeneous vision within the fovea”. In: *Current Biology* 23.17, pp. 1691–1695 (cit. on pp. 19, 20, 105).

– (2010). “Stability of the visual world during eye drift”. In: *Journal of Neuroscience* 30.33, pp. 11143–11150 (cit. on pp. 74, 78).

Poletti, Martina and Michele Rucci (2016). “A compact field guide to the study of microsaccades: Challenges and functions”. In: *Vision research* 118, pp. 83–97 (cit. on pp. 19, 20, 73, 104).

Posner, Michael I (1980). “Orienting of attention”. In: *Quarterly journal of experimental psychology* 32.1, pp. 3–25 (cit. on pp. 54, 56).

Posner, Michael I and Stanislas Dehaene (1994). “Attentional networks”. In: *Trends in neurosciences* 17.2, pp. 75–79 (cit. on pp. 53, 54).

Posner, Michael I and Steven E Petersen (1990). “The attention system of the human brain”. In: *Annual review of neuroscience* 13.1, pp. 25–42 (cit. on p. 53).

Press, William H and Saul A Teukolsky (1990). “Savitzky-Golay smoothing filters”. In: *Computers in Physics* 4.6, pp. 669–672 (cit. on p. 313).

Pressnitzer, Daniel and Jean-Michel Hupé (2005). “Is auditory streaming a bistable percept”. In: *Forum Acusticum, Budapest*, pp. 1557–1561 (cit. on pp. 36, 285).

– (2006). “Temporal dynamics of auditory and visual bistability reveal common principles of perceptual organization”. In: *Current biology* 16.13, pp. 1351–1357 (cit. on pp. 35–37, 286).

Putnam, Nicole M, Heidi J Hofer, Nathan Doble, et al. (2005). “The locus of fixation and the foveal cone mosaic”. In: *Journal of Vision* 5.7, pp. 3–3 (cit. on p. 21).

Quaia, Christian, Boris M Sheliga, Edmond J FitzGibbon, and Lance M Optican (2012). “Ocular following in humans: Spatial properties”. In: *Journal of Vision* 12.4, pp. 13–13 (cit. on pp. 22, 72).

Rajashekar, Umesh, Lawrence K Cormack, and Alan C Bovik (2004). “Point-of-gaze analysis reveals visual search strategies”. In: *Human vision and electronic imaging IX*. Vol. 5292. International Society for Optics and Photonics, pp. 296–306 (cit. on p. 22).

Rashbass, C1 (1961). “The relationship between saccadic and smooth tracking eye movements”. In: *The Journal of Physiology* 159.2, pp. 326–338 (cit. on p. 17).

Ratliff, Floyd and Lorrin A Riggs (1950). “Involuntary motions of the eye during monocular fixation.” In: *Journal of experimental psychology* 40.6, p. 687 (cit. on p. 74).

Rees, Geraint (2007). “Neural correlates of the contents of visual awareness in humans”. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 362.1481, pp. 877–886 (cit. on pp. 208, 280, 289, 298, 301).

Remington, Lee Ann (2012). “Chapter 13 - Visual Pathway”. In: *Clinical Anatomy and Physiology of the Visual System (Third Edition)*. Ed. by Lee Ann Remington. Third Edition. Saint Louis: Butterworth-Heinemann, pp. 233–252 (cit. on pp. 276, 278).

Riche, Nicolas, Matthieu Duvinage, Matei Mancas, Bernard Gosselin, and Thierry Dutoit (2013). “Saliency and human fixations: State-of-the-art and study of comparison metrics”. In: *Proceedings of the IEEE international conference on computer vision*, pp. 1153–1160 (cit. on p. 22).

Riggs, Lorrin A, Floyd Ratliff, Janet C Cornsweet, and Tom N Cornsweet (1953). “The disappearance of steadily fixated visual test objects”. In: *JOSA* 43.6, pp. 495–501 (cit. on p. 264).

Rivet, Bertrand, Marc Duda, Anne Guérin-Dugué, Christian Jutten, and Pierre Comon (2015). “Multimodal approach to estimate the ocular movements during EEG recordings: a coupled tensor factorization method”. In: *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, pp. 6983–6986 (cit. on p. 304).

Rizzolatti, Giacomo, Laila Craighero, et al. (1998). “Spatial attention: Mechanisms and theories”. In: *Advances in psychological science* 2, pp. 171–198 (cit. on pp. 53, 54, 141).

Rizzolatti, Giacomo, Lucia Riggio, Isabella Dascola, and Carlo Umiltá (1987). “Reorienting attention across the horizontal and vertical meridians: evidence in favor of a premotor theory of attention”. In: *Neuropsychologia* 25.1, pp. 31–40 (cit. on pp. 53, 54, 141).

Rizzolatti, Giacomo, Lucia Riggio, Boris M Sheliga, et al. (1994). "Space and selective attention". In: *Attention and performance XV* 15, pp. 231–265 (cit. on p. 54).

Robinson, Do (1965). "The mechanics of human smooth pursuit eye movement." In: *The Journal of Physiology* 180.3, pp. 569–591 (cit. on p. 17).

Rock, Irvin and Kurt Mitchener (1992). "Further evidence of failure of reversal of ambiguous figures by uninformed subjects". In: *Perception* 21.1, pp. 39–45 (cit. on p. 288).

Rolfs, Martin (2015). "Attention in active vision: A perspective on perceptual continuity across saccades". In: *Perception* 44.8-9, pp. 900–919 (cit. on p. 56).

– (2009). "Microsaccades: small steps on a long way". In: *Vision research* 49.20, pp. 2415–2441 (cit. on pp. 19, 20, 73, 104, 142).

Rubin, Nava, Jean-Michel Hupé, et al. (2005). "Dynamics of perceptual bistability: Plaids and binocular rivalry compared". In: *Binocular rivalry*, pp. 137–154 (cit. on pp. 37, 144, 148, 149, 171, 173, 207).

Rucci, Michele and Antonino Casile (2005). "Fixational instability and natural image statistics: implications for early visual representations". In: *Network: Computation in Neural Systems* 16.2-3, pp. 121–138 (cit. on pp. 21, 25).

Rucci, Michele, Ramon Iovin, Martina Poletti, and Fabrizio Santini (2007). "Miniature eye movements enhance fine spatial detail". In: *Nature* 447.7146, p. 852 (cit. on pp. 19, 21, 107).

Rucci, Michele and Jonathan D Victor (2015). "The unsteady eye: an information-processing stage, not a bug". In: *Trends in neurosciences* 38.4, pp. 195–206 (cit. on pp. 12, 21, 64, 107).

Ruff, Christian C, Felix Blankenburg, Otto Bjoertomt, et al. (2006). "Concurrent TMS-fMRI and psychophysics reveal frontal influences on human retinotopic visual cortex". In: *Current Biology* 16.15, pp. 1479–1488 (cit. on p. 56).

Ruiz-Mirazo, Kepa, Juli Peretó, and Alvaro Moreno (2004). "A universal definition of life: autonomy and open-ended evolution". In: *Origins of Life and Evolution of the Biosphere* 34.3, pp. 323–346 (cit. on p. 255).

Rumelhart, David E and James L McClelland (1982). "An interactive activation model of context effects in letter perception: II. The contextual enhancement effect and some tests and extensions of the model." In: *Psychological review* 89.1, p. 60 (cit. on p. 263).

– (1986). "On learning the past tenses of English verbs". In: (cit. on p. 259).

Sabrin, Howard W and Andrew E Kertesz (1980). "Microsaccadic eye movements and binocular rivalry". In: *Perception & psychophysics* 28.2, pp. 150–154 (cit. on p. 295).

Sacks, Oliver (1985). "The man who mistook his wifefor a hat". In: *London: Duckworth* (cit. on p. 278).

Santini, Fabrizio, Gabriel Redner, Ramon Iovin, and Michele Rucci (2007). "EyeRIS: a general-purpose system for eye-movement-contingent display control". In: *Behavior Research Methods* 39.3, pp. 350–364 (cit. on p. 283).

Scase, Mark O, Oliver J Braddick, and Jane E Raymond (1996). "What is noise for the motion system?" In: *Vision research* 36.16, pp. 2579–2586 (cit. on pp. 204, 205, 302, 309, 310).

Schröder, H (1858). "Ueber eine optische Inversion bei Betrachtung verkehrter, durch optische Vorrichtung entworfener, physischer Bilder". In: *Annalen der Physik* 181.10, pp. 298–311 (cit. on p. 35).

Schütz, Alexander C, Doris I Braun, and Karl R Gegenfurtner (2011). "Eye movements and perception: A selective review". In: *Journal of vision* 11.5, pp. 9–9 (cit. on p. 12).

Schütz, Alexander C, Doris I Braun, J Anthony Movshon, and Karl R Gegenfurtner (2010). "Does the noise matter? Effects of different kinematogram types on smooth pursuit eye movements and perception". In: *Journal of vision* 10.13, pp. 26–26 (cit. on pp. 204, 309).

Schwartz, Jean-Luc, Nicolas Grimault, Jean-Michel Hupé, Brian CJ Moore, and Daniel Pressnitzer (2012). *Multistability in perception: binding sensory modalities, an overview* (cit. on pp. 13, 108, 125, 148, 208, 217, 288).

Shaikh, Aasef G and David S Zee (2018). *Eye movement research in the twenty-first century—a window to the brain, mind, and more* (cit. on pp. 12, 14).

Shams, Ladan and Ulrik R Beierholm (2010). "Causal inference in perception". In: *Trends in cognitive sciences* 14.9, pp. 425–432 (cit. on pp. 25, 26).

Sheiman, IM, EV Zubina, and ND Kreshchenko (2002). "Regulation of the feeding behavior of the planarian Dugesia (Girardia) tigrina". In: *Journal of Evolutionary Biochemistry and Physiology* 38.4, pp. 414–418 (cit. on p. 258).

Sheliga, BM, L Riggio, and G Rizzolatti (1995). "Spatial attention and eye movements". In: *Experimental brain research* 105.2, pp. 261–275 (cit. on pp. 54, 55).

Shpiro, Asya, Rodica Curtu, John Rinzel, and Nava Rubin (2007). "Dynamical characteristics common to neuronal competition models". In: *Journal of neurophysiology* 97.1, pp. 462–473 (cit. on pp. 46, 48, 51, 126).

Shpiro, Asya, Rubén Moreno-Bote, Nava Rubin, and John Rinzel (2009). "Balance between noise and adaptation in competition models of perceptual bistability". In: *Journal of computational neuroscience* 27.1, p. 37 (cit. on pp. 47–49, 51, 63, 107, 121, 123, 126, 141, 298).

Sinn, Petra and Ralf Engbert (2016). "Small saccades versus microsaccades: Experimental distinction and model-based unification". In: *Vision research* 118, pp. 132–143 (cit. on pp. 20, 73, 103, 104).

Skinner, Julianne, Antimo Buonocore, and Ziad M Hafed (2018). "Transfer function of the rhesus macaque oculomotor system for small-amplitude slow motion trajectories". In: *Journal of neurophysiology* 121.2, pp. 513–529 (cit. on pp. 74, 78, 79, 81, 82, 105).

Smith, Carolyn L, Natalia Pivovarova, and Thomas S Reese (2015). "Coordinated feeding behavior in Trichoplax, an animal without synapses". In: *PloS one* 10.9, e0136098 (cit. on p. 258).

Smith, Marie, Frédéric Gosselin, and Philippe G Schyns (2006). "Perceptual moments of conscious visual experience inferred from oscillatory brain activity". In: *Proceedings of the National Academy of Sciences* 103.14, pp. 5626–5631 (cit. on pp. 291, 298, 300).

Smolensky, Paul (1988). "On the proper treatment of connectionism". In: *Behavioral and brain sciences* 11.1, pp. 1–23 (cit. on p. 259).

Song, Joo-Hyun (2019). "The role of attention in motor control and learning". In: *Current opinion in psychology* (cit. on pp. 41, 56, 208, 303).

Spering, Miriam and Marisa Carrasco (2015). "Acting without seeing: eye movements reveal visual processing without awareness". In: *Trends in neurosciences* 38.4, pp. 247–258 (cit. on p. 12).

Spering, Miriam and Karl R Gegenfurtner (2008). "Contextual effects on motion perception and smooth pursuit eye movements". In: *Brain research* 1225, pp. 76–85 (cit. on pp. 296, 308).

Spering, Miriam and Anna Montagnini (2011). "Do we track what we see? Common versus independent processing for motion perception and smooth pursuit eye movements: A review". In: *Vision research* 51.8, pp. 836–852 (cit. on pp. 17, 18, 71, 72, 106, 142).

Spering, Miriam, Marc Pomplun, and Marisa Carrasco (2011). "Tracking without perceiving: a dissociation between eye movements and motion perception". In: *Psychological science* 22.2, pp. 216–225 (cit. on pp. 14, 296).

Sporns, Olaf, Giulio Tononi, and Gerald M Edelman (1991). "Modeling perceptual grouping and figure-ground segregation by means of active reentrant connections". In: *Proceedings of the National Academy of Sciences* 88.1, pp. 129–133 (cit. on p. 279).

Stern, John A, Donna Boyer, and David Schroeder (1994). "Blink rate: a possible measure of fatigue". In: *Human factors* 36.2, pp. 285–297 (cit. on p. 282).

Sterzer, Philipp, Andreas Kleinschmidt, and Geraint Rees (2009). "The neural bases of multistable perception". In: *Trends in cognitive sciences* 13.7, pp. 310–318 (cit. on pp. 27, 39, 208, 280, 289, 293, 298, 301).

Stone, Leland S, Brent R Beutter, and Jean Lorenceau (2000). "Visual motion integration for perception and pursuit". In: *Perception* 29.7, pp. 771–787 (cit. on pp. 18, 72).

Stone, LS, AB Watson, and JB Mulligan (1990). "Effect of contrast on the perceived direction of a moving plaid". In: *Vision research* 30.7, pp. 1049–1067 (cit. on p. 145).

Stoner, Gene R and Thomas D Albright (1992). "Neural correlates of perceptual motion coherence". In: *Nature* 358.6385, p. 412 (cit. on pp. 145, 300).

Strüber, Daniel, Canan Basar-Eroglu, Edwin Hoff, and Michael Stadler (2000). "Reversal-rate dependent differences in the EEG gamma-band during multistable visual perception". In: *International journal of psychophysiology* 38.3, pp. 243–252 (cit. on p. 301).

Strüber, Daniel and Christoph S Herrmann (2002). "MEG alpha activity decrease reflects destabilization of multistable percepts". In: *Cognitive Brain Research* 14.3, pp. 370–382 (cit. on pp. 280, 301).

Strüber, Daniel, Stefan Rach, Sina A Trautmann-Lengsfeld, Andreas K Engel, and Christoph S Herrmann (2014). "Antiphasic 40 Hz oscillatory current stimulation affects bistable motion perception". In: *Brain topography* 27.1, pp. 158–171 (cit. on p. 301).

Szulborski, Robert G and Larry A Palmer (1990). "The two-dimensional spatial structure of nonlinear subunits in the receptive fields of complex cells". In: *Vision research* 30.2, pp. 249–254 (cit. on p. 277).

Tafaj, Enkelejda, Thomas C Kübler, Gjergji Kasneci, Wolfgang Rosenstiel, and Martin Bogdan (2013). "Online classification of eye tracking data for automated analysis of traffic hazard perception". In: *International Conference on Artificial Neural Networks*. Springer, pp. 442–450 (cit. on p. 16).

Taouali, Wahiba, Laurent Goffart, Frédéric Alexandre, and Nicolas P Rougier (2015). "A parsimonious computational model of visual target position encoding in the superior colliculus". In: *Biological cybernetics* 109.4-5, pp. 549–559 (cit. on pp. 13, 124).

Tatler, Benjamin W, Mary M Hayhoe, Michael F Land, and Dana H Ballard (2011). "Eye guidance in natural vision: Reinterpreting salience". In: *Journal of vision* 11.5, pp. 5–5 (cit. on p. 113).

Thaler, Lore, Alexander C Schütz, Melvyn A Goodale, and Karl R Gegenfurtner (2013). "What is the best fixation target? The effect of target shape on stability of fixational eye movements". In: *Vision Research* 76, pp. 31–42 (cit. on pp. 114, 308).

Thiele, Alexander, Karen R Dobkins, and Thomas D Albright (2000). "Neural correlates of contrast detection at threshold". In: *Neuron* 26.3, pp. 715–724 (cit. on p. 300).

Toet, Alexander (2011). "Computational versus psychophysical bottom-up image saliency: A comparative evaluation study". In: *IEEE transactions on pattern analysis and machine intelligence* 33.11, pp. 2131–2146 (cit. on p. 22).

Tomasello, Michael (2000). "First steps toward a usage-based theory of language acquisition". In: *Cognitive linguistics* 11.1/2, pp. 61–82 (cit. on p. 259).

Tomczak, Maciej and Ewa Tomczak (2014). "The need to report effect size estimates revisited. An overview of some recommended measures of effect size". In: (cit. on p. 88).

Tononi, Giulio, Olaf Sporns, and Gerald M Edelman (1994). "A measure for brain complexity: relating functional segregation and integration in the nervous system". In: *Proceedings of the National Academy of Sciences* 91.11, pp. 5033–5037 (cit. on p. 60).

Tootell, RB, MS Silverman, SL Hamilton, RL De Valois, and E Switkes (1988). "Functional anatomy of macaque striate cortex. III. Color". In: *Journal of Neuroscience* 8.5, pp. 1569–1593 (cit. on p. 275).

Tootell, Roger B, MS Silverman, SL Hamilton, E Switkes, and RL De Valois (1988). "Functional anatomy of macaque striate cortex. V. Spatial frequency". In: *Journal of Neuroscience* 8.5, pp. 1610–1624 (cit. on p. 275).

Toppino, Thomas C and Gerald M Long (1987). "Selective adaptation with reversible figures: Don't change that channel". In: *Perception & Psychophysics* 42.1, pp. 37–48 (cit. on p. 288).

Toyoda, Jun-Ichi, Hiroshi Nosaki, and Tsuneo Tomita (1969). “Light-induced resistance changes in single photoreceptors of Necturus and Gekko”. In: *Vision research* 9.4, pp. 453–463 (cit. on p. 274).

Treutwein, Bernhard (1995). “Adaptive psychophysical procedures”. In: *Vision research* 35.17, pp. 2503–2522 (cit. on p. 193).

Tsuchiya, Naotsugu, Melanie Wilke, Stefan Frässle, and Victor AF Lamme (2015). “No-report paradigms: extracting the true neural correlates of consciousness”. In: *Trends in cognitive sciences* 19.12, pp. 757–770 (cit. on p. 280).

Turing, Alan M (2009). “Computing machinery and intelligence”. In: *Parsing the Turing Test*. Springer, pp. 23–65 (cit. on p. 259).

Ullman, Tomer D, Noah D Goodman, and Joshua B Tenenbaum (2012). “Theory learning as stochastic search in the language of thought”. In: *Cognitive Development* 27.4, pp. 455–480 (cit. on p. 259).

Van Essen, David C and John HR Maunsell (1983). “Hierarchical organization and functional streams in the visual cortex”. In: *Trends in neurosciences* 6, pp. 370–375 (cit. on p. 279).

Van Gulick, Robert (2018). “Consciousness”. In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Spring 2018. Metaphysics Research Lab, Stanford University (cit. on p. 263).

Van Vugt, Bram, Bruno Dagnino, Devavrat Vartak, et al. (2018). “The threshold for conscious report: Signal loss and response bias in visual and frontal cortex”. In: *Science* 360.6388, pp. 537–542 (cit. on p. 280).

VanRullen, Rufin, Niko Busch, Jan Drewes, and Julien Dubois (2011). “Ongoing EEG phase as a trial-by-trial predictor of perceptual and attentional variability”. In: *Frontiers in psychology* 2, p. 60 (cit. on pp. 208, 281, 300).

VanRullen, Rufin and Julien Dubois (2011). “The psychophysics of brain rhythms”. In: *Frontiers in psychology* 2, p. 203 (cit. on p. 300).

VanRullen, Rufin and Christof Koch (2003). “Is perception discrete or continuous?” In: *Trends in cognitive sciences* 7.5, pp. 207–213 (cit. on pp. 208, 281, 300).

VanRullen, Rufin, Lavanya Reddy, and Christof Koch (2004). “Visual search and dual tasks reveal two distinct attentional resources”. In: *Journal of Cognitive Neuroscience* 16.1, pp. 4–14 (cit. on p. 106).

Varela, Francisco J (1996a). “Invitation aux sciences cognitives (Nouv. ed.)” In: *Paris: Éd. du Seuil* (cit. on pp. 12, 260).

– (1996b). “Neurophenomenology: A methodological remedy for the hard problem”. In: *Journal of consciousness studies* 3.4, pp. 330–349 (cit. on p. 264).

Veltman, JA and AWK Gaillard (1998). “Physiological workload reactions to increasing levels of task difficulty”. In: *Ergonomics* 41.5, pp. 656–669 (cit. on pp. 41, 207).

Verhulst, Pierre-François (1838). “Correspondance mathématique et physique”. In: *Ghent and Brussels* 10, p. 113 (cit. on p. 173).

Wade, Nicholas J (1998). “Early studies of eye dominances”. In: *Laterality: Asymmetries of body, brain and cognition* 3.2, pp. 97–108 (cit. on p. 33).

Wandell, B (1995). “Foundations of Human Vision”. In: *Sunderland, USA, Sinauer* (cit. on pp. 11, 25, 260, 270, 271, 275, 277).

Wang, Dong, Fiona B Mulvey, Jeff B Pelz, and Kenneth Holmqvist (2017). “A study of artificial eyes for the measurement of precision in eye-trackers”. In: *Behavior research methods* 49.3, pp. 947–959 (cit. on p. 283).

Watanabe, Masayuki, Ken-ichi Okada, Yuta Hamasaki, et al. (2019). “Ocular drift reflects volitional action preparation”. In: *European Journal of Neuroscience* 50.2, pp. 1892–1910 (cit. on pp. 74, 76, 105).

Watanabe, Takamitsu, Naoki Masuda, Fukuda Megumi, Ryota Kanai, and Geraint Rees (2014). “Energy landscape and dynamics of brain activity during human bistable perception”. In: *Nature communications* 5, p. 4765 (cit. on pp. 121, 215, 216).

Watson, Andrew B (2017). “QUEST+: A general multidimensional Bayesian adaptive psychometric method”. In: *Journal of Vision* 17.3, pp. 10–10 (cit. on pp. 174, 193).

Watson, Andrew B and Denis G Pelli (1983). “QUEST: A Bayesian adaptive psychometric method”. In: *Perception & psychophysics* 33.2, pp. 113–120 (cit. on pp. 174, 193).

Welch, Leslie (1989). “The perception of moving plaids reveals two motion-processing stages”. In: *Nature* 337.6209, p. 734 (cit. on pp. 144, 145).

Werkhoven, Peter, HP Snippe, and JJ Koenderink (1990). “Effects of element orientation on apparent motion perception”. In: *Perception & Psychophysics* 47.6, pp. 509–525 (cit. on p. 157).

Wernery, Jannis (2013). “Bistable perception of the Necker cube: In the context of cognition & personality”. PhD thesis. ETH Zurich (cit. on pp. 29, 32–37, 39, 46, 288).

Wertheimer, Max (1938). “Gestalt theory.” In: (cit. on p. 25).

White, Olivier and Robert M French (2017). “Pupil diameter may reflect motor control and learning”. In: *Journal of motor behavior* 49.2, pp. 141–149 (cit. on p. 282).

Whittington, James CR and Rafal Bogacz (2017). “An approximation of the error backpropagation algorithm in a predictive coding network with local Hebbian synaptic plasticity”. In: *Neural computation* 29.5, pp. 1229–1262 (cit. on p. 60).

Williams, Ziv M, John C Elfar, Emad N Eskandar, Louis J Toth, and John A Assad (2003). “Parietal activity and the perceived direction of ambiguous apparent motion”. In: *Nature neuroscience* 6.6, p. 616 (cit. on pp. 280, 300).

Windmann, Sabine, Michaela Wehrmann, Pasquale Calabrese, and Onur Güntürkün (2006). “Role of the prefrontal cortex in attentional control over bistable vision”. In: *Journal of cognitive neuroscience* 18.3, pp. 456–471 (cit. on p. 280).

Wuerger, Sophie, Robert Shapley, and Nava Rubin (1996). ““On the visually perceived direction of motion” by Hans Wallach: 60 years later”. In: *Perception* 25.11, pp. 1317–1367 (cit. on pp. 25, 144–146, 156).

Wyart, Valentin and Catherine Tallon-Baudry (2008). “Neural dissociation between visual awareness and spatial attention”. In: *Journal of Neuroscience* 28.10, pp. 2667–2679 (cit. on p. 54).

Wyatt, Harry J (2010). “The human pupil and the use of video-based eyetrackers”. In: *Vision research* 50.19, pp. 1982–1988 (cit. on pp. 74, 103).

Yanulevskaya, Victoria, Jasper Uijlings, Jan-Mark Geusebroek, Nicu Sebe, and Arnold Smeulders (2013). “A proto-object-based computational model for visual saliency”. In: *Journal of vision* 13.13, pp. 27–27 (cit. on p. 53).

Yarbus, Alfred L (1967). “Eye movements during perception of complex objects”. In: *Eye movements and vision*. Springer, pp. 171–211 (cit. on pp. 14, 15, 71, 73, 74, 77, 214, 264, 283).

Yeshurun, Yaffa and Marisa Carrasco (1998). “Attention improves or impairs visual performance by enhancing spatial resolution”. In: *Nature* 396.6706, p. 72 (cit. on pp. 41, 53).

Zeljko, Mick, Ada Kritikos, and Philip M Grove (2019). “Temporal dynamics of a perceptual decision”. In: *Journal of vision* 19.5, pp. 7–7 (cit. on p. 57).

Zhang, Jianming and Stan Sclaroff (2013). “Saliency detection: A boolean map approach”. In: *Proceedings of the IEEE international conference on computer vision*, pp. 153–160 (cit. on p. 67).

Zhou, YH, JB Gao, Keith D White, I Merk, and Kung Yao (2004). “Perceptual dominance time distributions in multistable visual perception”. In: *Biological cybernetics* 90.4, pp. 256–263 (cit. on p. 39).

Zozor, Steeve, Pierre-Olivier Amblard, and Cédric Duchêne (2009). “Does eye tremor provide the hyperacuity phenomenon?” In: *Journal of Statistical Mechanics: Theory and Experiment* 2009.01, P01015 (cit. on pp. 21, 107).

Acronyms

2AFC 2-alternatives forced choice. 193

AOA ambient optic array. 268

BCEA bi-variate contour ellipse area. 76, 79, 82

BOLD blood oxygenation level dependent. 289, 293

CSP cumulative smooth pursuit. 207, 208

dERP differential event related potentials. 290, 291

EEG electro-encephalography. 26, 39, 42, 45, 62, 66, 154, 204, 212, 214, 278, 280, 289, 291, 292, 298, 300–304

EfRP eye fixation related potentials. 304

EK Engber-Kliegl. 88, 118–120, 312, 319

EMG electro-myography. 39, 42

ERP event relate potential. 39, 40, 290, 291, 300, 303, 304

EsRP eye saccade related potentials. 304

FEF frontal eye field. 13, 55, 56, 124, 141

FEM fixational eye movements. 19–21, 68, 69, 71, 73, 74, 78–83, 89, 94, 101, 103–107, 109, 113–115, 118, 122–124, 135, 137, 197, 199, 212, 214, 215, 295, 302, 304, 307

FFA fusiform face area. 293

fMRI functional magnetic resonance imaging. 27, 39, 42, 44, 55, 56, 215, 216, 278, 289, 293, 298, 300, 301

GraFEM gravitational fixational eye movements. 114, 117–123, 126, 127, 141, 203, 213–215, 315

ISI inter-stimulus-interval. 40, 41, 45, 284, 290, 303

IT infero-temporal. 294

LGN lateral geniculate nucleus. 13, 274–277, 279, 294

LIP lateral intra-parietal. 13, 55, 124, 141

LOC lateral occipital complex. 293

MA moving average. 116, 127

mad mean absolute deviation. 72, 91

MAE motion after-effect. 17, 18, 48

MCMC Markov-chain Monte Carlo. 193

MEG magneto-encephalography. 27, 39, 42, 66, 214, 278, 280, 289, 291, 292, 298, 300, 301

MPC maximally projected correlation. 78, 81–83, 87–94, 96, 98, 100–103, 105, 107, 109, 118, 120, 122, 213, 306, 307, 315

MST medial superior temporal. 294

MT middle temporal. 27, 294, 300, 301

NaN not a number. 312, 319

NCC neural correlates of consciousness. 66, 206, 214, 280, 281, 299

OFR oculo-following reflex. 22, 72–75, 205

OKN opto-kinetic nystagmus. 22, 44, 45, 65, 72, 73, 205–208, 212, 213, 295, 296, 308

PET positron emission tomography. 278

PFC pre-frontal cortex. 299

RDK random-dot kinematogram. 17, 18, 57, 203–205, 302, 308–315

SC superior colliculus. 13, 55, 124, 141, 274, 277

SNR signal to noise ratio. 306, 307

STS superior temporal sulcus. 294

TF time-frequency. 39, 40, 66, 292, 301

TMS trans-magnetic stimulation. 37, 56, 278, 289

VOR vestibulo-ocular reflex. 22, 72, 73, 77

A Complementary information on the literature review

A.1 Theoretical context

Before diving into the puzzling and complex mysteries of visual perception, one should step back to see where does this phenomenology stand in our current scientific culture. The study of visual phenomena has been linked to many disciplines covering physics, philosophy, biology, cognitive psychology, neurosciences, applied mathematics, signal processing, and more. This work is inscribed in this transdisciplinary tradition, now referred to as Cognitive Sciences, and aims to address how visual ambiguity is treated and processed in the human brain by combining methods from signal processing, physics, psychology and neuro-imaging, to research original approaches. Therefore, covering the context and background of this research work is essential to give light to the reader, regarding the inspiration it draws from, but also its potential application perspectives.

A brief review of how perception fits into the evolutionary process of living organism's complexity will be presented, followed by a more specific picture of the current understanding of perception in the larger study of the brain and intelligence.

A.1.1 Evolution of natural complex systems with perception

First and foremost, it is important to remind the reader that cognitive sciences relate to the study of cognitive systems, and in this work, we focus on biological, living systems, more precisely human cognition. It is necessary to first define what science considers and defines as **living**.

> "*A living thing, or a self, is a part of the universe that sustains itself and makes more of its kind.*"

The definition proposed above by Ruiz-Mirazo et al. (2004) is a recent and universal attempt to constrain the systems of study. In that work, the authors propose requirements around the boundaries of a living organism, the energy transduction

apparatus, and the presence of functionally interdependent macro-molecular components. Another aspect discussed lies in the relationship between the organism and its environment, as one needs to interact with it in order to grow and replicate. More simply, though there is no formal consensus and a unifying theory of biology, a system is considered as living when it tends to be complex, organised and with behavioural dynamics. Therefore, living organisms tend to develop and feature perceptual capacities and mechanisms to draw information from their environment, from which they can obtain the energy needed to grow and reproduce.

Sensing the environment

Perception has been intensely studied and theorised by philosophers for centuries, and even more so with the rise of psychology as a science as it poses many problems linked to the notions of reality, awareness or consciousness (T. Crane and C. French, 2017). The definition of **perception** as a phenomenon of study has evolved since Aristotle, Ptolemy and Ibn al-Haytham's intro-mission and extra-mission theories (Hatfield, 2001). Indeed, with the advent of modern science, and with increased knowledge drawn from psycho-physics, the phenomenon has been centred around the relationship between a manipulable parameter in a physical stimulus and the associated cognitive performance for detection, awareness, representation and inference. As defined in the Oxford English Dictionary:

> **Perception**: the neuro-physiological processes, including memory, by which an organism becomes aware of and interprets external stimuli.

To perceive, organisms need to interact with their physical environment and process sensation signals. The physical properties are extracted and converted into information through encoding processes. Perception can be described as a three component phenomenon (Delorme and Flückiger, 2003):

1. physical stimulation,
2. physiological sensing,
3. psychological and cognitive representation.

> **Essentially, information on the physics of one's environment varies. These variations are captured through different modes or channels, and the information is fused and interpreted by the brain in order to be used for other cognitive tasks, such as decision making for instance. Sensory information can be acquired in a variety of ways**

i.e., chemical reaction, acoustical vibrations, photon detection, etc, which are distinguished as modalities of perception; in other words, senses.

Perceiving the world can be achieved *via* various methods. These are often categorised as modalities or senses. For instance for humans, olfaction relies on chemical reactions to estimate the composition of the air surrounding, while audition is based on the vibration of the eardrums, and for the sense of touch, the signals come from the somatosensory system composed of a variety of receptors (mechanorepectors, thermoreceptors, pain receptors, etc). In the case of vision, light is sensed through photoreceptors in the eye, and more specifically, in the retina. The information from each modality have varying benefits and constraints for the survival of an organism. While touching enables one to gain information on the heat, the texture and the size of an object, it requires one to be close to the object of study. On the other hand, vision for instance enables an organism to obtain information from afar, but is vulnerable to obstruction, when line of sight is lost. Though these modalities vary in terms of the physics of the observed phenomenon and the physiology of their sensors, they are combined to form complex representations of the world that are referred to as **percepts**.

The evolutionary origins of sensations.
From an evolutionary perspective, the emergence of cognitive capacities to sense the environment can be explain through random mutation and natural selection of the most adapted behaviours. Most complex living organisms have evolved to process information about their environment in order to find and extract energy from it, in order to grow and reproduce. But some species—i.e., the Trichoplax—do not need to make decisions to move and use random motion to get food (C. L. Smith et al., 2015). Thus, developing a sense of a self, in contrast with the environment, can positively help make a moving living organism take a decision on its next movement. In the study of consciousness and perception, this is an elementary step to guarantee the survival of species with complex behaviours. In fact, the neuro-physiologic mechanisms on which conscious cognitive tasks are based use many evolutionary ancient neuro-biological structures that can also be found in early points of vertebrate brain evolution (Mashour and Alkire, 2013). This suggests that the development of neural capacities to represent a self is at least intertwined to, or precedes, the development of perception. This view might differ from the traditional and popular narrative in which consciousness is often placed as higher in an often unjustified hierarchy of cognitive functions. As we shall see, we will argue that the emergence of perception and consciousness are interlinked and complementary. The capacity to determine a self and process one's internal state, in other words to be **self-conscious**, can be seen in the example of the *Dugesia Tigrina* worm. It shows simple behaviours: it can sense its internal state of hunger, and if it is not hungry, it rests while if it is, it starts sensing its environment in search of food (Sheiman et al., 2002). Hence, perception serves a purpose for dynamic organisms; it allows them to sense the world around them when in search of food, but also many other behaviours as their perceptual capacities grow.

From sensation to perception

Why do living organisms perceive? Fundamentally, perception can be considered as a means for active and dynamic organisms to obtain and process information of their environment and themselves, in order to make decisions. In the evolutionary context, perception can be thought as a plural and complex phenomenon that has emerged in most, if not all, living species, and facilitates the task of finding new energy sources for reproduction and survival. It is also linked to having a sense of *self* and thus consciousness (Feinberg and Mallatt, 2013; Mashour and Alkire, 2013). For instance, with the rise of vision as a modality, distant information can be captured, thus making it possible for beings to distinguish their organism from their environment. The multi-modal aspect of perception and its necessity to constrain the information of its inputs generates percepts. These percepts are representations of the world that can be held in the brain and manipulated. Once theses representations are held, the phenomenon of memory arises as percepts need to be remembered to be manipulated and exploited (Klein et al., 2002). This is linked to the notion of

object permanence introduced by Piaget and Cook (1952) as a marker of percepts being maintained beyond the sensory information flow and raises many questions regarding how the brain and perception operate this. Furthermore, developmental studies show that perception is actually a learned phenomenon. Indeed, human children, for instance, seem to fine tune their perceptual systems before showcasing further complex cognitive capacities (Tomasello, 2000; Gervain and Mehler, 2010; Ullman et al., 2012).

> **To synthesise, perception can be seen as a complex information processing phenomenon that enables a living organism to make decision based on its sensing of itself and its environment. As perceptual modalities diversified, in the evolution of species, their cognitive systems adapted and developed novel cognitive functions—e.g., memory, learning, language, consciousness, meta-cognition, etc—as perception became more complex.**

Perception for building a mental representation of the world

In cognitive sciences, the direction of the information flow is essential in order to attempt to break down and understand the processes. This leads to representations of the process and phenomenon with simplified pathways and general views such as **top-down** and **bottom-up** approaches to sensory signal processing. The implied hierarchical view of perception suggests that at the bottom of the process, sensory information is captured through the variety of physiological receptors present in a living organism. At the higher end, cognitive functions (i.e., memory, learning, consciousness, etc) aggregate information from lower processes to make decisions.

This paradigm is mostly inherited from the *cognitivist* approach to cognition (Fodor and Pylyshyn, 1988; Turing, 2009), where the brain operates in an essentially modular fashion. Hence information can be driven upwards, from sensory to perception, and further so, the living organism can understand its environment and interact with it. Philosophers have also proposed alternative paradigms that affect the very definition of perception and how it can be conceived, and modelled.

In the *connexionist* approach (Rumelhart and McClelland, 1986; Smolensky, 1988; Andler, 1990), cognitive functions are considered as emerging phenomena from highly inter-connected complex networks of neurons and synapses. Though this view keeps a notion of sensory input being driven by physiological receptors, but the processing of information is much less modular in the later stages. Thus remains a notion of direction of information flow, with the nuance that it breaks down once

it reaches the cortex. In this paradigm, cognitive function are not considered to be separated and independent, especially spatially, but are driven by the biological and information processing properties of neurons and their inter-connections.

Finally, a more recent approach has been proposed, the *embodied* or *enacted* cognition (Varela, 1996a; Glenberg and Kaschak, 2003; M. L. Anderson, 2003), in which cognitive functions are considered as emerging phenomena as well, but the distinction that removed sensory receptors from consideration is removed. Thus, the properties and constraints of the physiological sensors are considered in the modelling and understanding of the studied cognitive functions. In other words, cognitive functions are profoundly active, with abstract and multi-modal representation of the world being driven and moulded by physiological characteristics. Overall, these approaches have a common view of perception as an information processing phenomenon in living organism where sensory information is captured and processed into representations, which can be used to make behavioural decisions.

In other words, cognitivists, connexionists and embodied cognition fields do not differ on the fundamental reasons that justify the presence of perception, however they vastly disagree on its implementation and conceptualisation[1]. Biological systems tend to do so in an energy efficient manner, as opposed to artificial systems (Friston, 2010; Glasgow, 2018). This can be justified by the evolutionary origins of such phenomena. Indeed, an efficient approach for a system with high energy constraints is to extract information from sensory inputs and maintain it in its working space by doing **inferences** (Delorme and Flückiger, 2003). Perception is, as of today, widely acknowledged as an inference mechanism, and therefore, opens up many questions regarding the computational methods associated to it. In fact, perception is an interpretation of the captured information, a form of encoding in which information is abstracted and transformed to be used to orient an organism's purpose (Wandell, 1995).

> **The work presented in this thesis is fundamentally derived from an active cognition perspective, where cognition is embodied or enacted and highly interlinked to physiological sensors that capture the visual information and the perceptual phenomena. The focus, here, is on how ambiguity arises for the perceptual system, and how is it dealt with. When considering the inferential mechanisms that can fit the observed behaviours, ambiguity might emerge from insufficiently clear sensory information and when competitive interpretations are present.**

[1] Influential researchers from cognitive sciences are mapped in Fig. A.1 (Núñez et al., 2019).

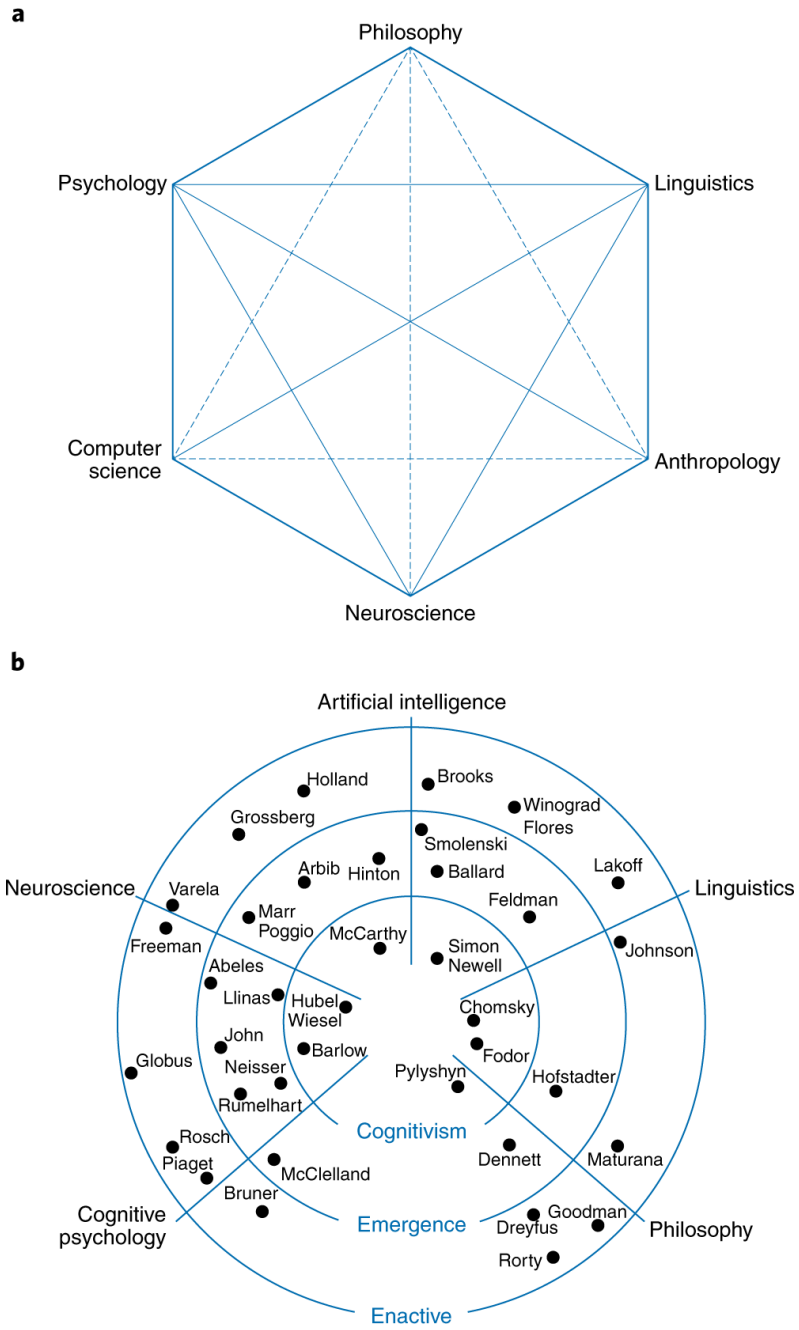


Figure A.1. **Map of cognitive sciences.**
a: an overview of the different research fields involved.
b: influential researchers according to research fields (angle) and philosophical positioning (scalar).
Figure taken from Núñez et al. (2019).

A.1.2 Perception as a cognitive process of information

Information flows in perception

Neuronal systems treat the information flow in the brain when perceiving. These biological systems are characterised by their diversity and multi-morphism (Eccles, 1965), but also by their plasticity (Conel, 1959; Fentress, 1999; Brown and P. M. Milner, 2003). An estimated 10^{12} neurons lay in the brain and they vary in terms of connectivity, size, shape, organisation, architectures, etc. In other words, there is a high diversity of types of neurons. Neurons are connected to other neurons through synapses which, depending on neurons, can vary between 1 000 and 10 000 synapses for a neuron. It is through these synapses that the information is transferred in the brain. Neurons react to the level of excitation at its dendrites, its inputs, and activate its own output through an electric discharge in its axon. The information is coded in *spike rates* and trains of incoming electric signals (Gabbiani and Koch, 1998).

Neurons and their synaptic connections form large complex networks that are able to produce many operations on the information flowing through them. As information is manipulated by the operations of neurons and their synaptic connections, the network applies a variety of filters, classifications, error correction, generalisation processes and more (Hopfield, 1982). The networks of neurons and their synaptic inter-connection lead to a variety of structures known to have different functional roles. For instance, in the olfactory system, the structure makes the sensory information converge and diverge, while also showcasing diverse inhibitory processes and a centralised control of the emerging percept (Hérault and Jutten, 1994). Another example, the visual system, shows a layer based structure with columnar organisation of the information flow, a parallel dispatch of the information content, and matrices of columns coding for the orientation of the visual information (D. H. Hubel and Wiesel, 1962; D. H. Hubel and Wiesel, 1968; Hérault, 2010). The difference of organisations and architectures of neural networks in the brain gives some of these networks specific functional roles, most notably near the sensory input. However, once the information is spread in the neo-cortex, the signals' processing is extremely vast, done in parallel and asynchronously. It is probably for that reason that complex behaviour and processes emerge.

Complexity in the brain.
The complexity, briefly described above, provides vertebrates' brains their capacities to act and interact with their environment. Complexity, here, is characterised by:

- the number of neural units and their connections,
- the large variability in the biological and chemical composition of these neurons,
- a highly complex metabolism,
- a system that operates at multiple scales (molecular, cellular, network and behavioural),
- and with a variety of architectures.

Other interesting properties lie in the robustness, energy efficiency and performances of the brain as a computing system. Indeed, vertebrates have functioning brains for long life spans, some loss of components does not necessarily degrade performances and mostly, brains show high plasticity and capacities to adapt and learn (Dayan et al., 2001). In an evolutionary context, these features have been useful and essential in the survival of species with such complex processing system. The highly diverse, parallel, and complex nature of mammal brains have made them capable to adapt and evolve rapidly to changes in their environment, giving them more chances to survive (Adami et al., 2000; Godfrey-Smith, 2002). Here again, nature offers different layers and methods for the evolution of behaviours—e.g., genetic changes, neural plasticity, social identity, etc. Perception and consciousness are phenomena that give organisms methods to treat information of their environment, and adapt these methods.

From perception to consciousness

Perceptual information can sometimes be experienced by human primates consciously. For instance, *you*, the reader, can be conscious of the visual information you experience in order to read this manuscript, when considering the text's visual forms (e.g., letters, words, spacing, etc) rather than its semantic meaning. This degree of consciousness is somewhat vague and troublesome to define as, for instance, experienced and efficient readers run through words with no need to consciously process letters individually (Rumelhart and McClelland, 1982). This is not the case for children learning to read who represent an extreme example of readers with little experience in the task (Casco et al., 1998; Blythe et al., 2009). The phenomenon of consciousness has remained a mystery to most researchers and intellectuals as far as written history goes (Van Gulick, 2018). A marking turn in philosophy has been the work of René Descartes, which led to the view of *dualism* (Descartes et al., 1970). The idea, that conscious thoughts and objects live in a world separated from the physical world, has since been discarded by *materialist* and scientific approaches (Dennett, 1993; Damasio, 2006). The range of solutions

are vast and the phenomenon is hard to study as it is deeply linked to introspection and methodological constraints (Dennett, 1993; Varela, 1996b; Chalmers, 2007), and many thinkers have proposed creative and wide paradigms—e.g., the absence of consciousness in Behaviourism, the presence of consciousness in all things in Pan-Psychism, consciousness as the interpretation of quantum gravity derived from Gödel's uncertainty theorem (Penrose and Mermin, 1990) or consciousness as a cultural construct (Jaynes, 2000).

> **The link between perception and consciousness is important in this thesis' work, as the phenomenon studied historically has relied on conscious subjective reports of dynamics experienced by a subject. Therefore, the methodological problems addressed in consciousness research are related to those faced in conscious perceptual dynamics studies.**

Of the philosophical frameworks presented in this section, one is particularly capable to provide researchers with solutions to these methodological issues: embodied cognition. It conceives cognition and perception as deeply dependant and intertwined with the agent's body. Though its premises might be found in the study of metaphors in languages (Glenberg and Kaschak, 2003; Lakoff and Johnson, 2008), embodied cognition can be defined more generally. The framework mostly considers the following notions.

- Constraints from the agent's body, as the sensory inputs to perception and cognition are parts of the body, their nature and features will mould how cognitive processes develop.
- Distribution of the agent's cognitive processes in the body: it implies that not all cognitive computations are done in the central neural system, the brain, but also some is done in other organs—e.g., the heart (Park et al., 2014).
- Regulation of the agent's cognitive processes by the body's state: meaning that action and energy from the body is needed to ensure cognitive processes.

To illustrate these general aspects, in the case of vision, the information that enters the eyes can be greatly affected by the dynamics of the eyes themselves. In Appendix A.2, it will also become evident to the reader, as we review our current understanding of human vision, that given the physiological properties and capacities of the sensors, the perceptual processes were likely developed to adapt to those. For instance, visual percepts may fade out if their image on the retina is stabilised for a period until the eyes move again (Riggs et al., 1953; Yarbus, 1967; Blakemore et al., 1971; Cohen et al., 1977; Martinez-Conde, Macknik, and D. Hubel, 2004).

Hence, it is in the context of the embodied and enacted cognition paradigm that the works presented in this thesis should be interpreted—though not necessarily incompatible with the other paradigms. It allows the consideration of the body and action as parts of the cognitive processes, and more specifically, in our case, we may look at the links between eye movements and visual perception. However, how is visual consciousness or awareness approached for research? In order to address this type of question, it is necessary to define conscious information, and to review what is considered as visual perception. This step is needed as the first relates to subjective experience and introspection, while the scientific method aims to model, predict and measure phenomena as objectively as possible.

The distinction has mostly been drawn through contrasting methods developed over the decades in psychology and neurosciences. For instance, differences between a subjective experienced of a perception being reported consciously versus unconscious perceptions (Baars, 1993). This approach must be reproducible even though it may have variability from one trial to another, or from one individual to another. One approach is to work on the masking of a stimulus so that it can be either be perceived consciously or not in a task, and see how behavioural performance or neural correlates are affected (Dehaene and Naccache, 2001; Dehaene, Naccache, et al., 2001). Another approach is to study phenomena where the conscious state of perception changes, or alternatively, the lack of conscious changes while the physical stimulus varies (D. A. Leopold, 1997; Mamassian, 2006; Chopin, 2012).

In this section, we have looked at how cognitive behaviours from increasingly complex organisms have led to the emergence of perception and consciousness. In the next section (Appendix A.2), we shall focus and present the visual modality of perception in humans, from photons to percepts.

A.2 From the eyes to the brain

This section offers an introduction to visual perception. It develops into how the brain, with the eyes, reconstructs a rich representation of the world, starting at the physics of light to the neural pathways that feed the visual cortex.

Terminology.
The terminology of vision provides a structure to phenomenon, based on the information flow and its processing. We shall refer to **low-level** vision, as is common in the literature, for the early stages of visual processing (e.g., the eyes, the retina, photoreceptors, lateral geniculate nucleus, etc). And **high-level** vision will refer to the perceptual experience and the processes associated to it in the brain.

The physics of vision

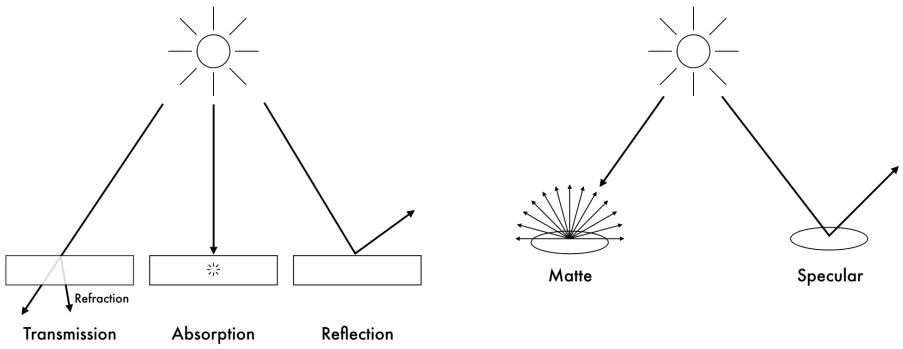
Vision relates to the act of sensing the light of one's environment. The field in physics that studies light is known as optics and has provided a strong understanding of the phenomenon over centuries of scientific research. It is known that light is composed of **photons** which act as *quantas*—i.e., minimal and granular units of energy—and behave like waves and particles. Photons are mostly considered to have particle behaviours in vision science (except for colour science) and will be considered as such in this work. In physics, photons are known to have their source in hot bodies (e.g., the sun, stars, fire, etc) and to radiate away from their origin. As small particles, they travel through the air in quasi-straight lines at high speeds (*more on photon displacement in the box below*). However, photons rarely hit the eyes directly: they are usually reflected on many parts of the environment, which correspond to the illumination of objects. In vision science, we refer to **luminance** for the physical quantity that corresponds to the amount of visible light, i.e., the number of photons, falling on a surface over time (S. Palmer, 1999) and it is expressed in candela per squared meters—abbreviated to $cd.m^{-2}$— (BIPM, 2006). It may be linked to a more familiar notion, *brightness*, though it is not done so in a straightforward manner.

Photons.
In vacuum, that speed has been established as the constant c of 299 792 458 meters per seconds (m.s^{-1}) and corresponds to a single Planck unit. The Planck units were proposed by Max Planck as the natural constants of physical cosmology because they were defined solely by properties of free spaces and are not of any chosen object or particle. They set the granular limits of the standard model of physics. The other four constants are the gravitational constant (G), the reduced Planck constant ($\hbar$), the Coulomb constant (k_e) and the Boltzmann constant (k_b) (Penrose and Jorgensen, 2006). However, in the air, the speed of light is affected by the *refractive index* defined as $n = c/v$, where v is the velocity of light in the air. Thus, given the refractive index of air ($n = 1.0003$), the speed of light in air is approximately 299 702 547.2358 m.s^{-1}.

Light has three main types of interaction with a surface (see Fig. A.2b):

1. transmission, in which the photons go through (and can be refracted),
2. absorption, in which the photons are trapped and,
3. reflection in which photons bounce back in different directions.

Reflection implies that light interacts with a surface and the direction and angle of photon propagation changes. For instance, in visual perception, the properties of reflection help infer on texture properties; a *matte* surface will be associated to a different type of surface than a *specular* (see Fig. A.2b). This is due to how the light is reflected—in a diffused fashion for the matte surface, a coherent one for the specular. Surfaces are considered as secondary light sources as they imply an interaction with the original light that can be perceived to infer an object's properties.



(a) Transmission, absorption & reflection. **(b)** Matte versus specular surfaces.

Figure A.2. Light interactions.
(a) Diagram showing the three types of interactions between light and an object.
(b) Diagram showing differences in light diffusion when interacting with a matte and a specular surface.

It is also worth noting that the photons captured can be considered as the ones converging towards the eyes. The converging flow of photons represents the available visual information for an observer and is named the ambient optic array (AOA) (Gibson, 2014). The AOA means that vision is possible because at a point of observation, the light converges and thus, environmental objects can be sensed, and given the properties of photons, the light provides, in most cases, a straight forward image of the environment—i.e., photons move directly to that point and have consistent behaviours. When adding a time component, we obtain a dynamic AOA which correspond to the optic flow of photons arriving at a point of observation. Thus, at such a point of observation, an image of the environment, based on the array, will provide visual information for a sensing entity, like a human being. There are two important considerations that are derived from this notion:

1. the observer has a *point of view* because he has an incomplete access to the visual information present in the environment and,
2. vision is based on an *image* of the environment at the sensors' position, namely the eyes' retinas.

The image is often named the **retinal image** and contains the visual information that will be treated by the visual system. That image is merely a projection of the optical flow available on a surface, the retina. Thus, the environmental objects perceived are considered as **stimuli** that can either be *distal*[2] or *proximal*[3]. The notions presented here mean that, in vision science, we will consider and specify the size of a stimulus—e.g., an object—as a **visual angle**. Visual angles are computed using the following equation and are measured in visual degrees (deg) or degrees of arc (Kaiser, 2004).

$$V = 2 \arctan\left(\frac{S}{2D}\right) \tag{A.1}$$

where V is the visual angle, S the frontal extent of the object in the environment and D the distance between the eyes and the object (see Fig. A.3). This corresponds to the size of the object as a proximal stimulus at the point of observation, given an AOA; in other words, the size on the retinal image for an observer.

> **The retinal image is two dimensional though we experience the world as three dimensional. Perception therefore goes beyond sensing; it is also an inverse problem. The brain reconstructs the observer's environment based on observations, but also what it has**

[2]Distant from the observer.
[3]The optical image on the retina, at the back of the eye.

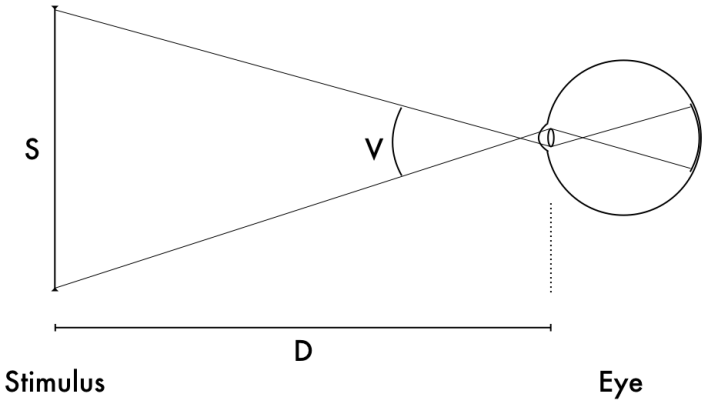


Figure A.3. Retinal image projection. Diagram showing the geometry of retinal image projection and the computation of visual angle.

learned in the past. This is where one of the key characteristics of visual perception intervenes; it is an inferential process.

The eye, a biological light sensor

Our understanding of what the eyes are composed of and how they work has evolved since Plato's emanation theory[4]. But the contemporary understanding of the eyes is founded in Alhazen's idea that conceived the eye as a pinhole camera (S. Palmer, 1999). As physicists progressed in their understanding of lenses, the derived physics of light impacted the understanding of how the eyes might work. Johannes Kepler combined these understandings to propose the first modern theory of physiological optics.

Anatomically, the eyes are composed of different biological tissues that allow light to be directed through the aqueous and vitreous humour to the retina (see Fig. A.4). This is where the retinal image is captured with **photoreceptors** – cells that can transform photonic information into neural information – and the visual information is sent to the visual cortex via the optic nerve. Photons enter the **pupil**, whose diameter is controlled by the ciliary muscles, in the iris through the pupil's cornea

[4] A view in which eyes were believed to have an "inner fire" propagating rays towards the perceived objects.

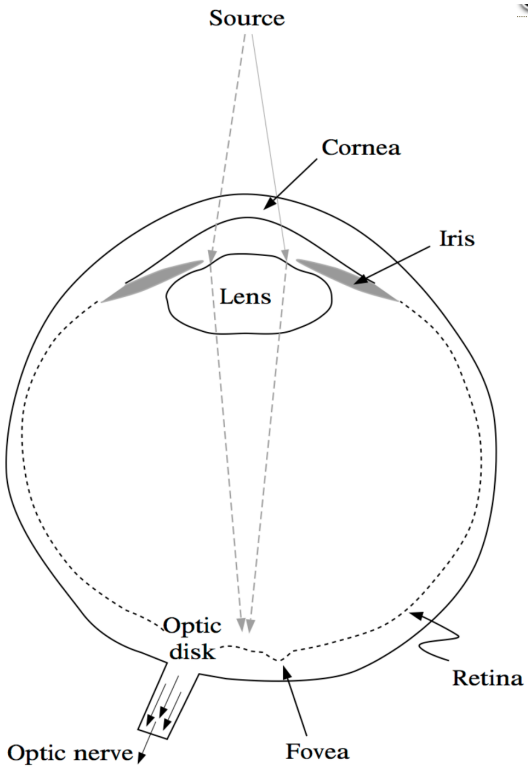


Figure A.4. **The eye.** Diagram of an eye and its structure: the cornea, the iris, the lens, the retina, the fovea, the optic disk and the optic nerve. Figure taken from Wandell (1995).

and through a lens that focuses the optic flow on the **fovea**. The fovea is the central region of the retina where the density of photoreceptors.

The retina. The retina is the region of interest in the eye as it is where the physical light is projected and captured by a dense heterogeneous lattice of photoreceptors[5]. The latter convert the light into neural activity, i.e., visual information, that is sent to the visual cortex and beyond. The retina's exposure to light is controlled by the iris and the pupil as physiological aperture controllers. In other words, if high quantities of light shine on to the retina, the pupil will constrict, and *vice versa*, it will dilate as luminance decreases. However, pupil dilation also depends on internal psychological factors and is known to be linked to emotional arousal (Hess and Polt, 1960) or concentrations (Hess and Polt, 1964). Though the pupil acts as an aperture hole, the cornea focuses the retinal image so that it is sharp[6]. These components of the

[5] Over 100 million light-sensitive photoreceptors cover the retina.

[6] The main ophthalmic conditions, e.g., myopia, hyperopia or presbyopia, are due to the focal point of the cornea being placed in front or behind the retina. Note that the cornea is elastic and can adjust the focal point's position depending on whether the observer focuses on distant or close-by objects.

eye manage the physiological optics, the front-end of the visual system. Once the photons hit the retina, the information is transformed into neural signals.

From photons to neurons. The photons interact with cells at the back of the eye called photoreceptors, composed of two classes: **rods** and **cones**. The former are more numerous (about 120 millions) and are highly sensitive[7] to light and located all across the retina, except at the centre. They are the peripheral photoreceptors and enable the detection of visual information in low luminance levels (also known as scotopic conditions), e.g., in the dark or at night. The latter, the cones, are less present in the retina (about 8 millions) and have lower sensitivity. They are densely positioned at the centre of the retina, in the fovea, with exponentially decreasing concentrations as shown by Fig. A.5 (Wandell, 1995). Cones are the receptors that capture most of our visual experience as they function at medium and high luminance levels (also known as photopic conditions). The fovea, where cones are concentrated, covers a small part of the visual field of approximately 2 degrees, the size of one's thumbnail at arm's length. However, this is where the visual system captures information the most precisely and in colour—i.e., adding more dimensions to light perceived referred to as *chrominance*.

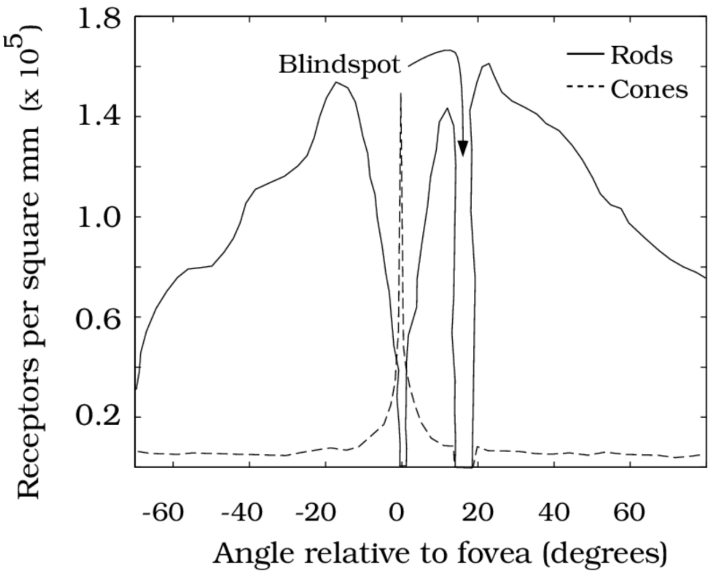


Figure A.5. Cones and rods. Density of rods and cones across the human retina. Cones are highly concentrated in the fovea at 0 degrees while rods are absent. On the other hand, rods dominate in the periphery. Note the blind spot which has no photoreceptors. Figures taken from Wandell (1995).

[7] Sensitivity means that the reactivity of the cell is higher. For instance, a rod needs less luminance, less photons, and less time to activate.

But, how do these photoreceptors transform a physical quantity of light, such as photons, into a neural signals? With a process called **transduction** (Corbett and J. Chen, 2018). Rods and cones have a similar structure, with an inner segment containing the nucleus and other cellular components, and the outer segment where billions of light-sensitive pigment molecules reside. These molecules are called *rhodopsin* in rods and are embedded in the membranes of the outer segment. They are responsible for transforming optical information, light, into electro-chemical energy that can be transferred in neural networks. They absorb photons as these strike them, thus changing the electric current flowing in and around the pigment molecule. The outcome is the production of an electrical change in the outer membrane of the receptor, which gets propagated to the synaptic part of the photoreceptor. From there on, chemical transmitters relay the charge to the next neuron.

Neural basics.
As the gateway of neural networks for visual perception, the retina is linked to **neurons**, the elementary unit of the brain and cognition. Neurons are composed of three parts (see Fig. A.6) that play different roles in its main function: the transmission of neural activity, an electric charge. The neuron is often presented in the direction of information flow, and its input are the **dendrites**. There are usually many dendrites deployed for a single neuron and they are sensitive to a graded potential, which correspond to the electrical difference between the inside and the outside of the dendrite. The central part, the **cell body** has the nucleus and integrates the potentials across the dendrites. If the graded potentials surpass a threshold, the body will execute an all-or-nothing reaction, thus activating the **action potential**—it may also be referred to as the nerve impulses or spikes. This has for consequence to activate the final part of the neuron, the **axon**, the thin, long and myelin sheathed—the myelin sheath speeds up the conduction of the action potential throughout the axon and is composed of Schwann cells interlinked by nodes of Ranvier—part of the cell. Information wise, the strength of the integrated signal at the dendrites and body is encoded in the frequency of axonal activity, i.e., the firing rate, which corresponds to the number of spikes per seconds. Finally, at the end of the axon, the electrical signal is transformed into a chemical signal by *neurotransmitters* situated on the terminals. These are connected to other neurons, thus propagating the neural signal on-wards. This region in between the terminal of a neuron and the dendrites of the next neuron, where the neurotransmitters flow, is called the **synapse**.

In the eyes, the photoreceptors act as the front-end of the neural visual network. Once the visual information is converted into neural activity, it is passed on by applying a first set of information processing in retinal neurons—e.g., horizontal, bipolar, macrine and ganglion cells. The information is integrated, encoded, concentrated and sent in a highway to the visual cortex through the optic nerve. Interestingly, the optic nerve is so dense with axons of ganglion cells that no photoreceptors are present at its location, thus generating an area, almost as large as the fovea, with no

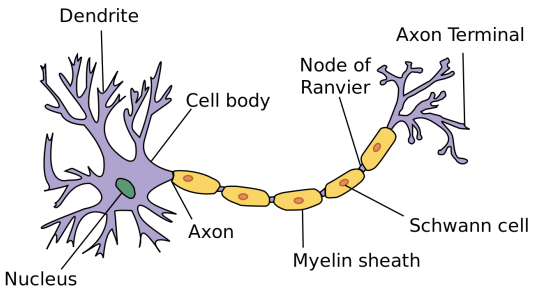


Figure A.6. Neuron structure. Diagram representing the structures and components of a neuron. The information flows from left to right, starting in the dendrites, with graded potentials being integrated and converted into an action potential in the cell body where the nucleus lies. The spikes are then transferred through the axon to the terminals where they can be outputted to other neurons using neurotransmitters. Taken from the internet (wikimedia).

Figure A.7. Blind spot illusion. The blind spot can be found by using the stimulus presented in this figure. Close one eye and fixate the cross with your eyes and move gradually the support on which you are reading this thesis towards or outwards until the dot disappears. When the dot disappears, hold it! This is the distance at which the retinal image of the stimulus is such that all the visual information of the dot is inside the blind spot and cannot be inferred, thus it disappears.

capacities for light detection. This area is called the optic disk or blind spot, and it is compensated by perception and visual processes. It can however be unmasked by doing the little experiment shown and presented in Fig. A.7.

The retinal design is actually full of counter intuitive properties. For instance, the photoreceptors are truly at the back of the eye, and the outer segments and the following cells are placed in front of them, in the direction of the photons' arrival. However, the retinal cells are relatively transparent which allow photons to be still detected. Another aspect is the presence of numerous dark blood vessels irrigating the eye's cells. They are however invisible in perceptual experience. This is explained by the fact that as light enters the eye, the shadow of such vessels is deported over different angles, thus the way more numerous photoreceptors still get light, more

so when the eyes move. Finally, an important counter intuitive properties of the lower visual system is in the dynamics of neural response to light stimulation of photoreceptors. One may expect the latter to increase synaptic output as more light is presented, but the opposite was shown for the receptors of vertebrates[8]: the flashing of a light on the eyes of Necturus maculosus and Gekkos showed a decrease in synaptic activity response (Toyoda et al., 1969). This dynamic is however recorded in a positive fashion at the next layer of synapses (S. Palmer, 1999).

Visual information pathways to the brain. Once the information has left the eye, through the axons of the ganglion cells in the optic fibre, it is crossed in the optic chiasm. The content on the right eye goes to the left hemisphere of the brain, and the left eye content goes the right one. Part of the information goes first through the SC, a nucleus in the brain stem, while the rest goes through the LGN of the thalamus. Following the LGN, the information reaches the occipital cortex (or primary visual cortex), a brain area at the back of the skull, from where it spreads further after to other cortical areas. The information that goes through the SC is smaller in quantity and most of the information flows towards the visual cortex.

> **Before moving to describe the ways the brain handle visual information and perception, it is necessary to review and look at how the visual sensors, the eyes, interact with the environment (in Section 1.1.2). Indeed, so far, we have described the motion of information through the low-levels of vision, but that the eyes are not static and passive components simply treating the visual flow. They are active and can move to interact with their environment, hence affecting and changing the retinal flow, and thus the visual content processed.**

A.2.1 Visual information in the cortex

The human cortex refers to the convoluted and folded biological tissue in the brain. It is composed of high concentrations of neurons, often organised in layers, and structured in two hemispheres that are approximately symmetrical and split along the neck to nose axis. Research on damaged brains[9] has shown that, though the brain is a set of highly parallel neural networks, some areas of the cortex are systematically used for some cognitive functions. For instance, the occipital lobe was discovered to

[8] More puzzling, the intuitive relationship actually occurs in invertebrate eyes.
[9] Often due to strokes, head injuries, or brain tissue removal.

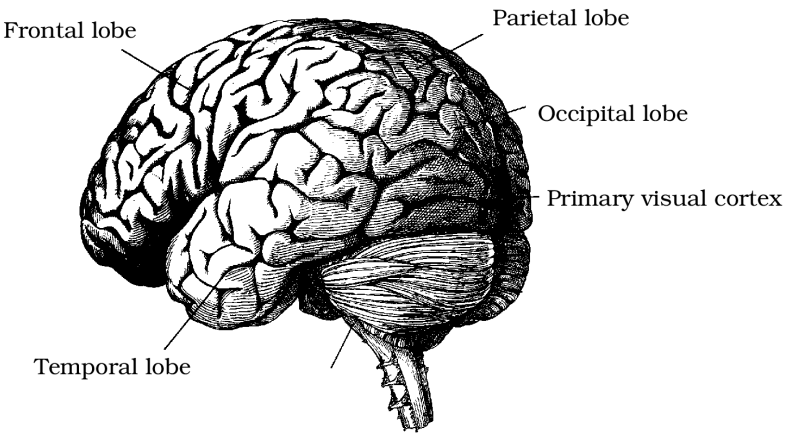


Figure A.8. Brain structure. Drawing of a brain showing the convolutions and folds of the cortical tissue. The lobes are localised. Image taken from Wandell (1995).

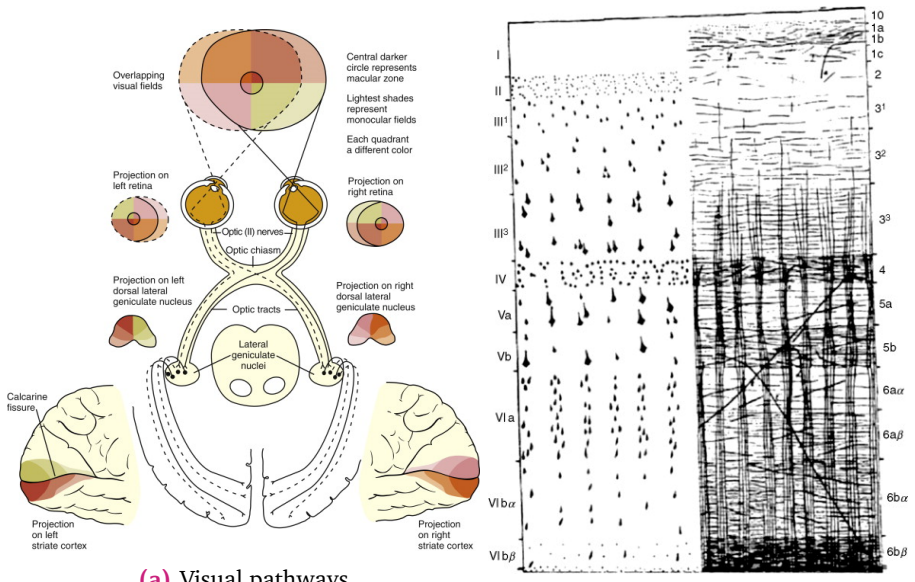
be essential for visual decoding, as it receives the visual information from the LGN for initial processing before transferring it to other parts of the cortex (Glickstein, 1988). In fact, the other sensory modalities have similar dedicated area—e.g., temporal for audition, parietal for motor control; see Fig. A.8 for a visual representation of the brain.

Primary visual cortex

The **occipital cortex** has been extensively studied with a lot of data on mammals, and more specifically macaque rhesus monkeys which have very close visual performances to humans (D. H. Hubel and Wiesel, 1959; D. H. Hubel and Wiesel, 1962; D. H. Hubel and Wiesel, 1968; R. Tootell et al., 1988; Roger B Tootell et al., 1988). Invasive studies on animals have allowed researchers to probe on cells as the information flows during tasks on which the animals are trained. Hence, insights on the functionality, combined with anatomy studies, are used to improve the literature's models of the human visual cortex. We know now that once the information has left the LGN, it reaches the *striate cortex*, or **primary visual cortex**, or even called **V1**, in which the information is crossed[10] as shown in Fig. A.11b. The information is therefore divided across the two hemispheres, but is also connected by the *corpus calosum*, a large fibre tract that permits information transfer from one hemisphere to the other. Moreover, the visual information is mapped topographically in the striate cortex, with respect to the retinal image, an aspect referred to as **retinotopic**

[10]The visual information from the left visual field is projected on the right striate cortex, and *vice versa*.

projection (*more details in the box below*). The visual cortex can then be separated into functional areas addressing separate areas of the visual fields.



(a) Visual pathways **(b)** Cortical layers

Figure A.9. Neurosciences of vision.
(a) Diagram of the visual pathways from the retina to the striate cortex, going through the optic nerve and the LGN. Coloured patches show how the image is manipulated to obtain retinotopic projection. Taken from Remington (2012). (b) Vogt's scheme of the fundamental plan of cellular (left) and fibre (right) layering in the cerebral cortex. The cellular layers (I through VI) are those given in the text and Vogt divides them into sub-layers. Taken from E. Jones (2004).

Retinotopic projection.
The projection is relatively distorted though it keeps a higher sensitivity for the central foveal area over the periphery. This is called the cortical magnification factor and is coherent with the much higher quantities of conic photoreceptors at the centre of the retinal image. However, once again, perception is not distorted, meaning that the brain operates computations to build a coherent and stable representation.

Receptive fields

Receptive fields in V1 correspond to areas in the cortex that treat visual information from an area of the visual field systematically. We know this since the pioneering work from D. H. Hubel and Wiesel (1959) in which the authors showed that a cat's occipital neurons responded specifically to the orientations of a bar stimulus, thus playing a role in detecting contour and orientation in the visual field. The receptive fields, each composed of multiple neurons, are connected to each other with some

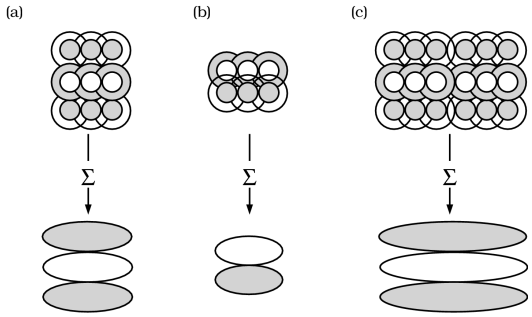


Figure A.10. Receptive fields. Diagram showing orientation selectivity of receptive fields in the primary visual cortex. Each set of central and peripheric circle corresponds to a neuron that react to an orientation. Each receptive field has an adjacent excitatory and inhibitory region. Their outputs are then combined by a summation mechanism and allow the extraction of orientation over a larger area of the retinotopic projection. Figure taken from Wandell (1995).

level of overlap, thus allowing the detection of contours and of object by using Gabor filter models for signal processing of early cells in V1 (J. P. Jones and L. A. Palmer, 1987b; J. P. Jones and L. A. Palmer, 1987a; Szulborski and L. A. Palmer, 1990; Nishimoto et al., 2006) and by combining receptive fields horizontally, and processing further up the information through a layered architecture vertically (Wandell, 1995; Breuil, 2018).

The layers are organised as in most parts of the cortex, with sensory information coming from the magnocellular and parvocellular pathways—two pathways that propagate the visual information at different speeds and spatial frequencies—on layer 6 (see **??**) towards the inner layer 1 and further cortical areas (Hendry and Yoshioka, 1994; Wandell, 1995). The visual signals are therefore carried through this series of layers, with a variety of selection based on spatial properties and decoded features. Since there is a retinotopic projection in V1, it acts as the basis before the information is dispatched across the rest of the brain to interact with other neural networks and cognitive functions—for instance, the content of visual awareness can be decoded using neuro-imaging techniques focused on V1 (Haynes and Rees, 2005; Parkkonen et al., 2008). One should note that, at this stage, the visual information has already been impacted by feedback mechanisms (in the LGN or SC for example) and will be in the following cortical areas where it spreads. The complexity of understanding human vision arises from these multiple loops and interactions with memory and predictions.

A.2.2 From visual perception to consciousness

The visual pathways

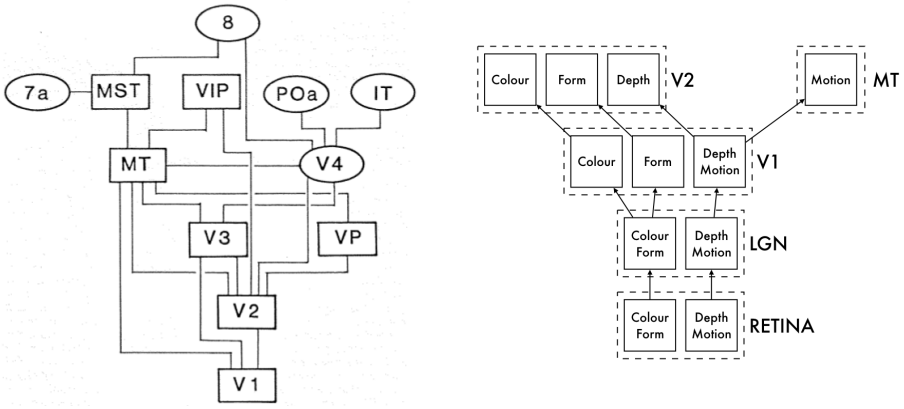
Once the visual information has been processed by the primary visual cortex in the occipital lobe, the information is disseminated in other cortical areas towards the temporal, parietal and frontal lobes. The visual system is thought to be composed of further areas (V2, V3, V4 and V5) spreading towards the temporal and parietal lobes (Remington, 2012).

The visual stream is then divided in two into the **ventral** and **dorsal** pathways (Mishkin et al., 1983). The ventral pathway, often described as the *what* system is mostly involved in perceptual object identification and involves areas that reach the temporal lobe. Meanwhile, the dorsal pathway is also considered as the *where* system, involving areas situated towards the parietal cortex, and that contribute to locating perceptual object.

However, the reader should be aware that interactions over each pathway remain, and that the information is not processed in a straightforward fashion—see the diagrams Fig. A.11 for a simplified model of cortical areas involved in vision. This was shown by the works of Melvyn A Goodale and A David Milner (1992), in which the visual information is considered in the action-perception dichotomy (A. Milner and M. Goodale, 1995; A David Milner and Melvyn A Goodale, 2008). Hence, the visual system is highly interlinked with interactions with most specialised cortical areas as has been shown by patients with *visual agnosia*[11] (S. Palmer, 1999). Another body of evidence that supports this view of the visual system comes from *unilateral neglect*, a syndrome in which patients are hemi-negligent when one side of their parietal lobe is damaged. In other words, they are unable to localise or interact with objects present on one side of their visual field (Sacks, 1985).

> **Overall, though neuro-imaging techniques such fMRI, EEG, MEG, positron emission tomography (PET) or electro-physiology exist, data from brain lesion often provide stronger conclusions with some level of causality[12]. These curious phenomena suggest that the brain handles many visual decoding operations unconsciously, and that visual awareness or consciousness is not systematic.**

[11] When a patient has a visual deficit for a specific task such as face recognition or reading, for instance.
[12] Neuro-imaging techniques mostly provide correlative evidence. But the TMS methods provide non-invasive techniques that permit artificial over-activation or inhibition of specified cortical areas at the surface of the brain.



(a) Diagram of the visual areas

(b) Diagram of the visual pathways

Figure A.11. Visual system.
(a) Diagram of the visual areas in the macaque. Each area provides specific functional properties in the processing of visual information. The pathways between areas represent known reciprocal interactions. Figure taken from Van Essen and Maunsell (1983).
(b) Diagram of the visual pathways hypothesis, showing how and where properties of the visual information are encoded into percepts from retina to the cortex. Reproduced from S. Palmer (1999).

Visual consciousness in the fronto-parietal networks

After tracing back the discourse of information in the visual system, from photons to the brain, going through the retina, the LGN and the visual cortex, we are confronted again with questions and problems linked to consciousness, as approached already at the end of Appendix A.1. Consciousness is considered through multiple definitions across the literature, with criteria that vary across authors. The notion of having a *self* and its capacity to do meta-cognition on perception for instance, is a key aspect. More precisely, the capacity for a thinking organism to have second-order and beyond thoughts, and from there, to be able to construct a narrative that explains perception and actions, is used to define consciousness. Another criterion is the necessity for conscious information to flow through multiple feedback connections also referred to as *recurrent processing* (Sporns et al., 1991; Lamme, 2006). Finally some authors refer to consciousness as a rich personal experience that is the qualitative particularities of sensory experience also known as *qualia* (Chalmers, 2007).

Recurrent processing.
Recurrent, or feedback, processing relates to feedback mechanisms in which the processed information cycles back through a core network and is updated by the various spread functions of the brain (Dehaene and Naccache, 2001). This differs from *feedforward* processes such as the ones described in the visual system so far, where the information is assumed to go in a unique direction, from the bottom to the top. Recurrent processing is thought to be necessary for conscious experience as it engenders conditions that satisfy Hebb's rule (Donald Olding Hebb and D. Hebb, 1949) of pre-synaptic and post-synaptic simultaneous activation leading to plasticity processes activation and providing the neural basis of leaning (Lamme, 2006).

In the case of visual perception, we will consider consciousness as visual awareness (D. A. Leopold, 1997). In other words, when a perceptual object can be subjectively reported on by an observer; this corresponds to the behaviour and introspective approach, and has helped develop the *masking* paradigm (Dehaene, Naccache, et al., 2001; Breitmeyer et al., 2006). These experimental protocols make it possible to estimate the degree of consciousness at which a stimulus is processed by inserting it in a series of frames and masking it, so that the time of exposition can be estimated to process it consciously. Thus, a contrast between consciously perceived and unconsciously perceived stimuli can be made. When combined with neuro-imaging techniques, neural correlates of consciousness can be identified. But the search for NCC has led scientists to take distances from introspection, and to base the detection of consciousness based on the detection recurrent processing network activation (Lamme, 2006).

Using neuro-imaging techniques, a body of works have contributed to show that a **fronto-parietal** network is activated when visual awareness occurs (Williams et al., 2003; Windmann et al., 2006; Rees, 2007; Sterzer et al., 2009; Kanai et al., 2010; Frässle et al., 2014; Tsuchiya et al., 2015; Michel and Morales, 2019). Such results were obtained using a key paradigm for visual awareness changes, **binocular rivalry**, in which conscious perception changes over time though the two stimuli exposed to each eye separately remain constant. However, the fronto-parietal network is not the only active area; the visual cortex is necessarily active during visual decoding (D. Leopold and Logothetis, 1999; Brouwer and Ee, 2007; Haynes and Rees, 2005; Parkkonen et al., 2008; Van Vugt et al., 2018).

Overall, the literature proposes that once the visual information has been encoded through the visual system, frontal cortical areas, involved in attention, and parietal cortical areas, involved in decision making, play a key role in the emergence of a new conscious representation for a perceptual object. Finally, EEG and MEG studies have shown long-distance cortico-cortical synchronisation of Beta and Gamma band oscillations, as well as the large scale activation in the fronto-parietal network (Strüber

and C. S. Herrmann, 2002; VanRullen and Koch, 2003; VanRullen, Busch, et al., 2011; Dubois and VanRullen, 2011; Dehaene and Changeux, 2011).

Though there is still a lot to uncover around NCC, researchers have started addressing methodological issues to orient data collection towards more objective methods. Visual consciousness shows particularly interesting features such as perceptual reversals when visual information is *ambiguous* but constant. Such changes are explained by one key property of human's visual system: representations are constructed by applying inference mechanisms.

A.3 Tracking the eyes

Eye tracking signals

Eye movement signals are multi-variate, which means that for a time series, there multiple dimensions associated to each sample. At the time of writing, the most studied signal is the bi-variate gaze, composed of the estimated foveal position on the screen in the horizontal and vertical spatial dimensions. Some video-based eye trackers now provide the gaze signal for each eye, when the binocular recording option is active. Moreover, the pupil diameter can be recorded, thus giving information on the state of pupil dilation and making it possible to combine gaze analysis with pupil analysis. This may also be applied to both eyes in modern high-end video eye trackers. Pupil size remains a difficult signal to use in cognitive studies as its dynamics is highly dependent on luminance; it is adjusted to reduce or increase the amount of light that reaches the retina. However studies have started characterising and contrasting the signatures in pupil dynamics of luminance adjustments and cognitive tasks (Knapen et al., 2016; White and R. M. French, 2017). Pupil dynamics are highly affected by blinks and eye movements, and their processing is challenging since these events become artefacts that degrade dramatically the signal over epochs (Hupé, Lamirel, et al., 2009).

Eye tracking signals are subject to artefacts, notably, eye blinks during which the eye lids cover completely or partially the eye. These events are usually detected by eye trackers as they have signatures that identify them accurately. However, blinks generate gaps in the temporal series, during which the visual information flow is physically interrupted, and so is the gaze position measurement, but visual experience is not necessarily. Our visual experience seems continuous most of the time though blinks are carried out. Blinks are a relatively less studied oculomotor phenomenon; they are linked to fatigue as their rates increase over time in a task (Stern et al., 1994) and they are also known to be dependent on cognitive tasks (Bentivoglio et al., 1997). Finally, the signals used to study eye movements are dependent on the technology and apparatus used.

Apparatus and technology

Eye movements can be measured with different techniques and the history of their development has affected how research has been carried out. The first systems were developed at the beginning of the 19th century, and since the methods developed

at the time were invasive, they were used mostly in medical contexts (Dodge and Cline, 1901; Huey, 1908). The systems used contact lenses on the cornea connected to an apparatus where light was projected and thus movement was detected, while other systems used electro-oculography (Heywood and Churcher, 1971; Heywood and Churcher, 1972).

Methodological apparatus evolved gradually but kept cumbersome and invasive constraints up until the 1970s (Yarbus, 1967; Cornsweet and H. D. Crane, 1973; Mele and Federici, 2012). As technology evolved towards digital instruments and with the advent of better cameras and computers in the 1980s, video-based eye trackers became more disseminated and accessible for research, and their accuracy and usability increased.

Eye tracker.
The systems used in this thesis are provided by *SR Research* and belong to the EyeLink category of eye trackers, more precisely, the *EyeLink1000* and the *EyeLink1000+*. They measure gaze and pupilometry signals by projecting infrared light and detecting its reflection on the cornea. The *EyeLink1000+* can sample the data at 2000 Hz when the head is stabilised on a chin-rest apparatus, with an accuracy between 0.25 deg and 0.5 deg and a resolution of 0.01 deg, in binocular recording.

The signals are combined with eye models and calibration data to estimate the position of gaze and the size of the pupils (Abramov and Harris, 1984). As instruments have become more off-the-shelf and commercial, the data has gained in standardisation. However, differences and imprecision persist at very small amplitude and has been shown to be related to the artificial eye models' noise structure (Coey et al., 2012; Wang et al., 2017). Researchers have also started adding their hardware and software proposals for digital signal processing systems that can provide fast and accurate real-time gaze position in order to implement gaze contingency[13] at intervals of 10 ms (Santini et al., 2007).

[13] An experimental paradigm in which the retinal image position is control by feeding back the position of the gaze to the experimental display computer in order to correct the stimulus.

A.4 Multi-stable perception detailed description

Properties, common to all multi-stable phenomena, are described in the following paragraphs.

A.4.1 Irrepressible

Multi-stable perception can, up to a certain extent, feel like it is voluntarily controlled. However results on continuous free viewing of stimuli show reversal always eventually occurs, giving the phenomenon its irrepressible property. Nevertheless, it is possible to stabilise such ambiguous perceptions by adjusting stimuli presentation time and ISI (D. Leopold, Wilke, et al., 2002) or by biasing the stimulus. Beyond given durations of presentation and ISI, which vary given the multi-stable stimuli used and the observer, reversals will be experienced; thus stimulus presentation does not need to be continuous, it only needs to be beyond a certain duration threshold. This characteristic is valid given the observer has perceived more than one interpretation of the stimuli in ambiguous figures but it has been showed that children, uninformed about the possible interpretations, between 3 and 4 years old, do not switch and that only 35% of uninformed children between 5 and 10 years old only experience reversals (Mitroff et al., 2006).

A.4.2 Mutual exclusivity

Mutual exclusivity of perception is clearly valid for ambiguous figures; however it is less clear in the case binocular rivalry. Indeed, in the latter experimental paradigm, subjects report their initial perception as an intertwined mixture of the two different images presented. The literature reports that alternations in binocular rivalry are progressive and spread along the visual receptive fields in a wave-like motion (Fig. A.12) invariant of stimuli visual angle size (Chopin, 2012). Thus, the following interpretation has been put forward: rivalry emerges from local coding competitions that have an impact on neighbouring neural populations, thus spreading the reversal. Hence, mutual exclusivity operates at a much more localised level in binocular rivalry.

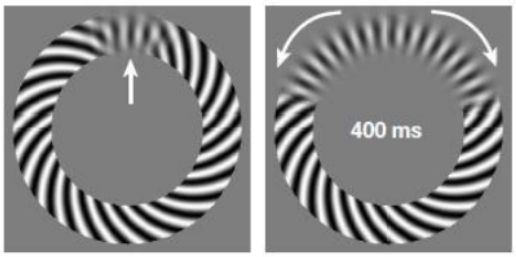


Figure A.12. Binocular rivalry. Illustration of the propagation of perceptual alternation in binocular rivalry. Figure taken from Chopin (2012).

A.4.3 Unpredictable

Multi-stability is characterised by the unpredictability of its perceptual reversals. Indeed, percept durations are thought to be independent from one to another (Fig. A.13) as it has been showed that one cannot predict the duration of a percept based on the previous one's duration. However, the unpredictability does not signify that multi-stability is random as it could be deterministic, more specifically chaotic. Lehky (1995) showed that binocular rivalry is stochastic rather than chaotic by using a method where percept durations data is separated in two blocks and where the first block of data is used to attempt to predict the second one. This method works in chaotic physical systems but fails for multi-stability. However, one should note that this has not be tried and reported yet on ambiguous figures (Chopin, 2012).

A.4.4 Percept durations distribution

A property that has been found across multi-stable phenomena is related to the density distribution of percept durations (Levelt, 1967; Chopin, 2012). For a long time, this aspect of the data was not examined into more details and the scientific community agreed that Gamma distributions modelled all bi-stable percept durations distribution (Fig. A.14). However, recently, it has been shown that Log-Normal distribution can sometimes fit the data more precisely—especially for the Necker Cube, rotating sphere and orientation rivalry stimuli (Hupé and Rubin, 2003; Pressnitzer and Hupé, 2005; Chopin, 2012). Furthermore, looking at reversal speeds[14] gives better results for Gamma distribution fitting (Jan Brascamp et al., 2005). Hence, reversal speeds have empirically been found to be modelled by a Gamma distribution while percept durations are better modelled by a log-normal distribution.

[14] Percept reversal speed corresponds to the inverse of percept's (x) durations: $v_x = \frac{1}{t_x}$.

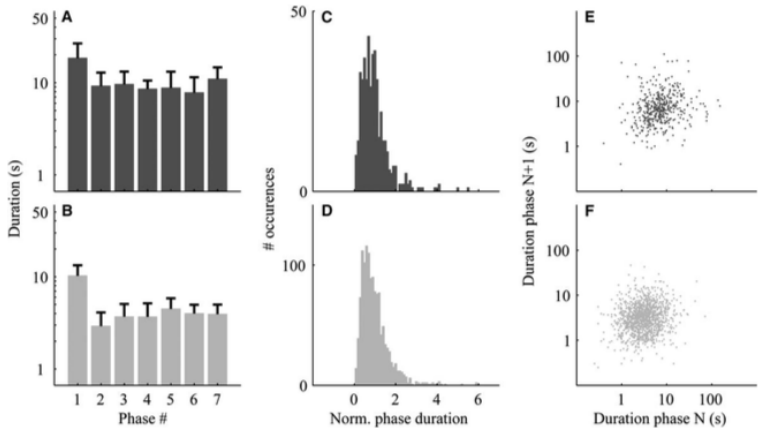


Figure A.13. Percept durations.
A & **B**: The durations of the first seven successive phases are presented for both the auditory modality (**A**, dark gray) and the visual modality (**B**, light gray).
C & **D**: The histograms of durations of grouped and split phases are presented, compiled for all participants (n = 23) and percept types (grouped and split).
E & **F**: The duration of a percept is shown as a function of the duration of the previous percept, for all participants (n = 23).
Figure taken from Pressnitzer and Hupé (2006).

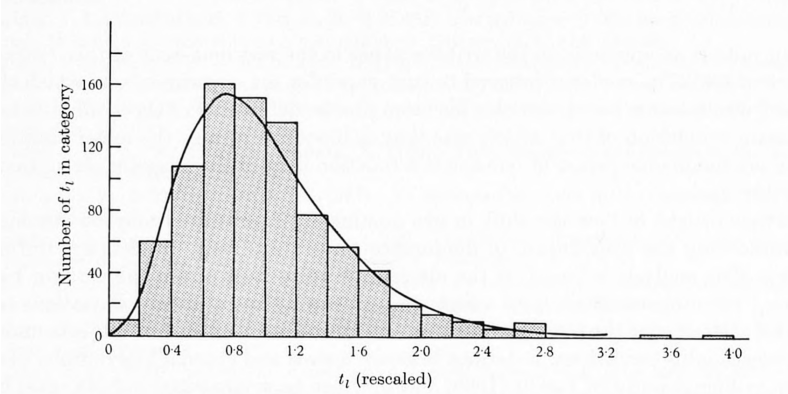


Figure A.14. Percept duration distribution. Example of histogram of phase duration distribution fitted to a Gamma distribution: $\phi(t) = t^3 \exp -t/6$. Taken from Levelt (1967).

A.4.5 First percept

The first percept at stimulus onset has been described as somewhat different and special from others (Hupé and Rubin, 2003): it is longer and tends to be idiosyncratic, e.g., biased for each subject (Fig. A.13). In binocular rivalry, the first percept has been linked to an ocular preference (Chopin, 2012) which disappears with the following phases. This has led some teams describe the temporal dynamics and discourse of bi-stable perception as a double regime phenomenon: an initial and a stationary one (Mamassian and Goutcher, 2005). Thus, using this approach, it is possible to describe bi-stable perception in a Bayesian framework (Ee, Adams, et al., 2003) by estimating probabilities *a posteriori*.

A.4.6 Theories & properties of bi-stable perception

Tab. A.1 provides a synthesis of theories and properties on bi-stable perception driven by empirical observations.

Bottom-up evidence	Description
Initial adaptation	Early trials can be subject to higher number of reversals reported as participants need to get used to the experimental setup.
Local adaptation	When a stimulus is rotated or moved in the visual field of the observer, perceptual reversal rates decrease as shown by Blake, Sobel, et al. (2003).
Multiple-figure presentation	Simultaneous observation of two or more bi-stable stimuli is characterised by independent reversals as well as independent adaptation (Toppino and Long, 1987).
Reverse-bias (priming)	A percept can be positively biased if the observer is exposed in a prolonged fashion to an unambiguous version of a percept (Long, Toppino, and Mondin, 1992).
(Dis-)Continuity of presentation	Introducing inter-stimulus-interval and making the presentation of the stimulus discontinuous affects reversal rates (D. Leopold, Wilke, et al., 2002).
Viewing parameters	The stimulus parameters (e.g., size, speed, etc) can influence the distribution of observed percept duration (Hupé and Rubin, 2003).
Top-down evidence	**Description**
Voluntary control	Volition has been shown to impact the speed of perceptual reversal in Wernery (2013). However it is not possible to prevent perceptual changes altogether, one can only decrease or increase the rates.
Knowledge of reversibility	Completely naive observers with no description of bi-stability are more likely to experience stable perception for long periods of time until the first change is experienced (Rock and Mitchener, 1992).
Priming (set effect)	Ambiguous perceptual experience can be influenced by presenting a cue that influence observers towards a percept, via binding for instance (Schwartz et al., 2012).
Cognitive load	Diverting attention to a distractor task has been shown to slow down the reversal process for ambiguous figures and binocular rivalry (Alais et al., 2010).

Table A.1. Bi-stability theories. Synthesis of bottom-up and top-down theories and associated empirical observations in bi-stable perception.

A.5 Multi-stability & neurosciences

Using fMRI has allowed researchers to identify regions and neural correlates of multi-stable perception. The fMRI studies of the blood oxygenation level dependent (BOLD) signals of V1 have shown that the spatio-temporal dynamics of perceptual alternations in multi-stability has strong effects on the measured activity. Indeed, as previously mentioned, this has allowed teams to predict perceptual states in rivalry based on the BOLD signal measured in V1 by using a multi-voxel pattern analysis (Haynes and Rees, 2005; Rees, 2007). Studies using MEG (Cosmelli et al., 2004; Parkkonen et al., 2008) have also shown that it is possible to read the neural bases of visual consciousness of participants. Interestingly, a team has shown that it is possible to induce perceptual alternations using TMS on V1, suggesting a causal role of the information coded at that level in modulating binocular rivalry (Pearson, Tadin, et al., 2007). However, this seems unlikely for ambiguous figures as competition seems to occur in higher cortical areas. Hence, these findings suggest that the early visual areas can be seen as a buffer for visual information and can be modulated by feedback mechanisms from higher-order areas that inhibit features (Sterzer et al., 2009).

The classic approach is to use a protocol that includes two perceptually equivalent conditions: natural rivalry and a biased-stimulus controlled rivalry (D. Leopold and Logothetis, 1999). The data of the two conditions are then subtracted from each other in order to mark out the areas where hemodynamic activity exists in bi-stable rivalry, which involve neural processes resolving conflict of perception, and the other condition, where vision is unambiguous. The results highlight activations in several areas of the frontal and parietal cortices in binocular rivalry, and seem to be systematically lateralised to the right hemisphere of the brain.

A.5.1 EEG: results from continuous viewing paradigms

EEG studies have attempted to characterise perceptual reversals more precisely by observing the time discourse with high temporal resolution. A classical approach is the manual response paradigm in which, a P-300 like parietal positivity, 250 milliseconds before key press has been consistently observed, as well as a decrease of energy in the alpha frequency band and an increase in energy in the gamma frequency band (Kornmeier and Bach, 2012). Gamma band increase has been interpreted as linked to attention and top-down processes. However, the paradigm shows some limits as the data show high inter and intra-individual variability and

applying signal processing through backward averaging damages the endogenous ERP.

A.5.2 EEG: results from discrete viewing paradigms

Another recent approach, the stimulus onset paradigm, consisting in having a discontinuous presentation of the stimulus with control of the ISI, has allowed researchers to address the issues mentioned above (Kornmeier, Ehm, et al., 2007; Kornmeier and Bach, 2012; Kornmeier and Bach, 2014). The stimulus is presented for a period and the participants report whether they have experienced a reversal in the following ISI. It offers the possibility to manipulate the stimulus presentation time, the ISI and the ambiguity of the stimulus. Thus, four types of ERP can be identified:

- Endogenous perceptual reversals (ambiguous stimulus)
- Endogenous perceptual stability (ambiguous stimulus)
- Exogenous perceptual reversals (unambiguous stimulus)
- Exogenous perceptual stability (unambiguous stimulus)

By subtracting reversal condition to stability condition data, one can extract the differential event related potentials (dERP) and characterise the signatures specific to endogenous reversals of ambiguous figures. Hence, a reversal positivity 130 ms after onset has been reported with better precision than in the manual response paradigm (see Fig. A.15). Moreover, early alpha modulations that start around 130 milliseconds after onset, are restricted to endogenous reversals, have an opposite sign as the subsequent components, and lasts for approximately 60 ms (see Fig. A.16). Kornmeier and Bach offered the following interpretation: reversal positivity could be an indicator of the visual system's detection of ambiguity while the alpha modulation could be the disambiguation or stabilisation process. In both endogenous and exogenous reversals, a reversal negativity is reported followed by fronto-polar and parietal positivities (Fig. A.15). The authors proposed that the fronto-polar positivity might indicate a role of working memory related to the delayed response in the subsequent ISI while the parietal positivity might be an indicator of attentional and cognitive processes during perceptual reversal and be a signature of the conscious recognition of the switch. Time frequency analysis showed Beta and Gamma band modulations have been measured in both exogenous and endogenous conditions with high similarities (Fig. A.16). Thus, these results provide

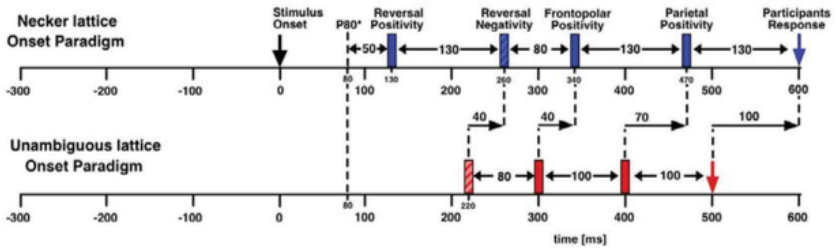


Figure A.15. Differential event related potentials (dERP) analysis. Diagram showing detected dERP in the discontinuous viewing onset paradigm. The ambiguous conditions shows events that are delayed compared to its unambiguous counterpart and an extra reversal positivity. Figure taken from Kornmeier and Bach (2012).

arguments that bring together the top-down and bottom-up understandings of multi-stability. Their integrated theory is founded upon two processes: a destabilisation, that can be initiated from top-down, bottom-up processes or noise over variable and potentially long periods, and a restabilisation or disambiguation, that follows the reversal positivity and that operates within approximately 60 ms to provide a new interpretation to the observer. Finally, following a similar paradigm, researchers have shown that working memory load modulated the observed ERP in reversals of the Necker cube though the behavioural reversal rates remained unchanged (Intaitė et al., 2014). Hence, this suggests that in discontinuous presentation of ambiguous figures, reversals are influenced by an early mechanism and that the effects of the load on the ERP might correspond to prefrontal cortex outbound top-down processes' impact on visual processing.

A.5.3 EEG & MEG: frequency tagging

An alternative paradigm has been explored in EEG and MEG by using a stimulus' frequency attributes to retrieve its content in cortical oscillations. This method is described as **frequency tagging** whereby the experimenters will add dynamic noise at two tagged frequencies for each part of the stimulus that corresponds to a percept. For instance, this was done on the Face/Vase stimulus which features a figure-ground ambiguity and percepts can be separated spatially (Parkkonen et al., 2008)—see Fig. A.17. Another method was to tag the motion of the rival stimuli to a frequency in binocular rivalry (Cosmelli et al., 2004). Finally, in a more subtle manner, researchers have shown that when the perceptual competition occurs over spatial frequencies, these can be exploited to retrieve perceptual information in EEG (M. Smith et al., 2006).

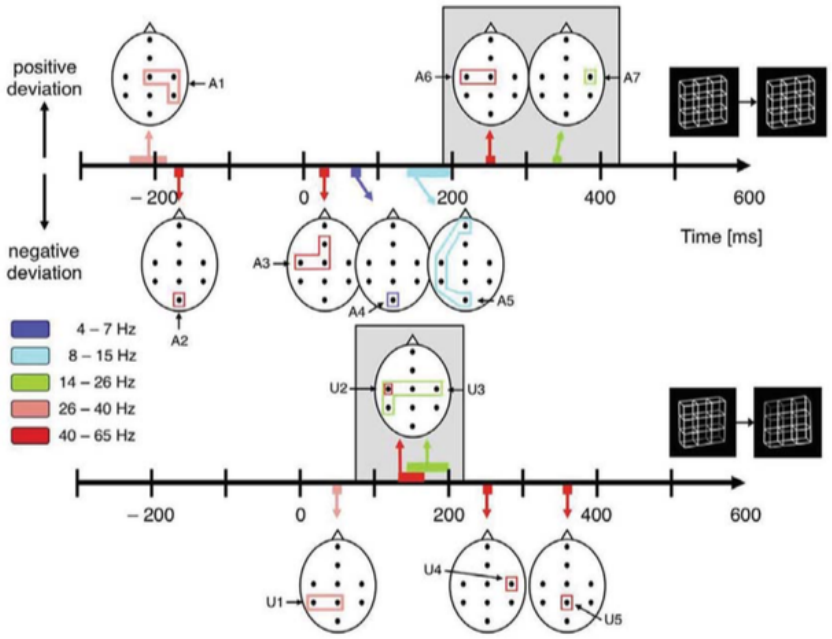


Figure A.16. Time-frequency (TF) analysis. Diagram showing the results over time of TF analysis and sources in the EEG data. On top, the dynamics for the ambiguous Necker lattice. Below, the results for the unambiguous condition. Figure taken from Kornmeier and Bach (2012).

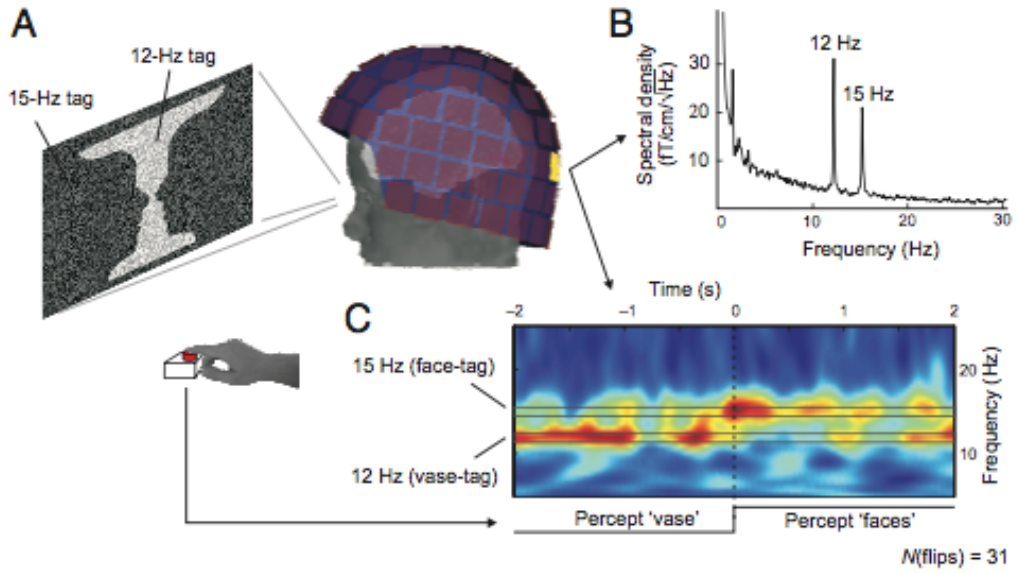


Figure A.17. Frequency tagging.
A: Experimental MEG setup with the Face/Vase frequency tagged stimulus.
B: Example of a spectral density at one sensor with peak power on the tagged frequencies.
C: Example of a time-frequency analysis plot with energy at the tagged frequency over a reported perceptual change. Figure taken from Parkkonen et al. (2008).

A.5.4 fMRI: percepts in early visual cortex

Neuro-imaging observations of the extrastriate visual cortex in binocular rivalry and ambiguous figures show a closer correlation with conscious perception[15]. Depending on the nature of the ambiguous figure and the features that initiate the bi-stability, BOLD activity in specialised areas such as the fusiform face area (FFA) is greater in Rubin's face-vase illusion (Sterzer et al., 2009). The lateral occipital complex (LOC) had higher BOLD activity when the conflict was centred on elements being perceived as grouped or randomly arranged. Alternatively, some studies have focused on neural activity in perceptual reversal events rather than perceptual states in multi-stability. Activity in extrastriate areas is correlated to these changes; for instance FFA goes through higher excitation before reversal to a face percept following a grating percept (Sterzer et al., 2009). Furthermore, ambiguous figures have been classified in two categories—intra-categorical type and cross-categorical type—and fMRI data has been analysed and contrasted to identify whether all the stimuli share a common reversal-related neural network (Ishizu and Zeki, 2014). This has been investigated and could be explained by the presence of coded information of multiple past percepts in a time window preceding the switch that has been described as a form of perceptual memory (Pearson and Jan Brascamp, 2008), which seems to play a role in how perceptual conflicts are resolved in ambiguous vision.

A.5.5 fMRI: blind decoding of percepts

Though it is possible to read the content of conscious perception using neuro-imaging techniques, the mechanisms involved in resolving conflict in the sensory inputs remain elusive. These mechanisms relate to unconscious inference processes that are characteristic to human vision. Studying reversals is more appropriate in understanding the reconstructions that lead to visual awareness than perceptual states. When contrasting ambiguous and unambiguous vision, it is mostly activations in the parietal and prefrontal cortex that stand out (Lumer et al., 1998; D. Leopold and Logothetis, 1999). Thus, it seems that top-down cognitive processes and networks seem to be highly involved in multi-stability and might operate through feedback mechanisms. These areas are typically engaged in cognitive functions such as working memory, attention and decision making. The feedback process could serve as a restabilisation of perception following sensory destabilisation (Sterzer et al., 2009) and take part in the inferential mechanisms. However fMRI studies

[15]Refer to the following paragraph on electro-physiology results for further evidence of this statement.

have shown their limits in studying the reversal processes due to their low temporal resolution.

A.5.6 Electro-physiology

Furthermore, invasive single neurons monitoring of monkeys in visual rivalry have been studied with probes in the striate visual cortex (V1), the extrastriate visual cortex (V2, V4), the MT area, the medial superior temporal (MST) *sulcus*, the infero-temporal (IT) cortex and the upper and lower bank of the superior temporal sulcus (STS). This revealed that even though the retinal image remained the same, some neurons in the observed visual information pathway described above were consistently modulated by the monkeys' perceptual changes. However, some neural populations' activity (V4, MT, MST) was more correlated to perceptual switches than others (V1, V2). Furthermore, neurons' activity patterns in IT and STS were closely matched to perceptual reporting while activity in the LGN showed no correlation. This suggests that conflicts in multi-stable perception are solved in higher cognitive processes. Lesion studies have shown that only patients with unilateral frontal lesions have an impairment in experiencing switches of subjective perception, while patients with posterior brain damages, including visual areas, did not have significantly different behaviours from control subjects.

A.6 Eye movements & the plaid

An aim of this work is to link motion ambiguity to eye movements. Indeed, these are physiological dynamics when measuring the gaze's motion, that can be linked to the perceptual multi-stability (García-Pérez, 1989). When perceiving motion, the visual system may lead the oculomotor system to program different types of eye movements such as saccades (Sabrin and Kertesz, 1980; L. v. Dam and Ee, 2005; L. C. v. Dam and Ee, 2006a), pursuits (Beutter and Leland Stone, 1998; Madelain and Krauzlis, 2003), FEM (Laubrock et al., 2008), pupilometry (Hupé, Lamirel, et al., 2009) and OKN (Einhäuser, Thomassen, et al., 2017; Aleshin et al., 2019).

Saccades were investigated as potential markers of perceptual changes since they operate great changes on the visual input. Studies on binocular rivalry found that saccades and micro-saccades are more likely to occur near reported perceptual changes (Sabrin and Kertesz, 1980; L. v. Dam and Ee, 2005). However, further analyses suggested that the retinal shift due to saccadic eye movement was a better criteria (L. C. v. Dam and Ee, 2006a), hence the authors concluded that spatial attention (highly coupled with eye movements) may be the key modulator of perceptual dynamics. These studies used a variety of stimuli in binocular rivalry including, at times, square wave gratings, rather than the more ecological approaches from ambiguous figures. Though saccades may not be the most correlated eye movements to perceptual changes, micro-saccades seem to provide insights on perceptual dynamics. In an ambiguous apparent motion discrimination task, Laubrock et al. (2008) reported an inhibition of micro-saccades before perceptual reversals and that orientation of the micro-saccades before stimulus onset biased the subsequent perceptual decision. They interpreted micro-saccadic movements to be used as a cue for the perceptual system when ambiguity is strong, in order to force a perceptual decision. Therefore, micro-saccades may be a key marker to analyse and identify in ambiguous motion perception.

The moving plaid, as a motion ambiguity stimulus, has also sparked investigative work focusing on pursuit eye movements. In a study looking for evidence on whether the a motion-processing system links perception and pursuit, using a plaid stimulus, the authors found shared biases[16] in the perceptual and pursuit analyses (Beutter and Leland Stone, 1998). This suggest that both systems share a motion processing stage which may be exploited in order to derive perceptual dynamics in oculomotor signals, by focusing on pursuit analysis. Moreover, Madelain

[16] Psycho-physical bias measured by how much, observers perceived the plaid motion to be higher or slower than its physical velocity direction. For eye movements, bias was a measure of the difference between processed gaze patterns and the physical velocity directions of the plaid.

and Krauzlis (2003) showed, using ambiguous apparent motion generated by a series of Kanizsa illusory squares, that pursuit can be a marker of perceived motion since it is highly correlated to motion-based percepts. They showed that humans are more likely to track their current percept and may reverse perceived motion with no need for saccadic eye movements. The motion of this dynamic stimulus has stable velocities within percepts and with directions of percepts being different, thus it is possible to conceive methods to detect changes in the gaze signals. Using OKN as a marker of perceptual states has been used before in binocular rivalry (Frässle et al., 2014; Aleshin et al., 2019) and similar methods can be applied to other motion based perceptual multi-stability. In these studies, the authors developed gaze analysis method to remove artefacts such as blinks and large saccades, in the eye movement signals, and shifted the slow pursuit components in order to estimate a cumulative smooth pursuit of the motion. They were then able to analyse their data by inferring perceived motion exploiting the accumulated pursuit events.

However, it has been shown that these oculomotor markers are not systematically present, and the responses from the visual and motor systems can be dissociated (Spering and Gegenfurtner, 2008; Spering, Pomplun, et al., 2011). By reviewing many studies investigating whether pursuit is impaired, enhanced or unaffected by stimuli with stationary context, pursuit direction motion context, opposite pursuit direction motion context and velocity perturbations contexts, the authors showed that when no orthogonal context perturbations are manipulated, experiments may generate no effects on pursuit.

But the results shown in these papers were obtained in short viewing paradigm with the stimulus being presented for less than 500 ms, and without multi-stable dynamics. Another existing approach has been based on pupil size, though its main results link most of the pupil changes preceding key press and thus perceptual reversal report have been linked to the motor programming rather than the endogenous change (Hupé, Lamirel, et al., 2009). Such results remind us that key press bias remains a factor present in most multi-stability experiment and may pose challenges to the interpretation of results.

> **Overall, one can expect to observe markers that show perceptual reversals and indicate which percept just appeared to the observer's visual awareness. A combination of methods detecting oculomotor markers predicting perceptual reversals, such as micro-saccades, and markers correlated to a percept, such as pursuit for motion based stimuli like the moving plaid, are needed to adequately infer perceptual dynamics in eye movement signals.**

Such methods are investigated in the *Eye movements* and *No Report* experiments (presented in Chapter 5, Appendix B.2 & Appendix B.3), where we test methods to (1) manipulate eye movement dynamics in order to investigate causal influences of gaze on perception and (2) read perception in gaze data in order to compare dynamics with and without key press reports of perception.

A.7 Can we remove subjective reports on the moving plaid?

Recent studies have attempted to remove subjective key press reports to find more objective physiological measures of perceptual dynamics (Frässle et al., 2014; Einhäuser, Thomassen, et al., 2017; Aleshin et al., 2019). If one considers also the decoding methods reported when combining multiple neural measures and machine learning classification methods on EEG and MEG signals, by focusing on detecting markers of perceptual changes (Cosmelli et al., 2004; M. Smith et al., 2006; Parkkonen et al., 2008) or in fMRI signals, by focusing on detecting markers of perceptual states (Rees, 2007; Sterzer et al., 2009)—*for a summary of important neurosciences results for the moving plaid, read Appendix A.8*. Moreover, participants tend to report attentional shifts associated with the motor action of using the keyboard to report their perception (Mamassian and Goutcher, 2005). Similarly, observers tend to report perceptual switches linked to blinking (L. v. Dam and Ee, 2005). In other psychology paradigms, such as masking based on *Signal Detection Theory* (Green, Swets, et al., 1966), the development of methods, that make the measures more objective has been addressed extensively (King and Dehaene, 2014). The expected advantages of removing key press reports are the following:

- *perceptual noise reduction*—motor programming and attentional shift may act as large noisy signals in the perceptual systems as interpreted by the models presented in Chapter 1 & 3 (Moreno-Bote, Rinzel, et al., 2007; Shpiro, Moreno-Bote, et al., 2009; Moreno-Bote, Knill, et al., 2011);
- *neural correlates studies*—no report paradigm are being developed and used to explore the neural bases of conscious experience in consciousness research (Frässle et al., 2014; Michel and Morales, 2019);
- *evidence of an active perception*—the methods of such an approach impose the necessity to show a strong correlation link between physiological markers and perceptual events (Laubrock et al., 2008; Einhäuser, Thomassen, et al., 2017; Aleshin et al., 2019).

Using the theoretical framework presented in Chapter 3, it is possible to derive hypotheses and associate predictions on perceptual and oculomotor dynamics in coupled systems.

Neural correlates of consciousness.
No-report paradigms in experimental neuroscience profit from a growing interest as more researchers aim to investigate the NCC and this has been made possible by the development of neuro-imaging techniques (Crick and Koch, 1990). NCC theories differ and the topic is far from being settled with for instance some authors providing evidence for the *Global Work-Space Theory* associated to pre-frontal cortex (PFC) activity (Baars and Franklin, 2007; Dehaene and Changeux, 2011). *Higher-Order Theory* also view PFC activity as a NCC (Lau and Rosenthal, 2011). In opposition, other researchers defend that NCC are situated in posterior parts of the cortex (Lamme, 2006; Koch et al., 2016). PFC theories have been criticised as its activity does not correlate with consciousness *per se* but with cognitive processes that follow perceptual awareness (Michel and Morales, 2019). This argument emerges from methodological constraint because experiments require participants to report whether they have perceived the phenomenon of study consciously or not.
A solution to such arguments is to develop methods that make it possible for experimenters to infer the consciousness of a percept during a task, or the alteration of an observer's perceptual awareness without relying on explicit report. If this is achieved, observers will not have to introspectively reflect on the content of their consciousness, making them use their *meta-cognitive* abilities (Faivre, Filevich, et al., 2018; Faivre, Vuillaume, et al., 2020). To apply such methods, experimenters must find methods in which they can reliably infer conscious perceptual dynamics. Multi-stability offers many advantages for such problems. Indeed, the aspect in which we are interested is perceptual awareness changes; this is different from contrasting conscious *versus* unconscious perception. Changes in perceptual consciousness may reveal some of the cortical areas necessary to operate perceptual reorganisation (Kawamoto and J. A. Anderson, 1985; Baå-Eroglu et al., 1993).
The field of consciousness studies, and more specifically NCC, have tried to use binocular rivalry to develop no-report paradigms.

A.8 Gaze-EEG experimental design

Note.
The following sub-section provides insight on the Gaze-EEG experiment that was designed and motivated all the other experiments reported in this chapter. Unfortunately, due to administration delay—e.g., French "*comités de protection des personnes*" (CPP) procedure—and the COVID-19 sanitary crisis, the experiment has only been launched on one participant at the time of writing. This part aims to provide further details on the experimental design that motivated the following experiments described in Chapters 4 and 5. It is recommended to read these chapters before reading this section as some of their results are referred to.

A.8.1 Neurosciences and the plaid

Another set of data that can be investigated are the neural correlates of multi-stability. These can be obtained with techniques from neurosciences such as electro-physiology, EEG, MEG or fMRI.

Activity in the middle and higher levels of the visual system pathway has been correlated to motion perceptual processing, notably in the *middle temporal* (MT or V5) area of the cortex (Stoner and Albright, 1992; Ferrera and Lisberger, 1997; Born et al., 2000; Castelo-Branco et al., 2000; Thiele et al., 2000; Williams et al., 2003). This cortex area is in fact sensitive for motion perception as lesions in monkey primates has been shown to degrade significantly performances for direction detection tasks (Newsome and Pare, 1988). One can therefore expect to observe changes in activity patterns as the direction of perception changes for an observer, using neuro-imaging techniques with high temporal resolutions—e.g., EEG, invasive electro-physiology or MEG.

Perceptual changes can also be expected to generate the *P300* component in ERP analysis as it is associated to a measurement surprise effects and can be reliably detected (Donchin, 1981; Cecotti et al., 2011). Surprise here is referred to a revision of internal models. An extensive study of EEG components for a bi-stable lattice of Necker cubes has reported fronto-polar positivity a little over 300 ms after onset, in a non-continuous viewing paradigm (Kornmeier and Bach, 2012). This approach is supported by arguments proposing that perception's continuous appearance is an illusion in itself, in which brain or neural rhythmics play a role in attention and perceptual processing (D. Leopold, Wilke, et al., 2002; VanRullen and Koch, 2003; VanRullen, Busch, et al., 2011; VanRullen and Dubois, 2011; Dubois and VanRullen, 2011). The use of oscillatory brain activity has been used to relate percepts to spatial frequencies (M. Smith et al., 2006) and to decode percepts by means of frequency

tagging and TF analysis in MEG (Cosmelli et al., 2004; Parkkonen et al., 2008). The Gamma band at 30-80 Hz has been reported to be linked to perceptual changes and in fronto-parietal networks while alpha band energy is reported to be higher in stable perceptual phases (Başar-Eroglu et al., 1996; Strüber, Basar-Eroglu, et al., 2000; Strüber and C. S. Herrmann, 2002; Mathes et al., 2006; Kornmeier and Bach, 2012; Strüber, Rach, et al., 2014; C. S. Herrmann et al., 2016).

Experiments in fMRI have also shown that it is possible to predict and decode the perception based on activity in cortical areas (Haynes and Rees, 2005; Brouwer and Ee, 2007; Rees, 2007; Sterzer et al., 2009; Megumi et al., 2015). The methods presented also showed evidence that the competition in multi-stable perception occurs at different stages of the visual cortex. Motion rivalry is reported to be detected in the MT/V5 area for instance, while figure-ground dynamics are decoded from lower visual area (D. Leopold and Logothetis, 1999).

Overall, parietal and prefrontal cortex activity correlates to perceptual changes though frontal activity might be related to meta-cognitive and introspective activities (Kanai et al., 2010; Frässle et al., 2014).

The aims in this experiment are to (1) manipulate ambiguity, (2) manipulate oculomotor dynamics and (3) to be able to infer perception without relying on explicit motor commands, i.e., key presses. Note that this experiment adds EEG measurement to previous experimental setups. An expected challenge from such an experiment is to show that gaze-EEG coupled analysis can be applied on a continuous viewing paradigm (Devillez et al., 2015; Kristensen, 2017). Indeed, the issues for EEG and MEG studies on bi-stable perception studies are centred around temporal synchronisation for signal analysis (Kornmeier and Bach, 2012). Since most multi-stable perception experiments rely on key press reports to relate the dynamics of the subjective perceptual experience, and that the motor response latencies can vary between 200 ms and 600 ms, the signals are hard to align, unlike in classical on-set paradigms.

Ambiguity control based on the *Ambiguity* experiment's results

The *Ambiguity* experiment (Section 4.3) led to the development of methods that allow to control the level of ambiguity of the moving plaid stimulus by relying solely on gratings' transparencies ($\boldsymbol{\alpha}$) as control parameters. We reduced this manipulation to one of the two gratings' transparency by constraining the relationship between α_L and α_R to (4.9): $\alpha_L + \alpha_R = 1$. We introduced a psycho-physical observer model (Bak and Pillow, 2018) of multi-stable perception in this context, with an associated

method for its parameters' estimation. Using an experimental protocol focusing on the first percept, we could gather many data points in a relatively short amount of time for estimation. In a second phase, we verified that the maximal ambiguity point(s) could be established, given the parameters estimated, and whether longer trials at such points would lead to empirical perceptual equi-probability. Therefore, for the Gaze-EEG experiment, the aim is to implement the methods of the *Ambiguity* experiment such that the stimulus can be *calibrated* using a procedure composed of first-percept short trials. These data points can be fed to the estimation algorithm and a psycho-physical observer profiles of the participants can be used to control ambiguity.

Oculomotor control based on the *Eye Movements* experiment's results

In the *Eye Movements* experiment (Appendix B.2), we aimed to manipulate oculomotor dynamics, in order to able to generate micro-pursuits that are coherent with a percept motion or micro-saccades in the opposite direction. Using a RDK (Scase et al., 1996), micro-pursuit can be elicited and according to the direction of its signal, trials can either be excitatory or inhibitory with respect to a percept. Alternatively, we expected micro-saccades to be generated as participants would re-centre their gaze on the fixation dot. However, the lack of conclusive results does not guarantee that RDK, as manipulated in the *Eye Movements* experiment, is sufficiently effective to generated the expected FEM. Hence, not all the relationships from Fig. 4.6 can be investigated; namely, the influence of eye movements on perception cannot be controlled. Therefore, one can only expect micro-pursuits and micro-saccades to be generated according to the model's predictions, given in Section 4.1. The manipulation is still expected to be observed in trials, but it is not possible to directly distinguish the impact of the perceptual system from the oculomotor system—presented in the fusion model of Chapter 3. In other words, even if FEM and oculomotor dynamics cannot be controlled, they are still to be measured in this Gaze-EEG experiment, and their contribution to the phenomenon of multi-stable visual perception can still be studied.

No-report perceptual change detection based on the *Noisy Motor Events* experiment's results

The *Noisy Motor Events* experiment (Section 5.3) led to the development of gaze signals analysis algorithm that can detect epochs with dynamics suggesting a micro-pursuit associated to the emergence of a new percept, following a perceptual reversal.

The algorithm sequences the gaze signal such that periods can be analysed and interpreted in the following manner: changing oculomotor dynamics with similarities to the motion dynamics of one of the percepts reveals changes in perception. Though the results in the *Noisy Motor Events* experiment shows discordance between key press reports and oculomotor interpretation, such methods are being used to provide alternative approaches for perceptual reports (Frässle et al., 2014; Aleshin et al., 2019). In fact, report and no-report conditions can be used to contrast and adjust the sensitivity of the algorithm's interpretation. Though a no-report setup exposes the data to have higher uncertainty, subjective reports as reported in Chapter 1 also relies on trusting participants capacity to report their perceptual changes. And the latter is sensitive to variable reaction times due to variability in motor programming for key press. Finally, attention is shifted over key press action (Song, 2019), hence affecting perceptual dynamics (Li et al., 2017). This uncertainty can however be reduced if neural correlates of perceptual changes are detected jointly with gaze dynamics changes.

Gaze-EEG

The experimental setup in this experiment adds joint EEG measurements. The constraints associated with using a dynamic multi-stable stimulus such as the moving plaid makes the use of a continuous viewing paradigm necessary over discrete presentation methods (Kornmeier and Bach, 2012). Indeed, interrupting a video in an analogous fashion to blinks, but with exogenous origins, requires a series of additional investigations to answer questions related to issues with this approach. For instance, it is necessary to estimate how the stimulus should evolved over the ISI period, and whether the valid ISI duration is similar to that of other static bi-stable stimuli (D. Leopold, Wilke, et al., 2002).

Moreover, neural correlates of multi-stable perceptual dynamics have been studied (as shown by the literature review above), and can be exploited for no-report blind condition trials to infer perceptual changes. Indeed, changes in perception are expected to be marked by higher Gamma band (30-80 Hz) activity (Başar-Eroglu et al., 1996; C. S. Herrmann et al., 2016) and activity has been reported to occur in parietal and prefrontal cortices (Kanai et al., 2010; Frässle et al., 2014). Another component expected to be detected in endogenous perceptual changes is the P300, as it is an ERP that is detected with precision, so much so that it is exploited in brain computer interfaces (Donchin, 1981; Cecotti et al., 2011). Hence, EEG signals can provide additional information and help give an internal neural understanding of multi-stable perception. Though eye movements can lead to artefact generation in

the signals, this is less true for small amplitude movements such as the FEM that are expected to be identified. Moreover, recent algorithms have allowed the development of joint measures Gaze-EEG, using eye movements as onset for ERP analyses with eye fixation related potentials (EfRP) and eye saccade related potentials (EsRP) (Devillez et al., 2015; Rivet et al., 2015; Kristensen et al., 2017; Kristensen, 2017).

Hypotheses

The main hypothesis was that eye movements are physiological markers of multi-stable perception. More specifically, with the moving plaid stimulus, we expect perceptual changes to generate oculomotor markers such micro-pursuits in the direction of percept motion and micro-saccades in opposite direction to that. Our theoretical work from Chapter 3 also predicts that oculomotor dynamics lead to perceptual changes, but this aspect will be less controlled given the results of the investigation in the *Eye Movements* experiment (Appendix B.2). The main two contrasts chosen to obtain results with respect to these hypotheses are the ambiguity and report manipulations. We expect to obtain gaze-EEG markers of perceptual changes using the non-ambiguity and no-report condition and the ambiguous report condition. The non-ambiguity with key press report will act as a control condition where all data is clearly labelled, but stable perception rather than multi-stable perception occurs. The ambiguity no-report, blind condition is the most challenging one, where data will not be labelled but will also have the least attentional and motor programming undesired artefacts. Using the algorithms developed in the *Eye movements* and *Noisy Motor Events* experiments (Chapter 5), and with subsequent additional EEG analysis, perceptual dynamics will be inferred.

B Experimental metrics, modules and designs

B.1 Maximally Projected Correlation

B.1.1 More details on MPC

We use $\cdot^\top$ for the transpose operator and $\mathrm{trace}(\cdot)$ will denote the trace operator (sum over the diagonal elements of a matrix). The identity matrix in dimension 2 will be denoted by Id_2.

Variance-covariance and inertia

Let $\boldsymbol{q}_G(t) = (x_G(t), y_G(t))^\top$, $\boldsymbol{q}_S(t) = (x_S(t), y_S(t))^\top$, and $\boldsymbol{q}_R = (x_R(t), y_R(t))^\top$ be the screen Cartesian coordinates (column vectors) at time instant t of the Gaze, the stimulus, and the retinal image, respectively. Now, having samples at n discrete times $\{t_i\}_{i=1}^n$, we estimate the centre of gravity of a gaze trajectory $\{\boldsymbol{q}_G(t_i)\}_{i=1}^n$ by its empirical mean $\boldsymbol{m}_G = n^{-1}\sum_{i=1}^n \boldsymbol{q}_G(t_i)$. This estimate approaches the true centre of gravity if we sample sufficiently regularly and beyond twice the Nyquist frequency, conditions that are met when working with the EyeLink 1000+, sampling at about 1000 Hz for each eye.

A second-order statistic of interest is the empirical variance-covariance matrix, which gives the inertia of the gaze trajectory defined as

$$\boldsymbol{\Sigma}_G = n^{-1}\sum_{i=1}^{n} \boldsymbol{q}_G(t_i)\boldsymbol{q}_G^\top(t_i) - \boldsymbol{m}_G\boldsymbol{m}_G^\top \tag{B.1}$$

and analogously for the stimulus and retinal image empirical variance-covariance matrix. The inertia about its centre of gravity is then given by

$$I_{\boldsymbol{m}_G} = n^{-1}\sum_{i=1}^{n} \|\boldsymbol{q}_G(t_i) - \boldsymbol{m}_G\|^2 = \mathrm{trace}\left(\boldsymbol{\Sigma}_G\right) \quad . \tag{B.2}$$

The inertia I_r of the gaze trajectory $\boldsymbol{q}_G$ with respect to any fixed point $\boldsymbol{r}$ having screen coordinates $(x_r,\, y_r)$ is

$$I_{\boldsymbol{r}} = \operatorname{trace}\left(\boldsymbol{\Sigma}_G\right) + \left(\boldsymbol{m}_G - \boldsymbol{r}\right)^\top \left(\boldsymbol{m}_G - \boldsymbol{r}\right) \quad . \tag{B.3}$$

Maximally projected correlations

Taking now the simultaneously recorded gaze $\{\boldsymbol{q}_G(t_i)\}_{i=1}^n$ and stimulus $\{\boldsymbol{q}_S(t_i)\}_{i=1}^n$ signals, and their respective empirical variance-covariance matrices $\boldsymbol{\Sigma}_G$ and $\boldsymbol{\Sigma}_S$, we denote the inter-covariance matrix by

$$\boldsymbol{\Sigma}_{GS} \doteq n^{-1} \sum_{i=1}^{n} \left(\boldsymbol{q}_G(t_i) - \boldsymbol{m}_G\right) \left(\boldsymbol{q}_S(t_i) - \boldsymbol{m}_S\right)^\top = \boldsymbol{\Sigma}_{SG}^\top \quad . \tag{B.4}$$

This matrix is particularly useful when considering the inertia of gaze with respect to the time-changing coordinates of the stimulus. Indeed, after some manipulations, we obtain:

$$I_{GS} = n^{-1} \sum_{i=1}^{n} \left\| \boldsymbol{q}_G(t_i) - \boldsymbol{q}_S(t_i) \right\|^2$$
$$I_{GS} = \operatorname{trace}\left(\boldsymbol{\Sigma}_G + \boldsymbol{\Sigma}_S - \boldsymbol{\Sigma}_{GS} - \boldsymbol{\Sigma}_{SG}\right) + \left\| \boldsymbol{m}_G - \boldsymbol{m}_S \right\|^2 \quad .$$

Unfortunately, the inertia does not account for differences in scale, nor for coordinate translation, two characteristics that are typical aspects for pursuits and for which we require an invariance. Indeed, we suppose the stimulus will always be at a constant phase with respect to the gaze, either lacking behind in phase (catching up on the stimulus) or ahead of phase (prediction), the scale difference is our main objective, showing that the stimulus trajectory is reproduced at a smaller scale and, finally, the coordinate translation shows a systematic bias in the trajectories.

Noise robustness & signal size dependency

Fig. B.1 shows results of simulations operated on a Lissajous signal degraded noise on the position of the stimulus at different signal to noise ratio (SNR) and for different signal sizes. For signals with more than 167 samples, the behaviour of MPC scores over SNR remains stable and shows quasi-unchanged dynamics.

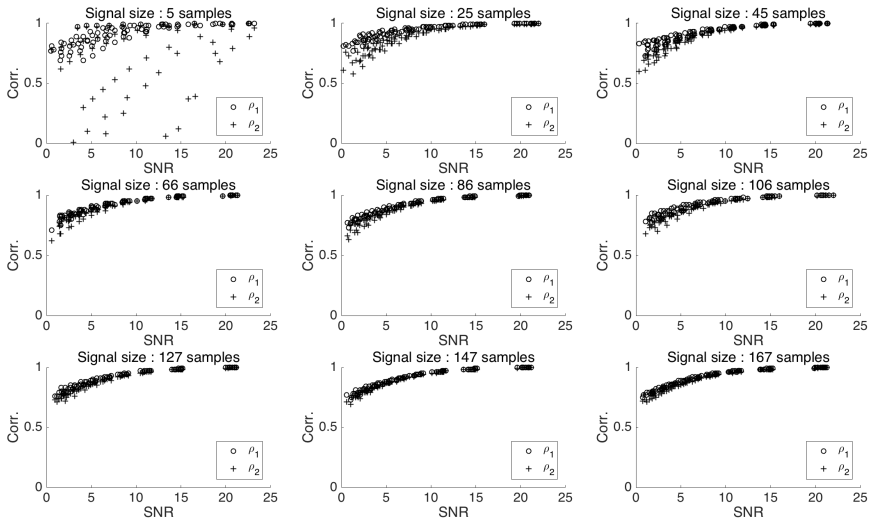


Figure B.1. Behaviour of MPC scores over SNR in simulated similarity computations with a Lissajous base signal from LJ.

B.2 Eye Movements experiment

B.2.1 Eye movement manipulations

To align our experiment with most of the past research on the stimulus, we chose to focus on manipulating the central, foveal, part of the stimulus in order to keep the retinal image relatively stable. By opting for an implicit manipulation method, we aimed to generate oculomotor markers of attractor dynamics that would generate FEM. The FEM expected, linked to attractor motion and inhibition levels, can either be, in this context, micro-saccades or micro-pursuits. Micro-saccades are expected to appear when attentional or intentional properties of the energy field, or in other words the parameters of the attractors, have highly unstable dynamics over time, or alternatively, when many attractors are integrated over a stimulus. On the other hand, we expected micro-pursuits to occur when attractors are relatively balanced, but the oculomotor system has sufficient noise to allow the gaze particle to jump into another attractor. After micro-pursuits, we expected that corrective micro-saccades or small amplitude saccades would occur as observers remember to focus on the fixation dot or the central part of the disk (see Fig. B.3). Finally, if the system does not have sufficient noise, we expected the gaze to stay stable in the fixation area with mostly some oculomotor drift and tremors FEM being detected.

Possible methods for eye movement manipulation in moving plaid stimulus

In the literature, moving plaid experiments use a disk area at the centre of the stimulus (see Fig. B.3) in which a fixation dot is placed, so that observers may keep their gaze fixated on it. Objects such as crosses, dots and combinations of them, can be used in psycho-physical experiments to help maintain the eyes in a steady position (Thaler et al., 2013). This is mostly done to reduce retinal image shifts and variability in the information flow for the visual system. Furthermore, the size of this disk is typically of 2-3 deg and added to minimise OKN occurrences, which means the gratings are present beyond central vision (Hupé and Rubin, 2003). This area of the stimulus can be manipulated more or less explicitly: for instance, the fixation dot could have movement that is coherent with percept motion to reinforce its probabilities of being perceived. Alternatively, it should be possible to generate motion in the periphery to increase oculomotor motion probabilities (Spering and Gegenfurtner, 2008). The difficulty lays in finding a balance between explicit methods that would be prone to top-down biases such as the *demand* and *response* biases (Firestone and Scholl, 2016) and implicit methods with small effect on the oculomotor dynamics, thus a high number of occurrences are needed because of low statistical power. A last argument to keep in mind in approaching such a problem is that in the context of moving plaid perceptual tasks, trials are long (over 30 s) in order to generate endogenous perceptual reversals.

One method to induce oculomotor biases is to use a RDK at the centre of the stimulus. Our aim was to establish a coherence ratio that would generate implicit micro-pursuit movements in the same direction as the signal and corrective micro-saccades in the opposite direction.

> **We expected that given the RDK signal has the same direction as one of the percepts, we would observe micro-pursuits in the direction of the signal and a bias with percept time reported for the percept with that same direction. Thus, this condition would be an excitation of that percept by manipulation of oculomotor dynamics. On the contrary, if the RDK signal has the opposite direction[1] as one of the percepts, we would observe micro-pursuits in both the direction of the RDK signal and the percept, combined with a bias on lower percept time reported. Here, this condition would be considered as an inhibition of the percept by manipulation of eye movements.**

[1] Orthogonality, in retrospective, may be a better option.

The relationship was not expected to be necessarily symmetrical; indeed inhibition should be less powerful as excitation.

Random-dot kinematogram.
A RDK is a set of points that have random movements in a defined area. Different implementations exist (Scase et al., 1996) with the three main noise methods combined with two dot signal selection methods (see Fig. B.2).

- **Random position**—when a dot is selected as noise, its next position will be chosen using a random distribution—mostly a uniform distribution over an area, but not necessarily—and it will be relocated at that position at the next iteration.
- **Random walk**—when a dot is selected as noise, it will follow a random walk, also known as Brownian motion (Einstein, 1956), in which the dot's direction and amplitude are randomly picked over each iteration.
- **Random direction**—when a dot is selected as noise, it will have a fixed motion direction which is drawn once for its entire life time, at birth.
- **Same**—when a dot is selected as noise, it remains so for its entire life time.
- **Different**—dots are selected as noise at each iteration.

These methods generate different spatio-temporal dynamics, visually and perceptively, and the coherence ratio—e.g., the percentage of dots that are associated to the signal—affects visual decoding with varying degrees of efficacy (Scase et al., 1996; Pilly and Seitz, 2009). In other words, the threshold for the correct detection of the signal's direction varies across RDK methods. Moreover, it has been reported that the smooth pursuit latency and early eye acceleration are not affected by the type of RDK, but late eye acceleration, pursuit gain and perceived velocity were dependent on RDK type (Schütz, D. I. Braun, Movshon, et al., 2010). Perception and pursuit performance also showed correlated dynamics. The authors interpreted their results as the pursuit system showing a capacity to integrate across directions of the RDK's signal, but not velocity.

B.2.2 Methods

Apparatus

The stimulus was displayed on a 36 cm by 27.5 cm (19 inches) Dell M993s CRT screen of resolution 1280 by 1024 pixels and a 75 Hz refresh rate, located 59 cm from the participants. Eye tracking was done using an EyeLink 1000+ (SR Research). Calibration was applied using a 5 points procedure between each block and if drift correction failed. Drift correction was applied after each trial.

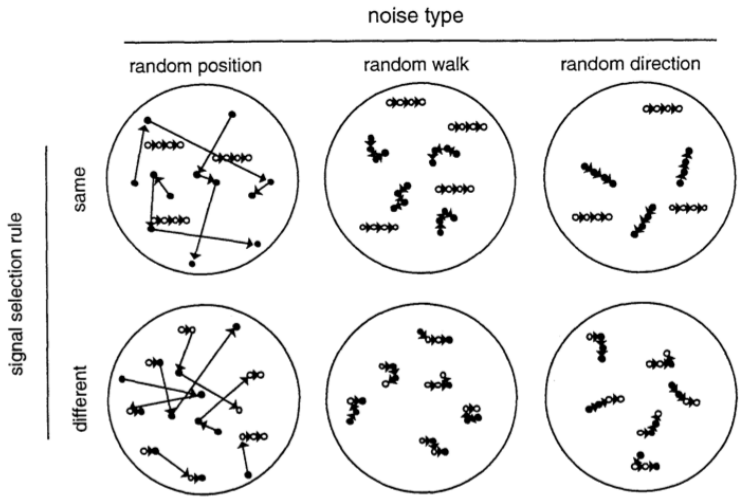


Figure B.2. Random-dot kinematogram. Schematic illustration of the six types of signal/noise display generated by the rules described in the text. The figure shows examples with 50% coherence and rightward signal motion in each case. Dots designated as signal dots for the following displacement are shown as open circles, those designated as noise dots are solid circles. In the random-position case, the displacement vectors shown join each noise dot to its new position selected by the plotting algorithm; for the visual system, these are not necessarily the most effective pairings for generating motion signals. Figure taken from Scase et al. (1996).

Stimulus

The moving plaid stimulus was presented in the same orientation setup as in the *Ambiguity* experiment (Section 4.3 in Chapter 4) with the coherent percept being perceived towards the top of the screen. In the central oculomotor fixation disk of a diameter of 4 deg, a RDK stimulus was implemented using the methods of **same random direction** (*see box on RDK for more details*) with the dots keeping the same label over their life time. 100 dots were generated with a life time of 40 s (see Fig. B.3), but they reappeared at new random initial location once they had left the display area of the RDK. The directions of each noise dot j were picked using a uniform random distribution over all angles: $\theta_j \sim \mathcal{U}(-\pi;\pi)$. All dots had the same velocity as the gratings of 1.5 deg.s^{-1}, perpendicular to the bars.

Protocol

Two levels of coherency (amplitude) for the RDK were used: 10% and 20% of dots being as signal. On the other hand, we varied the direction of the RDK signal by selecting amongst 8 possible levels: 0°, 60°, 90°, 120°, 180°, 240°, 270° and 300°

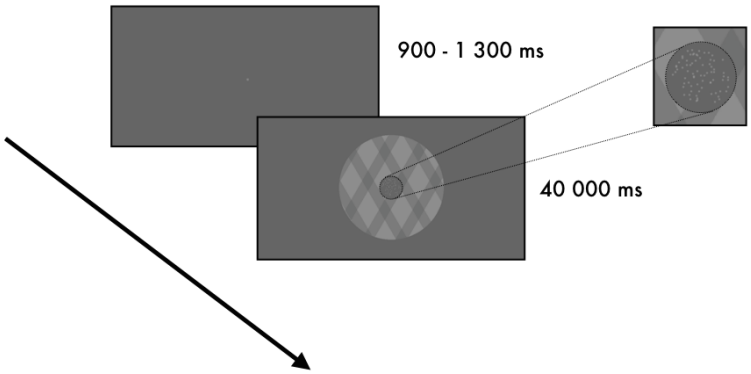


Figure B.3. Protocol. Experimental structure of a trial in the *Eye Movements* experiment.

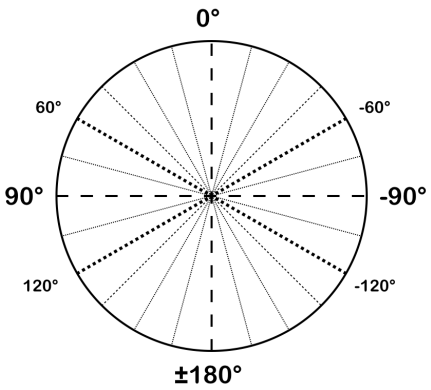


Figure B.4. RDK orientations. Diagram showing the orientations of the signals used for the RDK in order to verify the excitation and inhibition hypotheses of oculomotor dynamics on perception in the *Eye Movements* experiment.

(see Fig. B.4). There was one trial per combination of coherency level and RDK direction, and one control trial with no RDK and instead a simple dot (shown as 0% in the results). This summed up to a total of 18 trials per phase. However, two phases were carried out: one with the moving plaid (with a perceptual reporting task identical to the *Ambiguity experiment*) and RDK at the centre, the other with only the RDK and a grey empty space instead of the gratings. Phase order was selected randomly, and trials within each phase were shuffled randomly. Hence participants had to go through 36 trials in total, all with eye-tracking measurements.

Participants

10 individuals (3 males, 7 females; mean age of 26 years old and standard deviation of 8.52 years) participated in the experiment after signing declarations regard-

ing their consent and data anonymisation in accordance with the Declaration of Helsinki.

Eye movement analysis pre-processing.

In order to verify whether gaze was influenced as expected by the RDK manipulation of amplitude and direction, a method with the following steps of data processing was developed and implemented, based on the work on smooth pursuit analysis and perceptual inference in Aleshin et al. (2019). Data extraction was done using the *EDF mex* toolbox on *MATLAB* provided by SR Research Ltd, to extract data from EDF format to *MATLAB*. The data selected were samples recorded on both eyes at 1000 Hz on the EyeLink1000+. Raw gaze sample data beyond the screen size horizontally and vertically were replaced by not a number (NaN) entries such that the length of the data vector was left unchanged, but interruptions in the data time series were apparent.

Pursuit extraction.

We used the functions from the toolbox provided by Aleshin et al. (2019) to process the time series such that pursuits were extracted from the data—however, we did not generate cumulative smooth pursuits. The bi-variate signals for right and left eye signals were fed into a forward and a backward low-pass filter that applied convolutions with a 50 ms time kernel. The output signal was computed as the mean between both filters for each sample. Pursuit extraction was applied on the filtered data by selecting parts of the signal below a velocity threshold of 120 $deg.s^{-1}$ or an acceleration threshold of 471 $deg.s^{-2}$, and with durations longer than 50 ms. All parts of the signals above thresholds are replaced as NaN entries. To compute velocities, the removed parts of the vectors were obtained using a linear one-dimensional interpolation function.

The pursuit extraction step replaces eye blinks and saccades with NaN entries. We log these events in a table for signal division into *epochs*. Micro-saccades are detected, and logged in a table as well, using the EK algorithm—as presented in Chapter 2 (Engbert and Kliegl, 2003). This algorithm uses a relative thresholds based on 6 median absolute deviation of the eye velocity, a minimum duration of 3 ms and verifies that the micro-saccade is detected over both eyes.

Hence an *epochs table* was extracted with the epochs being the periods preceding micro-saccades, saccades or blinks.

Circular statistics were carried out on the polar representation of the velocity time series of the gaze using MATLAB's *CircStat* toolbox (Berens et al., 2009). For each epoch retained, we first tested whether velocity directions of gaze samples were uniform using a Rayleigh test, after smoothing the signal with a Savitzky-Golay filter (Press and Teukolsky, 1990) spanning over 50 ms. If the uniformity test was rejected at $\alpha = 0.05$, then we computed the mean velocity directions for left and right eye signals and verified that they did not differ by more than $\frac{\pi}{2}$. If velocity directions were conjugated, a one-sample circular t-test for the mean angle compared to the expected value based on the RDK direction was applied. The algorithm is described in **Algorithm** 1.

Algorithm 1: (Micro-)pursuit direction detection.

Result: Table of epochs classified as RDK (micro-)pursuits.
Extract left and right eye gaze signals as epochs $(\boldsymbol{g_L}, \boldsymbol{g_R})$;
for *each epoch* **do**
  if *Epoch duration* $\geq$ *100 ms* **then**
    Apply Savitzky-Golay filter with 50 ms span;
    Compute velocity directions $(\boldsymbol{\vartheta_L}, \boldsymbol{\vartheta_R}) = \arctan\left[(\boldsymbol{\dot{g}_L}, \boldsymbol{\dot{g}_R})\right]$;
    if *Rayleigh test null hypothesis is rejected at* $\alpha = 0.05$ **then**
      Compute velocity directions means $(\overline{\boldsymbol{\vartheta_L}}, \overline{\boldsymbol{\vartheta_R}})$;
      if *Absolute difference of left and right velocity direction means* $\|\overline{\boldsymbol{\vartheta_L}} - \overline{\boldsymbol{\vartheta_R}}\| \leq \frac{\pi}{2}$ **then**
        Compute conjugated velocity mean direction $\overline{\vartheta_{LR}}$;
        if *Conjugated velocity mean direction is in a 30° interval around the RDK directions* $\vartheta_{RDK} - \frac{\pi}{12} \leq \overline{\vartheta_{LR}} \leq \vartheta_{RDK} + \frac{\pi}{12}$ **then**
          Detection=true;
        else
          Detection=false;
        end
      else
        Detection=false;
      end
    else
      Detection=false;
    end
  else
    Detection=false;
  end
end

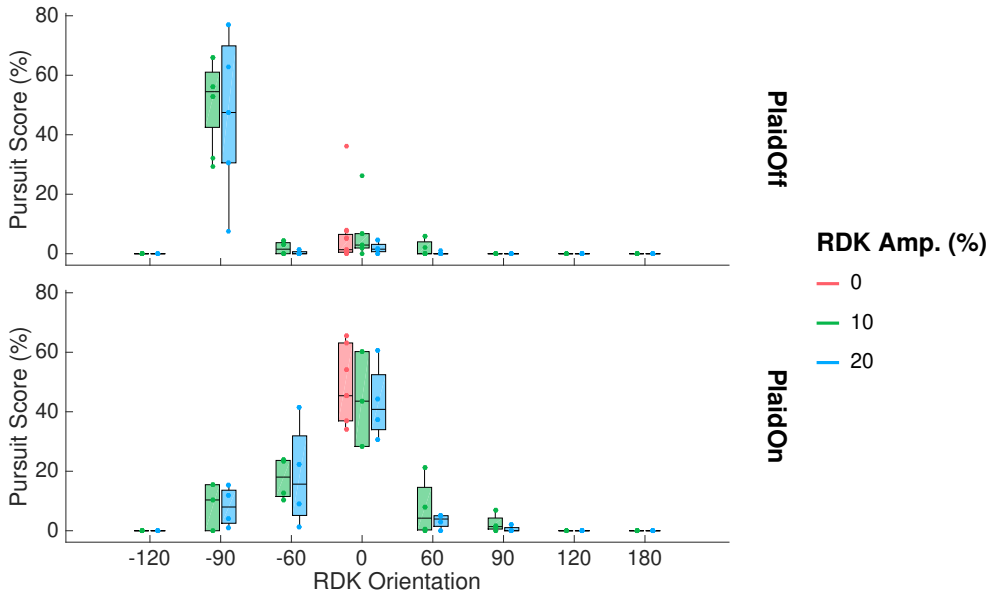


Figure B.5. Pursuit score results. Pursuit scores for all RDK orientations, amplitudes (colour) and in the Plaid On and Plaid Off phases. The control trials with no RDK had pursuit score computed against the 0° direction to measure the influence of the coherency percept motion direction.

Once the (micro-)pursuit direction detection was finished, a *pursuit score* was obtained by computing the ratio between pursuit duration and the sum of epochs durations in the *epochs table*. As this work was a pilot experiment and carried out on a small sample with many different conditions, no statistical analyses are presented and only a data description will be given.

B.2.3 Results

The data is visualised in Fig. B.5 where pursuit scores against RDK orientation are displayed, with RDK amplitude levels being discriminated by colours, for the phases with and without the moving plaid task. RDK seems to have generated few (micro-)pursuits in the phase without plaid task, except for the -90° trials in which participants made high pursuit scores. Meanwhile in its counterpart, we measured pursuits in some directions, namely 0°, -60° and 60°, when the plaid perceptual task was active. The effect appears to be irrelevant of RDK amplitudes across orientations and most potent at 0°. These measured pursuits thus occured at the directions coincidental with the perceived directions of the moving plaid percepts—i.e., 0° for coherency, -60° for right transparency and 60° for left transparency, as observed empirically in Section 4.2 in Chapter 4.

B.2.4 Discussion

Based on the descriptive data presented in Fig. B.5, we concluded that RDK manipulation, in the central fixation circle of the moving plaid stimulus was not efficient as no systematic pursuit were observed across orientation conditions. Moreover, the RDK amplitude (i.e. the percentage of dots having a determined direction) did not influence pursuit detection in both task conditions (Plaid On & Plaid Off).

A curious effect is observed at -90° in the Plaid Off phase, for both amplitude levels, where high pursuit scores were computed. As this direction coincides with the classic Latin language reading direction[2] and given the participants were french speaking and reading individuals, an oculomotor bias may exist that triggers (micro-)pursuits more easily in this direction.

In the Plaid On condition, we observed higher pursuit scores in the RDK direction orientations that matched the moving plaid's percepts—e.g., 0°, -60° and 60°. The RDK amplitude did not generated differences on the other hand, and moreover, the control trials where no RDK were displayed had similar pursuit scores to the 0° condition with Plaid On.

Therefore, RDK did not seem to create any oculomotor manipulation sufficiently potent to inhibit or excite the effects of perceived moving plaid directions. However, this data suggests that participants followed the moving plaid, providing encouraging evidence that verifies the hypotheses postulated in this thesis (Section 4.1 in Chapter 4). Namely, that perceived motion can be revealed by measuring (micro-)pursuits and interpreting them as active and embodied markers of ambiguous multi-stable perception.

The algorithm developed and used for this work has many restrictive criteria, i.e. many rejection outcomes, which could be a limitation of this work. Or at least, more signal processing methods – for instance approaches using MPC, or Quaternion Fourier transform (Flamant, 2018)—investigation could provide further insight on this data set. However, methods using the MPC measure, presented in Chapter 2, were not discriminant enough and threshold identification in the correlation space was not trivial to determine, whereas the threshold on velocity orientation used here were chosen through simpler hypotheses. Also, comparing the epochs' mean velocity values by computing gain in the Cartesian coordinate space yielded results with less interpretation potential. A promising future work perspective, for this work, would be to have the GraFEM model (presented in Chapter 3) generate a library of trajectories by varying its parameters across and specifying attractor motion, and

[2] As a reminder, our data has the 0° value at the top of the vertical axis of the polar circle.

to then compare and find the parameters set that show the most similarity with the observed data. We could then interpret the data based on the model's matched parameters and have further explanation of what occurs to generate these types of pursuits.

B.3 Noisy Motor Events experiment

B.3.1 Methods: no-report and no-fixation protocols with the moving plaid

Apparatus

The stimulus was displayed on a 36 cm by 27.5 cm (19 inches) Dell M993s CRT screen of resolution 1280 by 1024 pixels and a 75 Hz refresh rate, located 59 cm from the participants. Eye tracking was done using an EyeLink 1000+ (SR Research). Calibration was applied using a 5 points procedure between each block and if drift correction failed. Drift correction was applied between each trial.

Stimulus

The moving plaid stimulus was presented in the same setup as in the *Ambiguity* experiment with the coherent percept being perceived towards the top of the screen. Presentation time was of 40 seconds.

Protocol

The experiment was composed of two crossed conditions. On one side, we contrasted report *versus* no-report trials, and on the other, we had an oculomotor restriction task, i.e., fixing the central dot with the gaze, *versus* free oculomotor exploration of the stimulus. Each of these four conditions were organised in blocks of 12 trials with 2 trials being unambiguous. For each participants, the block order was sorted randomly. The report vs no-report contrast was meant to test the hypothesis on oculomotor markers of perception. The fixation vs free exploration contrast was meant to test the hypothesis on oculomotor marker amplification by manipulating the oculomotor task given to participants. Finally, non-ambiguous stimulus presentations were done in a few trials in order to obtain a contrast to investigate latencies between key press, eye movements and exogenous changes.

In non-ambiguous trials, the gratings' transparencies were dynamically changed based on exogenous percept durations drawn from a Gamma distribution with a

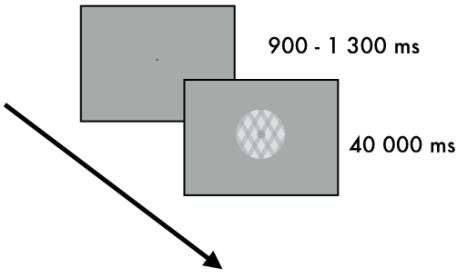


Figure B.6. Protocol. Diagram showing trial structure.

shape parameter $k = 1$ and a scale parameter $\theta = 5$ with the following density function:

$$f(x; k, \theta) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-\frac{x}{\theta}} \tag{B.5}$$

where $\Gamma(k)$ is the gamma function. The choice of which exogenous percept to display was done by flipping a coin between the other two percepts, for the following percept duration. Exogenous percepts were displayed as moving plaids with gratings' transparency values for coherency $(\alpha_L, \alpha_R) = (0.5, 0.5)$, for left transparency $(\alpha_L, \alpha_R) = (0.9, 0.1)$ and right transparency $(\alpha_L, \alpha_R) = (0.1, 0.9)$. Changes were done abruptly, from one frame to the other, when a percept duration was consumed.

Participants

19 individuals (17 women and 2 men, with an age mean of $\mu = 20.11$ and standard deviation of $\sigma = 1.14$) participated in the experiment after signing declarations regarding their consent and data anonymisation in accordance with the Declaration of Helsinki.

Eye movement analysis pre-processing

In order to verify whether the gaze was influenced as expected by the percepts and to detect to which percept belonged a part of the gaze time series, we used similar signal processing methods to that presented in Appendix B.2. Data extraction was done using the *EDF mex* toolbox on *MATLAB* provided by SR Research Ltd, to extract data from EDF format to *MATLAB*. The data selected were samples recorded on both eyes at 1000 Hz on the EyeLink1000+. Raw gaze sample data beyond the screen

size horizontally and vertically were replaced by NaN entries such that the length of the data vector was left unchanged, but interruptions in the data time series were apparent.

Pursuit extraction

We used the functions from the toolbox provided by Aleshin et al. (2019) to process the time series such that pursuits were extracted from the data—however, we did not create cumulative smooth pursuits. The bi-variate signals for right and left eye signals were fed into a forward and a backward low-pass filter that applied convolutions with a 50 ms time kernel. The output signal was computed as the mean between both filters for each sample. Pursuit extraction was applied on the filtered data by selecting parts of the signal below a velocity threshold of 120 deg.s^{-1} or an acceleration threshold of 471 deg.s^{-2}, and with durations longer than 50 ms. All parts of the signals above thresholds are replaced as NaN entries. To compute velocities, the removed parts of the vectors were obtained using a linear one-dimensional interpolation function.

The pursuit extraction step replaces eye blinks and saccades with NaN entries. We log these events in a table for signal division into *epochs*. Micro-saccades are detected, and logged in a table as well, using the EK algorithm—as presented in Chapter 2 (Engbert and Kliegl, 2003). This algorithm uses a relative velocity threshold of 6 standard deviations, a minimum duration of 3 ms and verifies that the micro-saccade is detected over both eyes.

Hence an *epochs table* was extracted with the epochs being the periods preceding micro-saccades, saccades or blinks.

Circular statistics were carried out on the polar representation of the velocity time series of the gaze using MATLAB's *CircStat* toolbox (Berens et al., 2009). For each epoch retained, we first tested whether velocity directions of gaze samples were uniform using a Rayleigh test, after smoothing the signal with a Savitzky-Golay filter spanning over 50 ms. If the uniformity test was rejected at $\alpha = 0.05$, then we computed the mean velocity directions for left and right eye signals and verified that they did not differ by more than $\frac{\pi}{2}$. If velocity directions were conjugated, a one-sample circular t-test for the mean angle compared to the expected value based on the possible percept direction ($\vartheta_l = 60°$, $\vartheta_c = 0°$ & $\vartheta_r = -60°$, based on

Percepts experiment results in Section 4.2) was applied. The algorithm is described in **Algorithm** 2.

Algorithm 2: (Micro-)pursuit direction detection.

Result: Table of epochs classified as perceptual (micro-)pursuits for percept inference.

Extract left and right eye gaze signals as epochs $(\boldsymbol{g_L}, \boldsymbol{g_R})$;
for *each epoch* **do**
  if *Epoch duration* $\geq$ *100 ms* **then**
    Apply Savitzky-Golay filter with 50 ms span;
    Compute velocity directions $(\boldsymbol{\vartheta_L}, \boldsymbol{\vartheta_R}) = \arctan\left[(\dot{\boldsymbol{g}}_{\boldsymbol{L}}, \dot{\boldsymbol{g}}_{\boldsymbol{R}})\right]$;
    if *Rayleigh test null hypothesis is rejected at* $\alpha = 0.05$ **then**
      Compute velocity directions means $(\overline{\boldsymbol{\vartheta_L}}, \overline{\boldsymbol{\vartheta_R}})$;
      if *Absolute difference of left and right velocity direction means* $\|\overline{\boldsymbol{\vartheta_L}} - \overline{\boldsymbol{\vartheta_R}}\| \leq \frac{\pi}{2}$ **then**
        Compute conjugated velocity mean direction $\overline{\vartheta_{LR}}$;
        if *Conjugated velocity mean direction is in a 30° interval around the percept directions* $\vartheta_{percept} - \frac{\pi}{12} \geq \overline{\vartheta_{LR}} \leq \vartheta_{percept} + \frac{\pi}{12}$ **then**
          Detection=true;
        else
          Detection=false;
        end
      else
        Detection=false;
      end
    else
      Detection=false;
    end
  else
    Detection=false;
  end
end

Inferred percept epoch table simplification

The epochs table with percepts inferred from eye movements was simplified (3) by (i) merging short identical and consecutive detected percepts, (ii) fusing percepts when the anterior and current match, (iii) identifying potential long percepts and

overriding smaller ones and (iv) taking a decision on whether to leave remaining rejected epochs as such or to integrate them in inferred percepts. The merger and fusion gives priority to the preceding percept when repetitions are found. Afterwards, the algorithm applies a fusion of repeated percepts in consecutive epochs, thus looking for long percepts. Finally, the algorithm looks at un-classified epochs and takes a decision based on duration, with short epochs (≤ 1000 ms) being fused and the others being kept as periods without percepts.

Percept inference scores

The percept inference process was evaluated by comparing the reported key press data to the inferred eye movement data. To do so, we computed the percentage of time in a reported key press percept, during which, the inferred eye movement percept matched. Scores were then averaged over trials for each percept. We also looked at the non ambiguous condition, during which exogenous, reported key press and eye movement inferred perceptual changes were compared. To do so, we estimated, using a maximum likelihood estimator, the parameters of Gamma and Log-Normal distribution functions in each trial for the exogenous, key press reported and eye movement inferred percept durations.

B.3.2 Preliminary results

As a reminder, the results presented here remain preliminary and further analyses are needed for a better understanding of the data.

Percept inference scores in the condition in which participants reported their percepts with key press were computed by obtaining the percentage of time in a key press percept that was correctly inferred from eye movements by the procedure described above. The results presented in Fig. B.7 show that overall, in the ambiguity contrast and the oculomotor task contrast, similar patterns are visible in the data. In particular, coherent percept detection was correctly detected more often than the other transparency percepts. In fact coherency scores seem to be above chance level—25% as percepts could be either c, l, r or none, though these levels are not necessarily uniform as they depend on the probability of percept observation—but the algorithms' detection seems to agree less with the reported key press data, for the transparency percepts.

We compared estimated Gamma and Log-Normal distribution parameters based on the different data types that could be analysed in this experiment: eye movements

Algorithm 3: Gaze-based percept timeline inference simplification.

Result: Table of simplified epochs classified as perceptual timeline inference based on gaze.

Provide a table of oculomotor epochs with associated perceptual detection;
for *each epoch* $\boldsymbol{e}(i) = (e_p(i), e_d(i))$ *with a percept* $p = (l, c, r, n)$ *&* *duration* d *(ms), respectively* **do**
  if *Percepts are the same for* $e_p(i-1), e_p(i), e_p(i+1)$ **then**
    if $e_d(i-1) > e_d(i) \wedge e_d(i+1) > e_d(i)$ **then**
      Merge: $e_p(i) = e_p(i-1)$;
    end
    if $e_d(i-1) > e_d(i) \wedge e_d(i+1) \leq e_d(i)$ **then**
      Merge: $e_p(i) = e_p(i-1)$;
    end
    if $e_d(i-1) \leq e_d(i) \wedge e_d(i+1) > e_d(i)$ **then**
      Merge: $e_p(i) = e_p(i+1)$;
    end
    if $e_d(i-1) \leq e_d(i) \wedge e_d(i+1) \leq e_d(i)$ **then**
      Do not merge;
    end
  end
  if *Only anterior & current percepts are the same* $e_p(i-1), e_p(i)$ **then**
    Fusion: $e_p(i) = e_p(i-1)$; $e_d(i-1) = e_d(i-1) + e_d(i)$;
    Remove $\boldsymbol{e}(i)$;
  end
end
while *There remain consecutive repeated percepts for epochs* **do**
  For each percept $p = (l, c, r, n)$, apply fusion on repeated epochs $\boldsymbol{e}$ into the first of the same percept;
end
for *Remaining un-classified epochs* $e_p(i) = n$ **do**
  if $e_d(i) \leq 1000$ *ms* **then**
    Fusion with $\boldsymbol{e}(i-1)$;
  else
    Keep epoch with $e_p(i) = n$;
  end
end

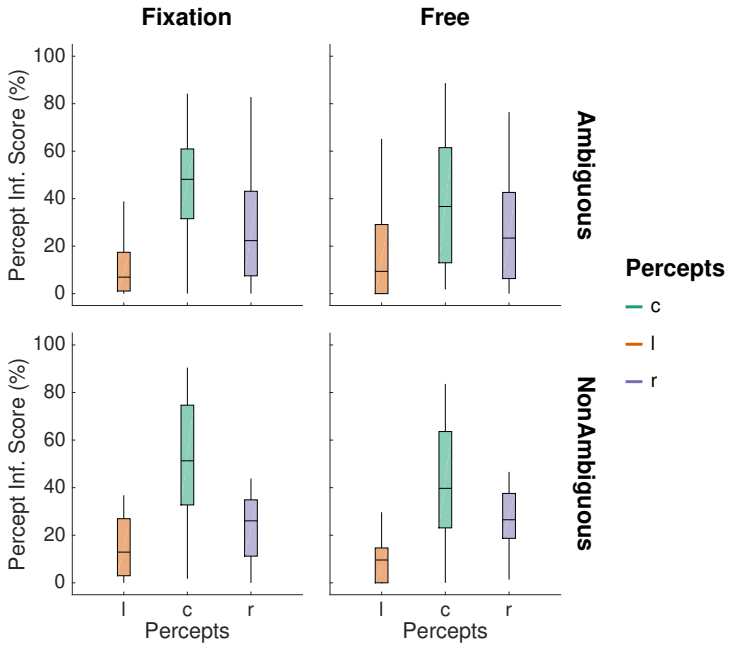


Figure B.7. **Percept inference.** Box plots of the computed percept inference scores, for each percept: coherency c, left transparency l and right transparency r. The fixation task and ambiguity contrast are presented and only data from the trials with key press report are presented. Note that the box plots for Ambiguous trials are composed of 557 data points, while the Non-Ambiguous 144 data points.

inferred (EM), exogenous changes (EX) in non ambiguous trials and key press reported (KP) in report trials. The trial-level data for both functions are shown in Fig. B.8. In both cases, the estimated parameters form clusters and do not generate similar distributions across data types. This approach shows that the data types provide very different perceptual information, with the key press data showcasing shorter percept durations while the eye movement inferred longer ones, based on the k and μ estimated parameters. It also shows that the ambiguity manipulation in this experiment did not reproduce patterns similar to key press reports.

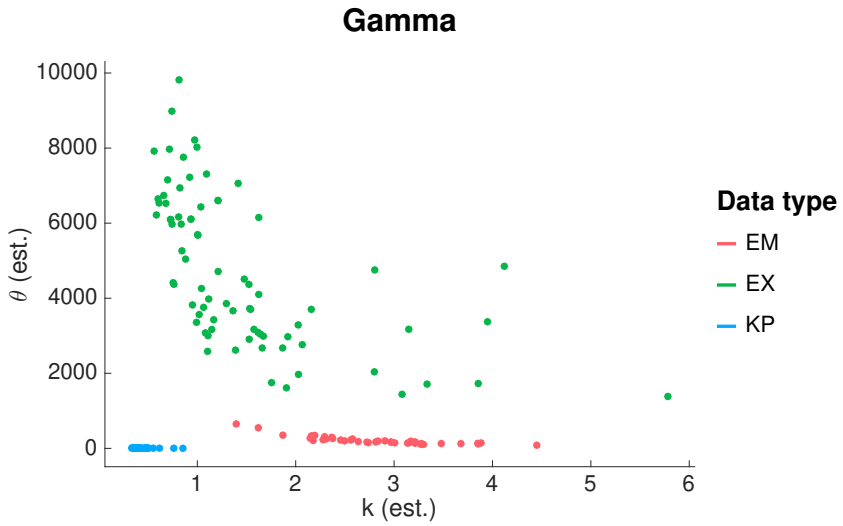
B.3.3 Discussing preliminary results

In this experiment, we saw a number of exploratory manipulations—e.g., ambiguity, oculomotor task, and perceptual report task—and though the results presented above remain preliminary, they provide interesting insights, useful to the objective of creating a paradigm where perceptual and (oculo-)motor systems can be studied jointly, as presented in Section 4.1 in Chapter 4. Some of these points are discussed in the following paragraphs.

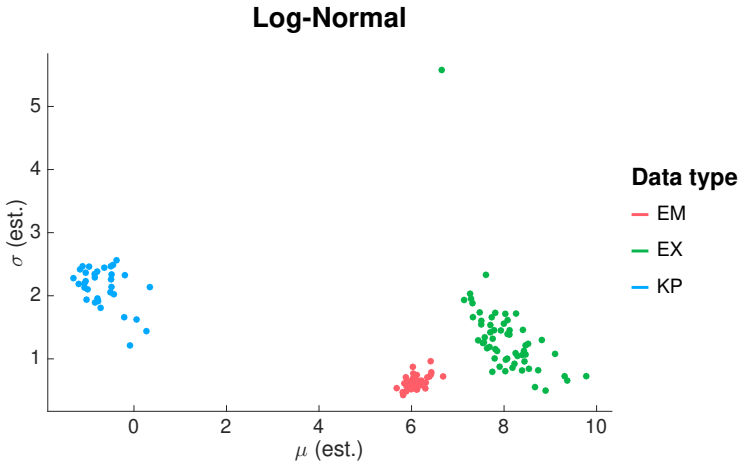
Limitations of EM percept detection: towards a better oculomotor task?

This experiment's data shows that percept detection in eye movement gaze signals is challenging. Though we based our methods for gaze processing on the one developed in Appendix B.2, the reference here were the dynamics percepts, whereas in the former, the reference was a velocity direction, constant across trial duration. The inference process was thus subject to sequencing error and temporal shifts, as key press reports are notoriously late with latencies of approximately 400 ms, whereas eye movements show much faster responses (Kornmeier and Bach, 2012; Aleshin et al., 2019). Hence, the score computation is also less trivial than in Appendix B.2 and generates three scores per trial: one per percept. For instance, some participants, at times, do not report all three percepts in a trial, which tends to negatively bias the score.

Furthermore, the results show that the *free exploration* oculomotor task was not effective as scores did not improve; in fact they are further dispersed for the coherency in the ambiguous free trials—see Fig. B.7. The reduced overall performance on the transparency percepts can be explained by the change of perceived moving objects' size, leading to eye movements in more directions. Indeed, in coherency, one sees diamond patterns moving upwards, creating smaller perceived objects to track. The



(a) Estimated Gamma distributions.



(b) Estimated Log-Normal distribution.

Figure B.8. Percepts data comparison.
(a) Estimated parameter space (k, θ) for Gamma distributions function on eye movement inferred (EM), exogenous changes (EX) and key press reports (KP) data sets (colours).
(b) Estimated parameter space (μ, σ) for Log-Normal distributions function on eye movement inferred (EM), exogenous changes (EX) and key press reports (KP) data sets (colours).

speed integration also make the motion faster, and potentially more adequate for the pursuit system. In transparency, the entire un-occluded gratings are perceived to move; participants may move along the tracked bars. Hence, this makes our detection algorithm ineffective as it searches for epochs with a velocity direction in an interval surrounding the percept's direction.

Therefore, a perspective would be to test the protocol with a more adequate oculomotor task: "*follow the perceived motion*". This would be a more adequate first step to gather the needed data to develop a robust percept inference algorithm.

Limitations of brutal ambiguity manipulation

Another limitation of this work resides in the method used to manipulate the stimulus for non ambiguous trials. Because of time constraints and the internship context, the psycho-observer model developed in Chapter 4 was not integrated and a simple, brutal approach was used. The spread in distribution parameters shown in Fig. B.8 provides evidence that the use of an arbitrary Gamma distribution, to generate exogenous percept changes, has critical limitations in reproducing key press reports dynamics. Thus, this data set, and the presented analysis, provides further motivation to estimate an observer's probabilistic model of the moving plaid ambiguity with respect to the transparency parameters. This can allow to infer and identify points of maximal ambiguity and furthermore, to use the points where the model's probabilities are highest for a percept to bias the stimulus towards that percept.

Journal of Vision article

C

The following pages present the corrected and revised results from Chapter 2 as they appear in the accepted *Journal of Vision* article. The modification followed a peer review exchange by the article's reviewers and comments by the examiner Laurent Madelain, during the defence.

Micro-pursuit: a class of fixational eye movements correlating with smooth, predictable, small-scale target trajectories

Kevin Parisot Univ. Grenoble Alpes, CNRS, Grenoble INP*, GIPSA-Lab, 38000 Grenoble, France
*Institute of Engineering Univ. Grenoble Alpes

Steeve Zozor Univ. Grenoble Alpes, CNRS, Grenoble INP*, GIPSA-Lab, 38000 Grenoble, France
*Institute of Engineering Univ. Grenoble Alpes

Anne Guérin-Dugué Univ. Grenoble Alpes, CNRS, Grenoble INP*, GIPSA-Lab, 38000 Grenoble, France
*Institute of Engineering Univ. Grenoble Alpes

Ronald Phlypo Univ. Grenoble Alpes, CNRS, Grenoble INP*, GIPSA-Lab, 38000 Grenoble, France
*Institute of Engineering Univ. Grenoble Alpes

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

Humans generate ocular pursuit movements when a moving target is tracked throughout the visual field. In this article, we show that pursuit can be generated and measured at small amplitudes, at the scale of fixational eye movements and tag these eye movements as *micro-pursuits*. During micro-pursuits, gaze direction correlates with a target's smooth, predictable target trajectory. We measure similarity between gaze and target trajectories using a so-called *maximally projected correlation*, and provide results in three experimental data sets. A first observation of micro-pursuit is provided in an implicit pursuit task, where observers were tasked to maintain their gaze fixed on a static cross at the center of screen, while reporting changes in perception of an ambiguous, moving (Necker) cube. We then provide two experimental paradigms and their corresponding data sets: a first replicating micro-pursuits in an explicit pursuit task, where observers had to follow a moving fixation cross (Cross), and a second with an unambiguous square (Square). Individual and group analyses provide evidence that micro-pursuits exist in both the Necker and Cross experiments, but not in the Square experiment. The inter-experiment analysis results suggest that the manipulation of stimulus target motion, task and/or the nature of the stimulus may play a role in the generation of micro-pursuits.

Keywords: **fixational eye movements, micro-pursuits, micro-saccades, maximally projected correlation, attractor-based model**

doi: Received: December 30, 2020 ISSN 1534–7362 © 20?? ARVO

Introduction

Eye movements are typically classified at macroscopic scale as either fixation, pursuit, saccade, or reflexive eye movements. But, even during fixations, eyes never stay still and a variety of fixational eye movements have been observed and studied (Martinez-Conde, Macknik, & Hubel, 2004). As an example, micro-saccades have been defined as small amplitude, ballistic movements, similar to large scale saccades (Rolfs, 2009). Based on the hypothesis that eye movements are consistent observations in an oculomotor continuum (Otero-Millan, Macknik, Langston, & Martinez-Conde, 2013), and in line with micro-saccades, one can thus expect to observe small-amplitude pursuits within fixations. Here, we will focus on this subclass of slow fixational eye movements, which we term micro-pursuit eye movements. We provide evidence of micro-pursuit eye movements at a fixation level, with an adapted metric that reveals their existence. Three different experiments are presented, two where micro-pursuit occurs and one where it does not. In what follows, we will first describe the current classes of macro-scale eye movements, with their functions and metrics, to provide a starting point for the oculomotor continuum hypothesis that we defend.

The main function of eye movements is to orient the gaze towards parts of a visual scene (Yarbus, 1967; Palmer, 1999; Liversedge, Gilchrist, & Everling, 2011). To accomplish this goal, the human oculomotor system has the capacity to generate a wide variety of movements that can be categorized based on their spatio-temporal dynamics: amplitude, velocity, and acceleration.

Rapid and ballistic eye movements (saccades): classified based on displacement, speed, and acceleration thresholds, e.g., displacement above 0.15 degrees (deg), velocity above 30 deg.s^{-1}, acceleration above 9500 deg.s^{-2}, though other detection criteria exist (Nyström & Holmqvist, 2010; Behrens, MacKeben, & Schröder-Preikschat, 2010; Mihali, Opheusden, & Ma, 2017). These criteria have become their definition. But, absolute threshold criteria have been criticized for their lack of functional, physiological or formal justifications. For example: the clear dichotomy between fixations and saccades has been loosened (Ko, Poletti, & Rucci, 2010).

Slow eye movements (smooth eye pursuits, slow oculomotor control): classified based on a simple velocity criterion, e.g., smooth pursuit ranges from 20 to 90 or 20 to 100 deg.s^{-1} (Krauzlis, 2004; Komogortsev & Karpov, 2013; Spering & Montagnini, 2011), though pursuits are considered smooth and precise only at speeds up to 30 deg.s^{-1}. If target velocity is too high for the pursuit system, catch-up saccades can compensate for the accumulated position error created by the difference between target and gaze velocities, also known as the retinal slip (De Brouwer, Yuksel, Blohm, Missal, & Lefèvre, 2002).

Eye fixations: usually defined as any eye movement with an amplitude below 1 deg. They specifically include fixational eye movements which form a generic class of small-amplitude eye movements (ocular drift, tremor and micro-saccades) sharing dynamic characteristics with regular (macro) eye-movements at smaller scale (Otero-Millan et al., 2013; Krauzlis, Goffart, & Hafed, 2017).

The article is organized as follows: First, slow eye movements are described with associated with their dimension and metrics. Secondly, small-amplitude, slow eye movements and their dependencies on the visual stimulation, the task, and the experimental paradigm are detailed as well as the metrics used for their detection. Then, we introduce a metric for target-dependent eye movement, maximally projected correlation (MPC), a scale- and translation-invariant metric that measures similarity between the gaze and a target 2D motion during small amplitude smooth movement. Finally, we propose three experiments and their results: a first experiment (Necker) that allows for the detection of micro-pursuit and two other experiments (Square and Cross) that have been built to replicate the generation of smooth pursuit with different stimuli and tasks.

Slow eye movements: different kinds of motion

The functional role of (smooth) pursuit is to maintain a—usually moving—target of interest on the high acuity foveal region of the retina (Spering & Montagnini, 2011). Tracking is believed to be controlled by retinal errors, the difference between gaze and target positions, or retinal slip[1], i.e., $\boldsymbol{q}_R \doteq \boldsymbol{q}_G - \boldsymbol{q}_S$, the difference between gaze and target velocities or speed vectors of the gaze and of the target stimulus, i.e., $\dot{\boldsymbol{q}}_R \doteq \dot{\boldsymbol{q}}_G - \dot{\boldsymbol{q}}_S$. According to Orban de Xivry and colleagues (Xivry & Lefevre, 2007), pursuit relies mostly on reducing retinal slip and is modulated, in a smaller way, by position and acceleration errors.

In order to detect and measure the quality of slow eye movements, metrics have been defined that associate gaze with the target stimulus position. For smooth pursuit, tracking quality is measured through *gain* (see **Micro-pursuits** section for more details). This measure has shown its effectiveness in experimental protocols where a target appears on screen and participants are tasked to follow its motion. Pursuit is mostly studied for tracking a single point on a uniform background, although other stimuli in motion also lead to pursuit movements, for instance, random-dot kinematograms (Heinen & Watamaniuk, 1998), line figures (Masson & Stone, 2002), illusory perceptual motion (Madelain & Krauzlis, 2003), or after-effect motion (D. I. Braun, Pracejus, & Gegenfurtner, 2006). In tasks where a percept is pursued, rather than a stimulus, the measure of gain and the associated models have been questioned (Stone, Beutter, & Lorenceau, 2000).

Among the slow eye movements, we also find reflexive movements such as the vestibulo-ocular reflex (VOR), the oculo-

[1] For the use of the notations in this manuscript the reader is referred to Appendix A.

following reflex (OFR), or the opto-kinetic nystagmus (OKN). The VOR is a reflex eye movement that compensates head motion in order to maintain a stable retinal image. Though the VOR expression may be similar to pursuit, it is only generated when the head is free to move. The OFR is a reflexive eye movement in response to a sudden change of a wide-field image (Michalski, Kossut, & Żernicki, 1977; Miles, Kawano, & Optican, 1986; Gellman, Carl, & Miles, 1990; Quaia, Sheliga, FitzGibbon, & Optican, 2012). The reflex is mainly attributed to the tracking of motion in peripheral vision (Ilg, 1997). The OKN is a composite gaze pattern in which an object is followed by smooth pursuit until the object leaves the visual field. At this point, the gaze returns to the object's initial position (fast saccadic response) at the starting position of the pursuit. VOR, OFR and OKN are eye movements solicited in specific visual stimulation and experimental contexts, which require the manipulation of a large part of the visual field, not a smaller perceptual target, as with pursuit.

To summarize, pursuits have been studied as large-scale eye movements with amplitudes exceeding 1 deg (60 min-arc) in which a target with motion is tracked by the gaze, such that the retinal slip is minimized. The metric used to measure pursuit has been velocity gain.

Do small amplitude pursuits exist?

Fixational eye movements

We have just described the three principal classes of macroscopic eye movements, where saccades and pursuits are distinguished from fixations based on the amplitudes and velocities involved. However, the fact that during the fixation the eye never stands still (Ditchburn & Ginsborg, 1953) and continuously produces fixational eye movements further subdivides *fixations* into the following sub-classes (Kowler, 2011): **Micro-saccades**, are ballistic small amplitude and fast gaze shifts (Rolfs, 2009; Poletti & Rucci, 2016). **Slow drifts** are small velocity ($<$ 0.5 deg.s^{-1}) displacements of the gaze (Nachmias, 1961; Yarbus, 1967) and **tremors (or physiological nystagmus)** are aperiodic high-frequency oscillations of the eye (30-80 Hz and amplitudes of up to 50 seconds of arc) (Nachmias, 1961; Martinez-Conde et al., 2004). Research has also been conducted on tremor, but due to their small amplitude and high frequency it is impossible to distinguish them from noise using video-based eye-trackers (Ko, Snodderly, & Poletti, 2016). Therefore, tremors will not be considered in our study. The class of slow drifts, and more particular small-amplitude pursuits, seems less covered in the literature, which can be explained by the technical difficulties associated with eye-tracker precision, especially video-based ones, at such small scales (Wyatt, 2010; Choe, Blake, & Lee, 2016). As we want to focus on the latter, we will give a detailed review of literature on slow drifts small-amplitude movements.

Micro-saccades

Micro-saccade is a class of fixational eye movements characterized by (i) ballistic properties—like saccades,—(ii) small amplitudes, and (iii) a linear relationship between peak velocity and amplitude, also known as a main sequence (Bahill, Clark, & Stark, 1975). The latter stipulates that as micro-saccades have larger amplitudes, their associated (measured) peak velocity increases, and this relationship is linear. In essence, the fast, ballistic nature of micro-saccades allow to quickly—typically under 80 ms—re-position the fovea in the context of visual perception (Rolfs, 2009; Ko et al., 2010; Poletti & Rucci, 2016; Sinn & Engbert, 2016), similar to saccades at larger scales (i.e., not contained within fixational eye movements). Physical properties of the oculomotor system constrain these ballistic movements of the eye to exhibit the linear peak velocity–amplitude relationship.

The main sequence has been very reproducible, and appears in over decades of eye movement research (Rolfs, 2009; Hicheur, Zozor, Campagne, & Chauvin, 2013). Other than providing insight into the oculomotor control system's properties (Bahill et al., 1975) it also supports the hypothesis of an oculomotor continuum (Rolfs, Kliegl, & Engbert, 2008; Sinn & Engbert, 2016). In Engbert and Kliegl (Engbert & Kliegl, 2003), detection of micro-saccades is based on a lower velocity threshold computed relatively to the overall velocities in an observation window. As such, the detection threshold is dependent on the contextual oculomotor activity. This is combined with a binocularity criterion to avoid spurious detections. This is also the approach we have followed in this work.

Ocular drift: a simple random process or stimulus-dependent?

These slow and small movements are the consequence of a slow control system of eye position (Cunitz, 1970) described in literature as a mere drift of the eye (Dodge, 1907), OFR (Chen & Hafed, 2013), or—more recently—as small amplitude pursuits (Skinner, Buonocore, & Hafed, 2018).

In early studies of fixational eye movements, when subjects had to fixate a static dot, eyes drifted slowly with an upper velocity limit at 0.5 deg.s^{-1} and mean velocity of 5 min-arc.s^{-1} (Yarbus, 1967). Their trajectories were considered as random and involuntary processes since they showed dynamics similar to Brownian random walks (Ratliff & Riggs, 1950; Engbert & Kliegl, 2004) as well as independence between the two eyes (Cornsweet, 1956). However, Ditchburn and Ginsborg's work (Ditchburn & Ginsborg, 1953) provided evidence that direction of eye movement is not completely random during drift; it is idiosyncratic. Nachmias (Nachmias, 1961) replicated this finding in an experiment where a fixation target was switched on and off during 3 seconds cycles. He found that each of the 2 subjects have preferred drifting direction but this preferred direction can be modified by changing the visual environment. The author interpreted the idiosyncratic direction preference as specific to muscular response and reasserted that nonrandom ocular drifts occur in fixations while providing evidence that drift direction can be modulated by the visual environment. More recently, a variety of experiments have shown that drift can take properties and charac-

teristics close to other known oculomotor phenomena (Poletti, Listorti, & Rucci, 2010; Chen & Hafed, 2013; Skinner et al., 2018; M. Watanabe et al., 2019).

As mentioned, drift can be viewed as part of a slow control system, enabling gaze to capture a target, whether static or dynamic. Here, we will discuss two studies that show evidence of slow eye movements correlating with the target stimulus, and as such related to our proposition of adding a subclass to the fixational eye movements: that of micro-pursuits.

Chen and Hafed (Chen & Hafed, 2013) studied the impact of micro-saccades on visual perception and investigated the relationship between micro-saccades and drift. Their experiment contained two major tasks. The first task required two monkeys to stare at a fixation dot where a change in luminance of the dot or a peripheral white flash was introduced to induce a higher probability of micro-saccade generation. Drift velocity was analyzed before and after the micro-saccades using either direct velocity measurements or spatial dispersion (by spatial binning and box counts). Both measures showed an increase in drift velocity post-micro-saccadic movements with respect to pre-micro-saccadic movements or baseline movements. They also showed that eye drift mainly occurs in the direction opposite to the micro-saccade, which is interpreted as corrective slow control of the gaze position. The second task consisted of a sinusoidal grating that started moving at predefined delays after the onset of a micro-saccade (or after 500 ms if no micro-saccade was detected). The authors analyzed the speed and direction of early drift of the eye, namely the OFR, according to the direction of the grating and the time of grating onset based on micro-saccade detection. Indeed, they reported that (i) the drift directions were in the opposite directions of the micro-saccades and (ii) the eye velocity was reduced when the grating's motion was initiated during micro-saccade and was enhanced when the motion was initiated after micro-saccade. The OFR being an indicator of "*the sensitivity of early motion processing to retinal-image slip after a micro-saccade*", the OFR, and thus motion perception, are suppressed during the saccade and enhanced after. Their overall findings suggest that there is a single slow gaze control system that control both fixation and eye movement position in the presence of a fixed target or a slow moving background linked to the motion perception system. Conclusions suggesting a subtle coupling between micro-saccades and drifts are also reinforce by previous reports (Engbert & Mergenthaler, 2006).

Part of this idea had already put forward by Murphy and colleagues (Murphy, Kowler, & Steinman, 1975). In their experiment, they asked participants to maintain their gaze on a present or absent fixation dot while a grating in the background moved horizontally at velocity ranging from 0.08 deg.s^{-1} to 8 deg.s^{-1}. In a second condition, the participants had to follow the grating. Eye movement velocities were analyzed for trials without saccades. The study shows that when participants have to stare at the fixation dot (i) they have an ability to keep gaze fixed when the fixation dot was present, and (ii) an OFR——a smooth displacement of the eye in the direction of the grating's movement but with smaller velocities—is detected when the fixation dot was absent. In contrast, when the task was to follow the grating, participants showed clear smooth, slow movement in the direction of motion with velocity as low as 0.08 deg.s^{-1}.

Both these studies confirm the existence of a slow movement within a fixation that track a slow velocity target or counteract the displacement of a micro-saccade. These slow movement of pursuit or fixation stabilization are thought to be under a same slow control system, although the tracking mechanism seems not to be triggered when the movement is initiated during a micro-saccade.

Ocular drift and slow motor control

Drift has been linked to slow control of the eyes during fixation in the context of investigating links between visual stimulation and drift motion.

In a series of experiments, Kowler and Steinman (Kowler & Steinman, 1979a, 1979b) have investigated how expectation, over a stimulus and task, can induce anticipatory smooth and slow eye movements. The authors implemented a task in which participants had to track a dot moving by steps (with three frequencies: 0.25, 0.375 or 0.5 Hz) along a horizontal segment of 3.3 deg amplitude. They showed that eye movements' direction and latency depend on predictability of target displacement. Furthermore, they showed this effect to remain even when the level of predictability was manipulated and when a distracting secondary task was imposed (Kowler & Steinman, 1981). In fact, they provided evidence that anticipatory eye movements—which they also named involuntary drifts in the direction of future target motion—depended on the history of prior target motions (Kowler, Martins, & Pavel, 1984). To understand whether the slow control of ocular drift is driven by position or velocity signals, they carried out an experiment in which they manipulated drift by changing the configuration of reference points, thus varying the difficulty of fixation of a central point (Epelboim & Kowler, 1993). Their analyses used gaze position data and bivariate contour ellipse area (BCEA) computation for quantification of gaze dispersion. As such, they provided evidence that the oculomotor system does not rely on visual position signals, but rather on retinal image slip, in order to implement slow motor control. This creates a parallel with the known models for smooth eye pursuit described above.

In addition, in a recent paper, Watanabe and colleagues (M. Watanabe et al., 2019) reported a study that links ocular drift, micro-saccades, and pupil area on voluntary eye movements preparation. They observed anticipatory drifts prior to stimulus appearance and they argue that these anticipatory eye movement may reflect volitional action preparation. Interestingly, the authors provide a replication of previous results on anticipatory drift with a video-based eye tracker while applying correction to their gaze signals for pupil deformation.

Overall, these studies show that slow eye movements are present during fixation. These movements can control for a fixation position, can track large target and depend on expectation.

Authors have postulated that all these behaviors are under control of a unique system.

Small amplitude pursuits

As mentioned higher, smooth pursuits are large-scale eye movements with amplitudes exceeding 1 deg (60 min-arc). A small set of studies found eye movements within a fixation that share characteristics with smooth pursuits, except for their amplitude. Though there are references to smooth pursuits of small amplitude as far as in Yarbus' book (Yarbus, 1967), most papers in the literature have reported the phenomenon in an indirect manner.

In a study on drift in the absence of visual stimulation or with afterimages, horizontal smooth drifts were reported (Heywood & Churcher, 1971). Although their description corresponds to pursuit dynamics, they did not define the observed movements as such. The authors published a follow-up paper showing that, depending on the eccentricity of the afterimage, oculomotor dynamics are more or less smooth and show low velocities, hence could be interpreted as pursuits (Heywood & Churcher, 1972). Further, while attempting to study oculomotor control capacities when presenting a moving grating background with a fixation point, Murphy and colleagues (Murphy et al., 1975) reported eye movements that correspond to small amplitude pursuits. When investigating the lack of compensation of the VOR when the head was free, Martins, Kowler and Palmer (Martins, Kowler, & Palmer, 1985) studied whether a smooth pursuit system might interact with the VOR. Their data provided a qualitative description that small amplitude pursuits are related to the velocity of target motion. The following finding was reported: foremost, the effectiveness of smooth pursuits varied with target velocities. At the lowest average velocities of a tracked point[2] (0.0025-0.125 deg.s^{-1}), smooth pursuit was the most effective, i.e., retinal-image speed during smooth pursuit was about the same as retinal-image speed during low target velocities. At higher target velocities (0.25-1 deg.s^{-1}), smooth pursuit was less effective for retinal image stabilization and at the highest velocities (1.5-2.5 deg.s^{-1}), smooth pursuit was totally ineffective.

More recently, small amplitude pursuits have been reported again, in very different contexts. In a study of eye drift and its relationship to retinal image motion—investigating whether the latter drives the former through retinal or extra-retinal information—Poletti and colleagues (Poletti et al., 2010) declared the following observation: "*small pursuit-like eye movement with amplitudes comparable to those of fixational drifts are under precise control of the oculomotor system*". Finally, a precise characterization of rhesus macaque oculomotor control for rectilinear sinusoidal motion of a target with amplitudes inferior to 0.5 deg and velocities below 2.5 deg.s^{-1} was recently reported (Skinner et al., 2018). The amplitude and frequency of the sinusoidal motion was modulated and gaze signals were analyzed using gain and compared to filter responses; filters are, here, used as models to show how the oculomotor system could display different behaviors based on input frequencies—on gaze position and velocity. Furthermore, they showed that the gaze signals had eye velocity spectrum with peaks at target frequency and that pursuit gain was highest at 1 deg.s^{-1}.

Overall, *pursuits* have been observed for a range of velocities (0.05-2 deg.s^{-1}) and amplitudes (1.9-30 min-of-arc) which qualifies them as fixational eye movements. Given the classification in the fixational eye movements research field—in which only micro-saccades, drifts, and tremors are considered—these observations raise questions on the nature and potential definition of micro-pursuits or fixational pursuits.

This article focuses on the presentation of micro-pursuits in three contexts: (i) presentation of metrics that fit the theoretical requirements to detect micro-pursuit, (ii) detection of the oculomotor phenomenon in (a) a dual task experiment (Necker) in which its elicitation was not explicitly made to participants, and (b) an explicit tracking experiment (Cross) and an implicit distractor setup (Square). Our hypothesis was that if the perceptual system has to detect a change in a moving stimulus with a predictable trajectory, the oculomotor system is likely to follow the target even if the participant is instructed not to do so (fixation task). But, since the fixation task inhibits large deviations, only small amplitude pursuit eye movements are generated. Furthermore, a computational model of pursuit eye movements based on gravitational energy fields is presented in the supplementary materials (**Appendix C**) that accounts for the two contrasting objectives (fixation vs. pursuit). In our data analyses, we made use of a measure of inertia for gaze dispersion and MPC for similarity, since they are simple methods that showcase clear advantages in our context. The latter also offers a metric that can be physically interpreted as it is able to capture similarity between two trajectories of different scales and spatial offsets.

Micro-pursuits

The study of micro-pursuit should aim to find consistent characteristics—like the main sequence for the micro-saccade—that can be measured through an adequate metric. Micro-pursuit being a slow eye movement, exhibiting strong similarity with the target (stimulus) trajectory, we will consider a fixation to be of the class micro-pursuit whenever the above criteria are met. In addition, if the oculomotor continuum holds true, these slow movements potentially alternate with small ballistic movements, called catch-up saccades, as is the case at macroscopic scale. It is clear that a thorough study of micro-pursuits thus needs a full characterization of fixational eye movements (especially micro-saccades) as well as the evaluation of a similarity measure between gaze and target.

[2] Here, we present the velocities rather than frequencies to provide comparable measures across reviewed articles. However, in most cases, the target signal corresponds to a sinusoidal movement, thus velocity is not constant over a period.

Quantifying pursuit movements (metrics)

To propose a definition of micro-pursuit movements, existing metrics for ocular movements will be discussed, since they will orient our choices for proposing metrics and hence our working definition.

Classical smooth pursuit is measured by velocity—or retinal slip—gain (gain = $\|\dot{\boldsymbol{q}}_G\|/\|\dot{\boldsymbol{q}}_S\|$ with $\dot{\boldsymbol{q}}_G$ the gaze velocity and $\dot{\boldsymbol{q}}_S$ the stimulus velocity), which is consistent with its closed-loop modeling (Liversedge et al., 2011). Position gain is also used—although to a lesser extent,—for instance, when dealing with catch-up saccades (Xivry & Lefevre, 2007). For the various drift phenomena described in the previous section, a variety of metrics have been used to study fixational eye movement dynamics (e.g., gaze position, velocity, acceleration, gain, and BCEA). For instance, gain measurement was used for analysis in the case of the small amplitude pursuits of monkeys on uni-variate sinusoidal motion (Skinner et al., 2018). But the authors went further and provided a spectral analysis using Fourier transform on eye signals to identify the fundamental frequency and harmonics with the expected target frequencies. However, gain is a univariate metric which does not extend to multi-variate problems. Thus, it can be used adequately only for pursuit of a target moving on a line, rather than a plane, like the visual field. Fourier analysis shares the same issue as it looks for a frequency in a univariate movement, typically horizontal.

In studies of ocular drift (Epelboim & Kowler, 1993), BCEA[3] was used to quantify the spatial variance—inertia, or spread—of the gaze. The authors obtain orientation preferences through the inferred relative anisotropy of the ellipse. Though this metric is clearly conceived for bi-variate signals, it does not provide spatio-temporal correlation between gaze and a target signal in the way gain does. Meanwhile, the box-count method used in more recent studies permits to compute dispersion of the gaze data over time, though it may suffer, like gain, from measurement noise, especially with video-based eye tracker (Engbert & Mergenthaler, 2006; Chen & Hafed, 2013). To summarize, (i) some metrics, e.g., BCEA, box count, inertia, can be used as quantifiers for the spread of a bi-variate gaze signal during an epoch, and these metrics are useful descriptors for drift and slow movements, and (ii) other metrics, e.g., gain, Fourier analysis, correlation, can be used to quantify similarity between two bivariate signals, to quantify the quality of a pursuit between gaze and a stimulus in motion. Each metric presents a trade-off that should be considered based on a theoretical definition and prediction.

Micro-pursuits: a working definition

Given the reported observations of small amplitude pursuits, the following constraints need to be considered to define a **micro-pursuit**.

[3]The surface area of the ellipse such that the data belong to this area with a probability of 68% when a two dimensional Gaussian fits the data; roughly speaking, up to a factor, it is the determinant of the empirical covariance matrix.

Amplitude as indicated by the prefix of its name, and as an analogy to micro-saccades, the micro-pursuit must be of small amplitude, within the range of fixational eye movements; typically below 1 deg;

Velocity micro-pursuit should consist of slow eye movements, similarly to drift, or smooth pursuit but at a smaller scale, with velocities below 2 deg.s^{-1};

Tracking micro-pursuits occur when a percept with motion across the observer's visual field is tracked. But, as pursuit involves matching the motion of a target by that of an observer in real time, micro-pursuit measurement of tracking should reflect the spatio-temporal interaction between the dynamics of two bi-variate signals. Hence, *similarity* between gaze dynamics should be evaluated. Because the eye movement amplitude is within the fovea's size, deformation may occur in the tracking of predictable bi-variate signals. Therefore, any similarity metric should exhibit both scale and translation invariances—spatial offset invariance may also be beneficial for measures from eye-trackers with lower precision and accuracy;

Duration the phenomenon of tracking a moving target requires by definition that it is done over a sufficiently long epoch. Thus, micro-pursuit should not occur over brief epochs such as saccades and micro-saccades;

Binocularity Conjugated movements on both the guiding and the complementary eye can be expected, being a strong indicator of oculomotor planning.

We propose that gaze signal epochs satisfying the above description be considered as **micro-pursuits**. As this is a proposed working definition, micro-pursuits may correspond to entire eye fixation periods, making it possible for micro-pursuit to be punctuated by other fixational eye movements. Once its properties are defined more precisely than above and detection algorithms can be developed, it will be possible discriminating micro-pursuits from other fixational eye movements, like micro-saccades.

Descriptive statistics for the classification of micro-pursuits

Choosing an adequate metric for analysis was key, given the constraints presented in the previous section and our experimental setup. Two metrics, **inertia** and **maximally projected correlation** (MPC), are used in this work; they provide complementary information about the data. The first is a measure of the spatial dispersion of the gaze within a fixation to investigate the marginal dynamics of the gaze during fixational eye movements. The second metric gives a quantification of similarity—and hence interaction—between the gaze and a target. Compared to works in the literature with similar observations (Martins et al., 1985; Skinner et al., 2018), an essential aspect was to have a metric that could reflect similarity with

noise robustness, as well as scale and translation invariance. Moreover, this was needed in the context of movements in the plane, rather than rectilinear ones for which uni-variate measures are sufficient. A benefit from such considerations is to propose a generalized metric for micro-pursuit that could be applied to track perceived motion in the two-dimensional visual field projected on the retina. MPC offers a method to quantify spatio-temporal similarity between two bi-variate signals. Furthermore, inertia and MPC can both be applied on the gaze signals in fixation epochs detected by video-based eye-tracker algorithm. Their mathematical relationship is detailed more in-depth in the **Appendix B**.

Measuring gaze dispersion with inertia

The dispersion of gaze within a fixation was computed using a measure of inertia, a metric used to quantify the spread of a cloud of data points with respect to a fixed point, usually its empirical mean. Here, we used a similar, but generalized formula based on the mean quadratic distance from an arbitrary reference point. As such, in the case of stimulus motion, we can compute inertia with respect to the stimulus' center of gravity. Let $\overline{\boldsymbol{q}}_U \doteq \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{q}_U^i$ be the empirical mean of a signal whose samples ($i = 1,\dots,N$) are given by $\boldsymbol{q}_U^i = \left[x_U^i, y_U^i\right]^\top$. We will use $U = G$ for the observed gaze and $U = S$ for the coordinates of the stimulus' (center of gravity). Gaze inertia I was computed over the stimulus trajectories over a trial as follows:

$$I = \frac{1}{N}\sum_{i=1}^{N}\left(\boldsymbol{q}_G^i - \boldsymbol{q}_O^i\right)^\top\left(\boldsymbol{q}_G^i - \boldsymbol{q}_O^i\right) = \frac{1}{N}\sum_{i=1}^{N}\|\boldsymbol{q}_G^i - \boldsymbol{q}_O^i\|^2 \qquad (1)$$

where N represents the total number of frames in the trial, $\boldsymbol{q}_G = [\boldsymbol{x}_G, \boldsymbol{y}_G]^\top$ the measured monocular bi-variate gaze signal coordinates and $\boldsymbol{q}_O = [\boldsymbol{x}_O, \boldsymbol{y}_O]^\top$ the origin reference point coordinates in the screen plane—however, one can compute inertia with respect to other points in space, e.g., stimulus center of gravity or the fixation's mean gaze position. Inertia quantifies gaze displacement as does BCEA (Epelboim & Kowler, 1993) and box-count measures (Engbert & Mergenthaler, 2006). Its key advantage over the former two is that inertia is a more intuitive measure of spatial displacement over a fixation period. The box-count metric is simple and provides similar insight in gaze dispersion over an epoch, it is dependent on the size of the box in space and time used for analysis. Hence, it corresponds to a down sampling measurement of inertia over a fixed time window. Finally, inertia provides the advantage of being a metric relative to a chosen origin or reference point—box count being independent of the origin—and thus it can be used to look at spatial displacement in the following three contexts: (1) absolute inertia (I_{screen}) is obtained by choosing the center of screen as a reference (absolute, like box count; $\boldsymbol{q}_O = [0,0]^\top$), (2) relative retinal image instability (I_{stimulus}) by choosing the stimulus' center of gravity (for pursuit; $\boldsymbol{q}_O = \boldsymbol{q}_S = [\boldsymbol{x}_S, \boldsymbol{y}_S]^\top$), and (3) general relative fixational eye movement instability (I_{fixation}) by referring to the fixation center of gravity (obtained by choosing $\boldsymbol{q}_O = \overline{\boldsymbol{q}}_G = [\overline{\boldsymbol{x}}_G, \overline{\boldsymbol{y}}_G]^\top$ with $\overline{\boldsymbol{q}}_G$, the empirical mean of the gaze for a N samples fixation epoch).

Measuring gaze-stimulus similarity with Maximally Projected Correlation (MPC)

Though humans can intuitively express a qualitative judgment of similarity between two trajectories, obtaining a quantified and objective value for any two bi-variate signals is not as trivial as one might suppose. Gain, of gaze velocity over stimulus velocity, has been used as a metric in pursuit data analysis (Skinner et al., 2018), though the stimulus moved in a univariate context: either horizontal or vertical. In bi-variate signals, however, a gain will be obtained for each dimension of the signal, and hence some form of projection to obtain a scalar metric is required. Although similarities between the stimulus and gaze trajectories can be quantified with a diversity of metrics, we will here focus on a measure based on multi-variate statistical theory (Anderson, 2003; Muirhead, 2009), quantifying the interaction between the stimulus ($\boldsymbol{q}_S$) and gaze ($\boldsymbol{q}_G$), in order to infer on the similarity of their trajectories during fixations. We choose to determine the direction of the plane for which correlation between gaze and target within a fixation are maximized, and report the such obtained correlation value, which we call MPC. Our metric hence inherits the ease of interpretability from (Pearson) correlation values and has low computational costs (just as gain). In addition, for unidirectional motion (see, e.g., (Skinner et al., 2018)), this exactly corresponds to Pearson's correlation coefficient between the two time-series.

Let $\boldsymbol{\Sigma}_{SG} \doteq \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{q}_S^i \boldsymbol{q}_G^{i\top} - \overline{\boldsymbol{q}}_S \overline{\boldsymbol{q}}_G^\top$ the empirical (variance-)covariance matrix between stimulus (S) and gaze (G). We then write ρ^* as the maximal absolute empirical correlation that can be obtained under simultaneous projections onto a one-dimensional space, i.e.,

$$\rho^* \doteq \max_{\boldsymbol{w}} \rho(\boldsymbol{w}) \quad \text{where} \quad \rho(\boldsymbol{w}) \doteq \frac{\boldsymbol{w}^\top \boldsymbol{\Sigma}_{SG} \boldsymbol{w}}{\sqrt{\boldsymbol{w}^\top \boldsymbol{\Sigma}_{SS} \boldsymbol{w}}\sqrt{\boldsymbol{w}^\top \boldsymbol{\Sigma}_{GG} \boldsymbol{w}}} \qquad (2)$$

and $\boldsymbol{w}$ represents the coordinates of the vector onto which both the gaze and the stimulus signal are projected. This method projects the data in a new space, and provides a quantity bounded between -1 and 1, where 1 shows perfect correlation and -1 perfect anti-correlation. By construction, MPC is invariant with respect to scale and to a translation of either or both of the signals.

To summarize this section, in this work, inertia with respect to screen (I_{screen}) was used as a measure of gaze displacement. Inertia with respect to stimulus (I_{stimulus}) was used as a measure of retinal image displacement. Inertia with respect to fixation (I_{fixation}) was used as a measure of fixational eye movement displacement. And finally, MPC (ρ^*) was used as a measure of similarity between gaze and stimulus trajectory, during a fixation.

Main Experiment: Necker cube

Micro-pursuits were observed and systematically detected at first in an experiment in which a moving ambiguous Necker cube stimulus was presented and participants had to report their perceived orientation. They were instructed to keep their gaze fixed on a static fixation cross at the center of the screen and report which side of the cube was perceived at the front; either lower-left or upper-right square. The main objectives of the experiment was to manipulate the rate of reversal by imposing different motion to the cube. In this paper, we focus solely on the oculomotor analysis of this data set, because the manipulation failed to induce any change in the reversal rate between the percept nor any observable percept modulation.

Methods

Apparatus

The display used was a 40 cm by 30 cm (20 inches) VisionMaster Pro 513 screen of resolution 1024 by 768 pixels and a 75 Hz refresh rate, located 57 cm from the participants, with mean gray luminance at 68 cd.m^{-2}. Eye movements were recorded using the Eyelink 1000 (SR Research, Ottawa, Ontario, Canada). Both eyes were tracked with a 1000 Hz sampling rate. The head was stabilized using a chin rest. A nine-point calibration routine was carried out at the beginning of each task and was repeated at the beginning of each block (every 15 trials) or when drift correction, performed every 5 trials, reported a mean error superior to 0.5 deg.

Stimulus & motion conditions

We imposed three type of motion to an ambiguous Necker cube of 2.6 by 2.5 deg (Fig. 1-A): (1) '**FX**' the control condition with no motion, (2) '**RW**' an unpredictable motion condition with a random walk and (3) '**LJ**' the predictable motion condition where the cube moved along Lissajous trajectories (see Fig. 1-B). **Random walk** trajectories were implemented by choosing at each time step an amplitude chosen from an exponential-Gaussian distribution and an orientation from a uniform distribution on $(-\pi,\pi)$. The exponential-Gaussian distribution was built from the sum of two independent variables, $\rho = G+E$ where $G \sim \mathcal{N}(\mu = 1.1;\ \sigma = 0.2)$ is the Gaussian component, and $E \sim \mathcal{E}(\lambda = 0.1)$ is the exponential one—units are in pixels (pix) and the $\sim$ symbol stands for "distributed according to". A radial limit of 10 pix (0.329 deg) with respect to the center of the screen was implemented so that a step that would exceed the limit would have its orientation reversed such that the step would bounce back towards the center. **Lissajous** trajectories in the LJ condition were defined by $x(t) = A\sin(c\theta t)$ and $y(t) = B\sin(d\theta t + \phi)$ with, in our setup, $A = B = 14$ pix (0.5497 deg) and $\phi = 0$ rad. The Lissajous ratio between signal frequencies randomly (uniformly) chosen across trials so that $(c,d) \in (2,3),(3,2),(-2,3),(-3,2)$ and $\theta = 2\pi\frac{(30/2.21)}{415} = 0.2$ Hz. The parameters' values were chosen empirically through *ad hoc* tests.

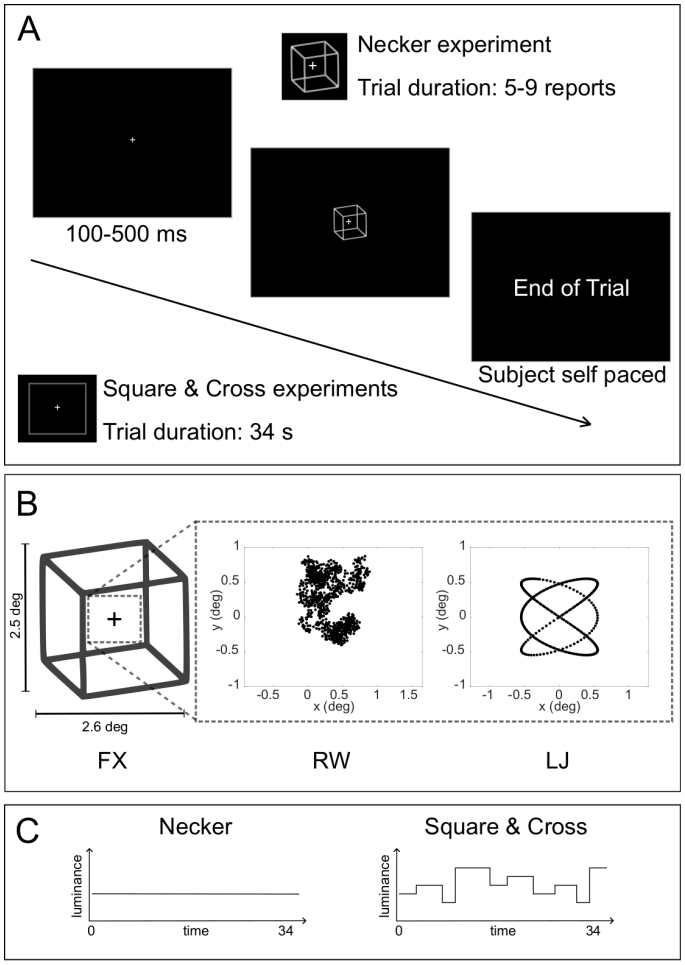


Figure 1: **Experimental protocols**. **A** is a timeline of a trial for all three experiments (Necker, Square, Cross). For the Necker experiment, a Necker cube was displayed and the trial finished if the participant had reported a randomly picked number of perceptual reversals. For the Square and Cross experiments, a plain square was displayed and trial lasted approximately 34 seconds. A fixation cross was shown during a randomly chosen interval between 100 and 500 ms. **B** shows the three different stimulus motion conditions; (1) FX, for the control no-motion condition, (2) RW, for the unpredictable random walk condition, and (3) LJ, for the predictable motion based on Lissajous trajectories. **C** are representations of the stumuli's luminance. For the Square & Cross experiments, luminance changed randomly between 5 levels in order to provide the participants with a perceptual report task, while the Necker cube always kept a constant luminance.

Stimulus spatial displacement due to movement was controlled across motion conditions. Indeed their inertia with respect to screen distribution were similar, with RW and LJ generating displacement of the same order of magnitude on average

over trials ($\bar{I}^{RW}_{\text{screen}} = 0.2995 \pm 0.1988$, $\bar{I}^{LJ}_{\text{screen}} = 0.2747 \pm 0.1372$).

Tasks & participants

23 adults, with normal or corrected-to-normal vision (self-assessed), participated in the experiment (15 females and 8 males; age range = 20–71 years, $\mu = 28.35 \pm 10.93$ years, whose tasks were two-fold:

- fixate a fixation cross at the center of the screen for a random interval between 100 and 500 ms (uniform distribution);
- report percept reversals of an ambiguous Necker cube by pressing the arrows of a keyboard when perceptual changes occurred.

The experiment followed a continuous viewing paradigm in which trials had variable (random) durations ($\mu = 34.00 \pm 13.26$ seconds, see Fig. 1-A) and ended based on which of the following condition happened first

number completion of a trial-based randomly (uniformly) set integer number ($n_{\text{rev}} \sim \mathcal{U}(5,9)$) of perceptual reversals on the ambiguous stimulus (see Fig. 1-A);

time-out maximal percept duration of 20 seconds.

The experiment was programmed using the *PsychToolBox* in MATLAB (Brainard, 1997). All participants gave their informed written consent before participating in the study, which was carried out in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki) for experiments involving humans and as approved by the ethics' committee of University Grenoble Alpes.

Data analysis

Data pre-processing: in our data analysis, only fixations of sufficient duration (> 80 ms) were considered. The duration threshold was set based on (1) the lack of significant fixations of interest in shorter time windows and (2) the necessity for the MPC metric to have a sufficient number of samples (see **Appendix B**). Guiding eye gaze signals were first passed through a corrective process to adjust for pupil area deformation as described in Choe and colleague's work (Choe et al., 2016). As the gaze and stimulus signals were systematically compared and computed together, we then applied a Butterworth filter (second order low-pass filter with a cut-off frequency of $f_c = 35$ Hz) to smooth the gaze data and down-sampled the gaze signal at the same frequency as the refresh rate of the stimulus (75Hz). Thus, all analyses are done with data down sampled from 1000 Hz to 75 Hz. Fixations generating inertia with respect to screen values beyond two standard deviation from the mean or NaN (due to missing samples) were considered as samples with faulty or jittery gaze recording and were removed from analyses. Data for Fig. 3 and statistical tests only consider fixations without micro-saccades, where the latter are detected by an algorithm proposed by Engbert and Kliegl (Engbert & Kliegl, 2003) based on the binocularity criterion. The algorithm uses relative thresholds based on median absolute deviation of the eye velocity, here over a fixation. Data for Fig. 2 and Fig. 5 are analyzed including fixations containing micro-saccades. Outliers were defined as data points[4] beyond two standard deviation from the mean, and were systematically removed from analyses. The results presented do not show these outliers, for better readability, but we also conduct the analyses with the outlier and found the same effects for all tests and experiences.

Statistical methods: statistical tests were conduct to assess difference between motion condition both within subjects (*group analysis*) and at the subject level (*individual analysis*). For both levels, we applied non parametric tests, since we did not have any priors on the data distribution for inertia and MPC. For group analysis, statistical tests were conducted using 10000 permutations on non parametric approximate (Monte Carlo) Friedman test for inertia, and if significant differences were inferred, approximate (Monte Carlo) Wilcoxon signed-rank tests were used for pairwise comparisons between conditions (with a decision criterion at $p = 0.05/3 = 0.017$). For MPC, a Wilcoxon signed-rank test was carried out All these tests were delivered using bootstraps based on 10000 permutations conditional on subjects for every experiments (Necker, Cross and Square) and metrics (I_{stimulus}, I_{fixation} and MPC) using the packages *coin* (Hothorn, Hornik, Van De Wiel, & Zeileis, 2006) and *rstatix* (Kassambara, 2020). Effect size were computed from the χ^2 statistics and using the transformation described by Tomczak and Tomczak (Tomczak & Tomczak, 2014) to get a Kendall W, that vary between 0 and 1, with 1 the maximum effect size :

$$W = \frac{\chi^2}{N(k-1)}. \tag{3}$$

With W, the Kendall's W value, χ^2 the Friedman test statistic value, N the sample size and k the number of measurements per subject. For each test, we report the χ^2 Friedman test statistic, with the p-value (p) computed with the bootstrap, it's effect size (Kendall W). For individual statistical analyses, we carried out an approximate Kruskal-Wallis test for inertia and an approximate Wilcoxon-Mann & Whitney test for MPC and pariwise comparisons using the same bootstrap package, with 10000 permutations. To compare experiments' data, Kruskal-Wallis tests were used over the three experiments' RW and LJ data, respectively, and Wilcoxon-Mann & Whitney tests were used to infer differences between pairs of experiment data-sets in each condition, with the same packages.

Results

Micro-saccades

We described peak velocities, amplitudes, and rate of occurrences of microsaccades detected during fixations (n =

[4] Here, data points refer to a statistic of a fixation period, for a given experiment, subject and condition. We also have outliers subject (71 years old) that is not removed.

21197, for Necker), using the algorithm from Engbert and Kliegl (Engbert & Kliegl, 2003). Distributions of micro-saccades' peak velocities and amplitudes across conditions and experiments are shown in Fig. 2-A. Detected micro-saccades showed similar main sequences across motion conditions. Moreover, when we add the MPC value of the fixation in which the micro-saccade was detected (color scale), we observe (i) a higher prevalence of fixations with high similarity between gaze and predictable motion (LJ) than in the random walk (RW) condition, and (ii) no apparent (qualitative) correlation between MPC and micro-saccadic properties can be established. Micro-saccade rates are described in Fig. 2-B, with bootstrapped 95% confidence intervals.

When fixations with detected micro-saccades were kept, data pre-processing led to the removal of 12.32% of fixations for the Necker experiment based on fixation duration and outlier removal based on inertia with respect to screen. When fixations with detected micro-saccades were removed, data pre-processing led to the removal of 63.39% of fixations. Results presented next were computed on fixations not containing micro-saccades, as they describe the purest form of micro-pursuits. However, when including fixations containing micro-saccades, results led to the exact same conclusions.

Inertia & MPC

We looked at the impact of the cube motion on eye movement and retinal image displacement. The former is made explicit through the inertia of gaze with respect to its average position within a fixation, see Fig. 3-B, whereas the latter is given by the inertia of the gaze with respect to the stimulus' center of gravity, see Fig. 3-A. Descriptive statistics and statistical tests' summary are given in Table 1.

Dispersion of eye movements around the fixation, computed with median inertia of the eye with respect to mean fixation position (I_{fixation}; see Fig. 3-C) differed with motion condition ($\chi^2 = 37.130; p < 0.0001; W = 0.807$). Paired comparisons of I_{fixation} showed differences between FX, RW and LJ ($Z_{FX-RW} = -2.4027, p = 0.016; Z_{RW-LJ} = -4.1973, p < 0.0001$ and $Z_{FX-LJ} = -4.1973; p < 0.0001$). Thus, when computing retinal image displacement, we found that the median inertia differed across cube motion conditions (see Fig. 3-A). Indeed, we find a difference in inertia computed with respect to the center of gravity of the stimulus (I_{stimulus}) with motion condition ($\chi^2 = 23.565; p < 0.0001; W = 0.512$). Median inertia differed in the conditions where the stimulus was in motion ($Z_{FX-RW} = -3.9844, p < 0.0001; Z_{FX-LJ} = -3.9539, p < 0.0001$ and $Z_{RW-LJ} = 0.09124, p = 0.9445$).

When considering that stimulus inertia was equivalent for both motion conditions ($\bar{I}^{RW}_{\text{screen}} = 0.2995 \pm 0.1988, \bar{I}^{LJ}_{\text{screen}} = 0.2747 \pm 0.1372$), the results suggest that both types of motion applied on the stimulus generated different effects on eye movements. Indeed, eye trajectories were more similar in the predictable LJ motion condition ($\tilde{\rho}^*_{LJ} = 0.921 \pm 0.047$) than in the unpredictable RW motion condition ($\tilde{\rho}^*_{RW} = 0.509 \pm 0.048$) with significant differences ($\chi^2 = 23; p < 0.0001; W = 1$ and $Z_{RW-LJ} = -4.1972; p < 0.0001$). The data is reported in Fig. 3-E. We evaluated the effect of the cube motion for every subject and found similar results (Fig. 3-B-D-F) that will be described in more details later.

Binocularity & velocity

As binocularity is an important criteria than can discriminate between erratic noisy movement and conjugate and functional movement (Fang, Gill, Poletti, & Rucci, 2018), we also looked at the similarity of gaze between the directing and non-directing eye, to look at how conjugated the eyes were. We found overall differences across conditions ($\chi^2 = 37.130; p < 0.0001; W = 0.807$). Paired comparisons of eye *versus* eye similarity showed differences between FX, RW and LJ ($Z_{FX-LJ} = -4.1973, p < 0.0001; Z_{FX-RW} = -2.2202, p = 0.023$ and $Z_{RW-LJ} = -4.1973, p < 0.0001$). Results are reported in Fig. 4-A, along with analyses for each participants Fig. 4-B.

To further investigate the pursuit description, we computed the MPC on the velocity signals, calculated on the position signals, down-sampled at 75 Hz, over 6 samples. In fact, as for the position analysis, LJ's predictable motion ($\tilde{\rho}^*_{LJ} = 0.798 \pm 0.096$) led to higher velocity similarity between the eyes and the target than for RW's unpredictable motion ($\tilde{\rho}^*_{RW} = 0.246 \pm 0.052$) with significant differences ($\chi^2 = 23; p < 0.0001; W = 1$ and $Z_{RW-LJ} = -4.1973; p < 0.0001$). The data is reported in Fig. 4-C, along with analyses for each participant in Fig. 4-D.

Intermediary discussion

When looking at our descriptive statistics (Table 1 and Fig. 3, A-B-C), participants' median similarity based on MPC is centered on values of high correlation in the predictable motion condition (LJ) compared to the other motion condition (RW). This means that fixational eye movement gaze trajectories were, for most subjects, highly similar to that of the stimulus moving on screen. On the other hand, the unpredictable motion condition (RW) led to much lower similarity measurements; an observation that can be explained by the incapacity of the oculomotor system to predict the motion of the Necker cube as motion followed random walk dynamics.

Therefore, globally, participants' gaze was influenced by the cube motion significantly more in LJ, where motion was predictable, than in RW, where motion was unpredictable, even though the oculomotor instructions were to fixate the cross in the middle of the screen for both. Moreover, the gaze in LJ showed similarity with the stimulus trajectories. All these measures were gathered on gaze data within fixation events and the difference between LJ and RW conditions show that oculomotor drift alone, as defined above, within fixational eye movements cannot account for this similarity. The oculomotor system would have to integrate visual information in order to quasi-systematically track the stimulus. We therefore refer to these detected fixational eye movements as micro-pursuits, in an effort to keep the analogy with the micro-saccades, while respect-

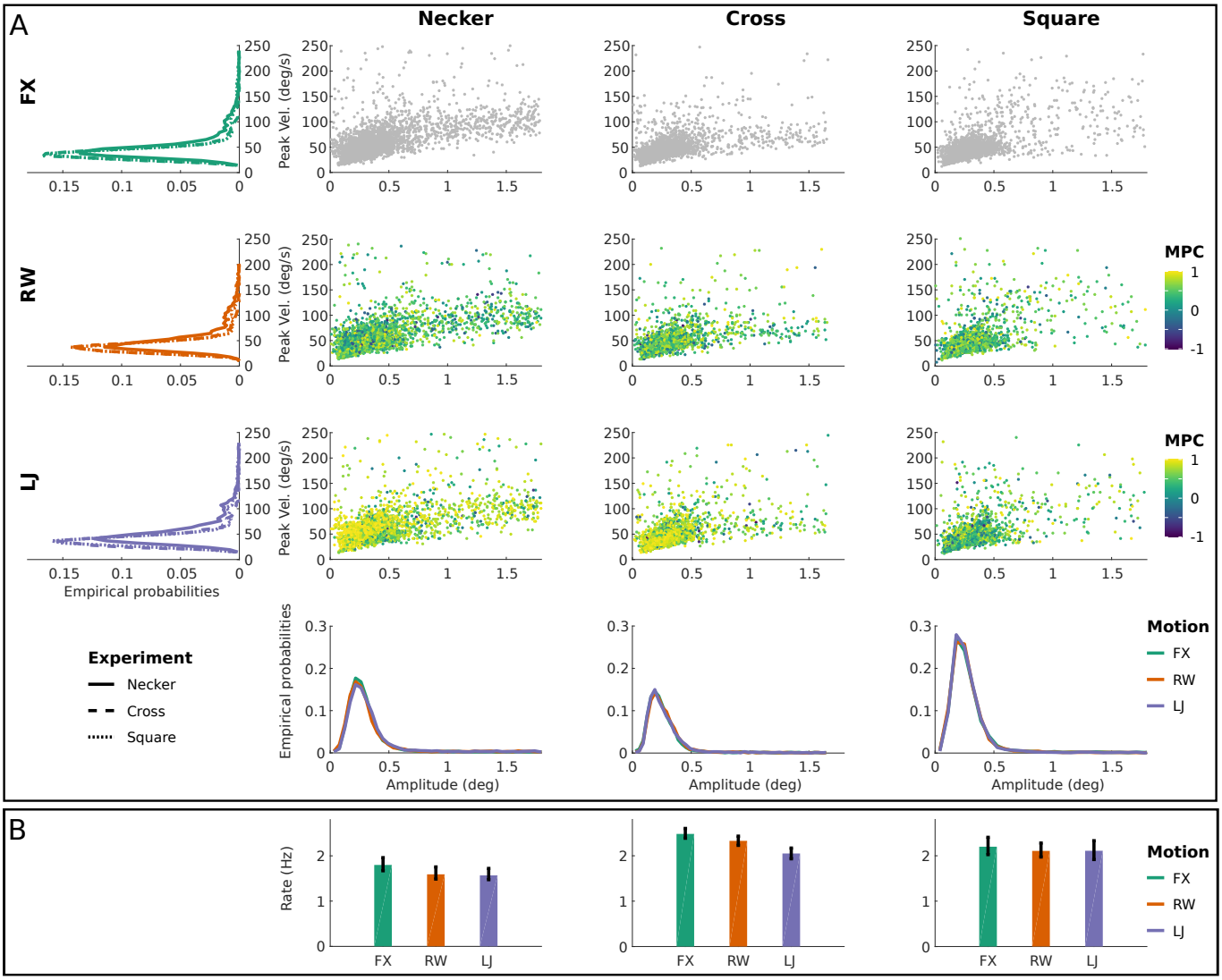


Figure 2: **Micro-saccade analysis**. **A** shows the main sequences when plotting micro-saccades' amplitudes versus peak velocities for all three experiments (Necker, Cross and Square) and conditions (FX, RW, LJ). The color encodes the micro-saccade's fixation similarity score (using MPC) in the LJ and RW conditions. *Left side*, marginal distributions of peak velocity depending on the experiment and condition are given, while *below*, marginal distributions for amplitudes are shown. **B** shows mean micro-saccade rates over experiments and conditions with, in black, 95% confidence intervals computed using bootstrap ($n = 200$ iterations).

Necker ($N = 23$)	FX	RW	LJ	χ^2	p	W
$\tilde{I}_{\text{stimulus}}$	0.488 ± 0.189	0.649 ± 0.190	0.629 ± 0.159	23.565	< 0.0001	0.512
$\tilde{I}_{\text{fixation}}$	0.019 ± 0.009	0.024 ± 0.015	0.071 ± 0.051	37.130	< 0.0001	0.807
$\tilde{\rho}^*$	n/a	0.509 ± 0.048	0.921 ± 0.047	23	< 0.0001	1

Table 1: *Left*, Summary statistics of three measures for the Necker experiments in the FX, RW and LJ motion conditions; inertia w.r.t. stimulus center of gravity (I_{stimulus}), inertia w.r.t. fixation center of gravity (I_{fixation}), and MPC (ρ^*). For each condition in the Necker experiment, median values over participants' data are given with median absolute deviation (mad) following the $\pm$ sign. *Right*, Approximate Friedman test results (χ^2; p) and size effect (W) are given.

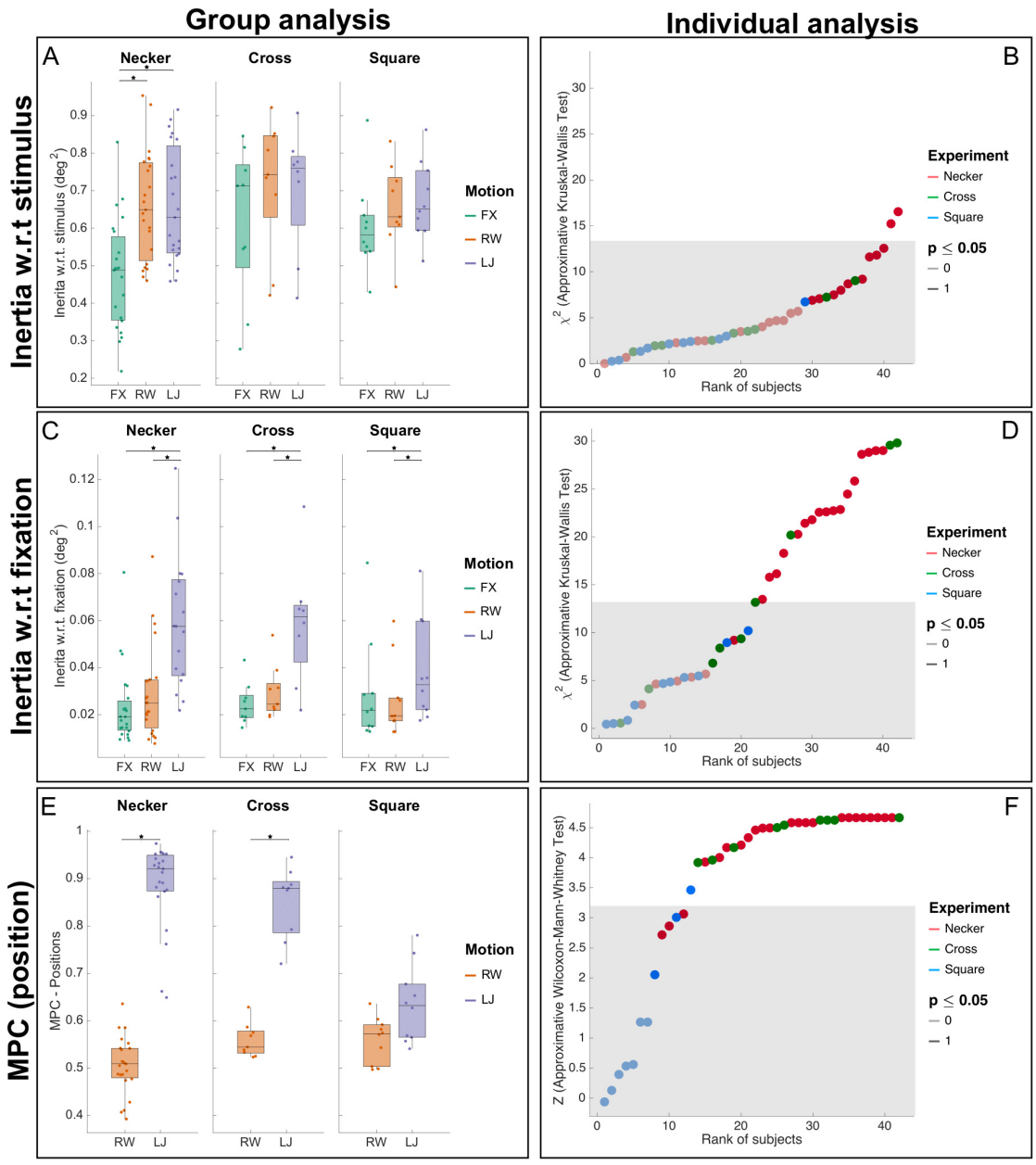


Figure 3: **Micro-pursuit analysis**. **A** is a box plot of I_{stimulus} over the three experiments (Necker, Cross and Square) and three motion condition (FX, RW and LJ). Stars represent significant differences in pairwise comparisons using the Wilcoxon-Mann-Whitney test in a bootstrap. **B** plots the individual analysis results for I_{stimulus} in all three experiments' participants using an approximate Kruskal-Wallis test in a bootstrap. All the participant have significant ($p < 0.05$) results. For individual analysis, statistics (Z score or χ^2) that fall inside the 95 % confidence interval were drawn with light color whereas statistics values outside the 95% confidence interval were drawn in plain color. The gray area defines a conservative confidence interval corrected for multiple comparisons (Bonferroni), i.e., 42 comparisons for the 42 tests computed on each subjects. **C** is a box plot of I_{fixation} over all experiments and conditions. **D** plots the individual analysis results for I_{fixation}. The outcome of the statistical test per participant are given through different lightness value, with 1 (darker) meaning that $p \leq 0.05$ and 0 (lighter) the opposite. **E** is a box plot of MPC (ρ^*) over all experiments and the RW and LJ motion conditions. **F** plots the individual analysis results for ρ^* in all participants using an approximate Wilcoxon-Mann-Whitney test.

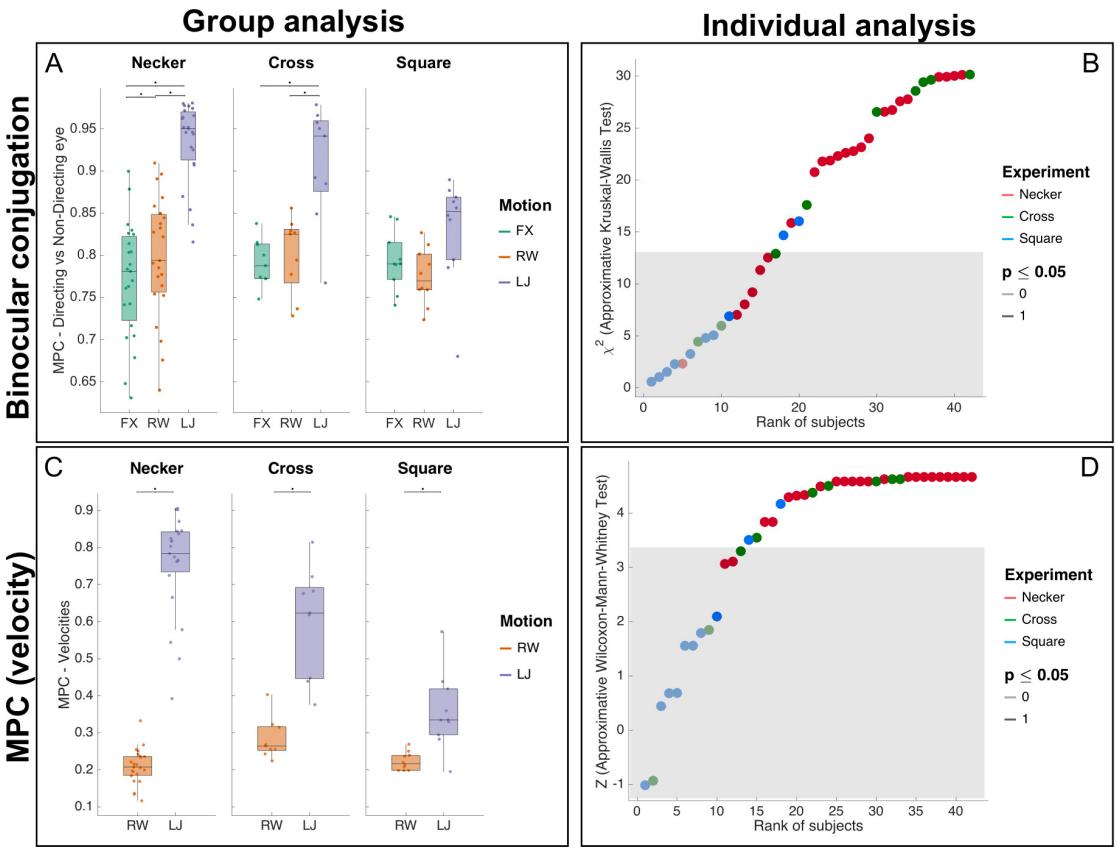


Figure 4: **Micro-pursuit additional analyses**. **A** is a box plot of directing vs non-directing eye similarity over the three experiments (Necker, Cross and Square) and three motion condition (FX, RW and LJ). Stars represent significant differences in pairwise comparisons using the Wilcoxon-Mann-Whitney test in a bootstrap. **B** plots the individual analysis results for directing vs non-directing eye similarity in all three experiments' participants using an approximate Kruskal-Wallis test in a bootstrap. For individual analysis, statistics (Z score or χ^2) that fall inside the 95% confidence interval were drawn with light color whereas statistics values outside the 95% confidence interval were drawn in plain color. The gray area defines a conservative confidence interval corrected for multiple comparisons (Bonferroni), i.e., 42 comparisons for the 42 tests computed on each subjects. **C** is a box plot of MPC based on velocity vectors over all experiments and conditions. **D** plots the individual analysis results for MPC based on velocity vectors. The outcome of the statistical test per participant are given through different lightness value, with 1 (darker) meaning that $p \leq 0.05$ and 0 (lighter) the opposite.

ing the definition and metrics given above. Given the non-dedicated and unpredicted observation of the oculomotor phenomenon in the Necker experiment, we carried out a second set of experiments to replicate the generation of micro-pursuit using a simpler stimulus, and to verify that the phenomenon is not caused by the presence of a bi-stable stimulus—namely the Necker cube.

Replication Experiments: Square & Cross

The experimental protocol is similar to the previous one (Necker experiment) except that the Necker cube is replaced by a gray square and subjects have to report changes in luminance in either the fixation cross (Cross experiment) or the square (Square experiment). In the Cross experiment, we set the participants' tasks and stimuli such that they had to follow a moving cross and detect changes of luminance on it. In the Square experiment, the setup aimed to investigate whether a complete reproduction of the Necker experiment, with a square instead of the Necker cube would still lead to the observation of micro-pursuits.

Methods

Material and stimuli were identical to the previous experiment unless specified.

Apparatus

The stimulus was displayed on a 36 cm by 27.5 cm (19 inches) Dell M993s CRT screen of resolution 1280 by 1024 pixels and a 75 Hz refresh rate, located 57 cm from the participants, with white luminance at 70.89 cd.m^{-2}, black at 0.09 cd.m^{-2} and mean gray at 15 cd.m^{-2}. Eye tracking was done using an EyeLink 1000+ (SRT Research). Calibration was applied using a 5 points procedure between each block and if drift correction failed. Drift correction was applied between each trial. Participants had their head stabilized by sitting and resting their chin on a rest and their forehead against a bar.

Stimulus & motion conditions

As in Experiment 1, we replicated the three motion conditions (FX, RW, & LJ) using the same parameters with balanced mean inertia. Trials lasted 34 seconds (the mean time duration of *Experiment 1: Necker Cube*) in which the same fixation cross was presented, and a moving object followed its trajectories depending on the condition (see Fig. 1-A).

Tasks & participants

The participants had to fixate a fixation cross surrounded by a square (2.5 deg by 2.5 deg), displayed in Fig. 1-A. They also had a perceptual task in which they had to report luminance changes using the same keys of the keyboard as in the Necker Experiment. However, here the alternations were randomly selected among 5 levels of luminance (levels at 30%, 40%, 50%, 60% and 70% of white) and duration of a level was selected using a log-normal probability law Log$-\mathcal{N} \sim (\mu = 1, \sigma = 1)$ seconds (see Fig. 1-C for a schematic representation of luminance over time). Two conditions were contrasted:

1. *Implicit pursuit - moving Square luminance change detection*: fixate the fixation cross at the center of screen, and report changes in luminance of the surrounding square moving with the three types of motions.

2. *Explicit pursuit - moving Cross luminance change detection*: fixate the fixation cross and report changes in luminance of the fixation cross moving with the three types of motions.

The 19 participants (17 females and 2 males; age range = 18-30 years, $\mu = 20.63 \pm 2.61$ years), with normal or corrected-to-normal vision, were randomly oriented in one of the two experiments (Cross; $n = 9$, and Square; $n = 10$) and provided their informed written consent before participating in the study, which was carried out in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki) for experiments involving humans and approved by the ethics' committee of University Grenoble Alpes. We estimated the number of participants to be included in the protocol based on a power analysis using *g*power* (Faul, Erdfelder, Buchner, & Lang, 2009) with $\alpha = 0.05$ and $1 - \beta = 0.95$. We found that we needed a minimum sample size of 9 participants (with 45 trials) to replicate the observations with a power of 0.95.

Data analysis

Data analyses were identical to the previous experiment.

Results

The data was analyzed by applying the same signal processing procedures and statistical methods as in the Necker experiment for inertia or MPC. When fixations with micro-saccades were kept, data pre-processing led to the removal of 8.79% and 9.23% of fixations for the Cross and Square experiments, respectively, based on fixation duration and outlier removal for inertia with respect to screen. Micro-saccade analysis (Fig. 2) led to the extraction of main-sequences with patterns showing no apparent qualitative differences between experiments (Necker, Cross and Square) for amplitude and peak velocity, across motion conditions (FX, RW and LJ).

When fixations with micro-saccades were removed as well, data pre-processing led to the removal of 65.43% and 72.73% of the data, in Cross and Square, respectively. Results presented in this section were computed on the fixations without micro-saccades, however when doing these analyses with fixations with micro-saccades, results led to the same conclusions.

Cross experiment: explicit micro-pursuits

When participants had to explicitly follow the fixation cross, on which the motion and luminance signals were applied, similar patterns to the Necker experiment were found for inertia of gaze. Dispersion of eye movements around the fixation, computed with median inertia of the eye with respect to mean fixation position (I_{fixation}; see Fig. 3-C) differed with motion condition ($\chi^2 = 8.667; p = 0.0096; W = 0.481$). Moreover, paired comparisons revealed differences between FX, RW and LJ ($Z_{FX-RW} = -2.403, p = 0.016; Z_{RW-LJ} = -2.5471; p = 0.0083$ and $Z_{FX-LJ} = -2.5471; p = 0.0085$). Retinal image displacement differed with cube motion (see Fig. 3-A). We also found no significant difference in inertia computed with respect to the center of gravity of the stimulus (I_{stimulus}) with motion condition ($\chi^2 = 4.667; p = 0.103; W = 0.704$).

Given the fact that stimulus inertia was equivalent for both motion conditions, this suggests that motion of the stimulus generated different effects on eye movements. Indeed, eye trajectories were more similar in the predictable LJ motion condition ($\tilde{\rho}^*_{LJ} = 0.880 \pm 0.050$) than in the unpredictable RW motion condition ($\tilde{\rho}^*_{RW} = 0.545 \pm 0.032$) with significant differences ($\chi^2 = 9; p = 0.0039; W = 1$ and $Z_{RW-LJ} = -2.6656; p = 0.0043$). The data is visualized in Fig. 3-E. We evaluated the effect of the cube motion for every subject and found similar results (Fig. 3-F).

Square experiment: implicit micro-pursuits

Dispersion of eye movements around the fixation, computed with median inertia of the eye with respect to mean fix-

ation position (I_{fixation}; see Fig. 3-C) differed with motion condition ($\chi^2 = 8.6; p = 0.0109; W = 0.43$). Moreover, paired comparisons revealed a difference between RW and LJ ($Z_{RW-LJ} = -2.3953; p = 0.0126$) but not with FX ($Z_{FX-RW} = 0.866, p = 0.4321$ and $Z_{FX-LJ} = -1.8857; p = 0.0609$). But retinal image displacement differed with cube motion (see Fig. 3-A). Indeed, we did not find a difference in inertia computed with respect to the center of gravity of the stimulus (I_{stimulus}) with motion condition ($\chi^2 = 2.4; p = 0.3621; W = 0.12$).

Given the fact that stimulus inertia was equivalent for both motion conditions, this suggests that motion of the stimulus did not generate different effects on eye movements. Unlike in the other experiments, eye trajectories were not more similar to stimulus trajectories in the predictable LJ motion condition ($\tilde{\rho}^*_{LJ} = 0.637 \pm 0.097$) or in the unpredictable RW motion condition ($\tilde{\rho}^*_{RW} = 0.573 \pm 0.044$) with no inferred statistical difference ($\chi^2 = 1.6; p = 0.3384; W = 0.16$). The data is visualized in Fig. 3-E. We evaluated the effect of the cube motion for every subject and found similar results (Fig. 3-F).

Individual analyses

We conducted the same analysis on every subject and results are displayed for the three experiments and three motion conditions in figure (Fig. 3-B-D-F). For every subject, we plotted the χ^2 or Z score statistics for the approximate Kruskal-Wallis and Wilcoxon-Mann-Whitney tests against their overall rank according to these statistics. For all subject we observed a main effect of inertia with reference to the stimulus (I_{stimulus}, with identical inertia between LJ and RW compare to FX. When looking at retinal displacement, we find the same pattern of result, i.e., a main effect of motion, with inertia with reference to the fixation (I_{fixation}) similar for FW and RW but lower to LJ for Necker and Cross experiments. For the Square experiment results were mixed within subject suggesting idiosyncratic behaviors. Finally, we observe more similar gaze pattern (high MPC) for the LJ condition both in the Necker and Cross experiments for every subject (except one out of nine in Cross) but mixed results for the square experiment. Thus individual analyses show that results observed at the group level are replicated at the subject level.

Comparing Necker, Cross, & Square experiments

To summarize, descriptive statistics of detected micro-saccades in terms of main sequences (amplitude and peak velocity; see Fig. 2-A) and micro-saccade rates (Fig. 2-B) show that overall, micro-saccades are consistent across Necker, Cross and Square, for all motion conditions. However, the Cross and Necker predictable (LJ) condition data seem to exhibit a different behavior than the other conditions and experiments when looking at gaze-target similarity (MPC). The micro-saccades' fixation MPC display many high correlation values, in contrast to the other conditions, and unlike the LJ condition in the Square experiment.

Fig. 5 provides a focus on MPC for fixations in all data sets, as well as for some selected signals that showcase some typical examples of gaze-stimulus pairs for different values of MPC. Since one cannot track the RW movements, the distribution of MPC under this condition serves as a baseline or null hypothesis control distribution. It can be seen that under RW, the empirically observed MPC distributions for all three experiments are confounded, indicating independence of MPC with respect to the experiment. Furthermore, it is also possible to observe a bias—the distribution is skewed toward the maximum value of 1—introduced by (i) the maximization of the correlation through the projection of the data into another coordinate system, and (ii) the RW movement being low-pass filtered by the observer, hence there exists a correlation at longer time scales. Indeed, the distribution under RW is not symmetric about 0 as would be the case for mere correlation between variables of multivariate independent Gaussian processes. On the other hand, under the LJ condition the distribution skews even further to one, resulting in a high probability for MPC values near one, specifically in Necker and Cross. This is less so in Square.

When we removed fixations with detected micro-saccades and carried out inertia and MPC analyses, we found a difference for MPC in the LJ condition across experiments ($\chi^2 = 20.876; p < 0.0001$). When looking at pairwise comparisons (subscripts N for Necker, C for Cross, and S for Square), no significant differences were found between Necker and Cross ($Z_{N-C} = -1.6136; p = 0.106$), but Square differed from the other two ($Z_{S-C} = 3.4293; p = 0.0002$ and $Z_{N-S} = 4.1915; p < 0.0001$).

For RW inter-experiment comparisons, we found an overall difference ($\chi^2 = 10.617; p = 0.0036$). Paired comparisons showed a difference between Necker and the two other experiments ($Z_{N-C} = 2.955; p = 0.0020$ and $Z_{N-S} = -2.076; p = 0.0350$) but none for Square versus Cross ($Z_{S-C} = 1.061; p = 0.3114$).

Finally, results for individual analyses show that most participants in the Square experiment had no significant differences between MPC in RW and LJ, while on the contrary, all 23 participants in the Necker and 8 out 9 participants in Cross do.

Overall, these results indicate that Cross did replicate the micro-pursuit phenomenon observed in the Necker experiment even with a smaller sample size, while Square did not.

Median inertia with respect to the stimulus' center of gravity ($\bar{I}_{\text{stimulus}}$) differed with motion conditions suggesting that the nature of stimulus motion, manipulated in each condition (fixed, unpredictable, and predictable) affects global spatio-temporal dynamics of fixational eye movements. Median inertia with respect to the fixation's mean gaze position ($\bar{I}_{\text{fixation}}$) showed the emerging pattern of a common oculomotor phenomenon occurring in Necker and Cross, where differences across conditions were measured. Again, this was not the case in Square (see Fig. 3-C). When looking at similarity between stimulus and gaze trajectories, integrated over fixation events using MPC, we found that the predictable motion condition (LJ) generated highly similar gaze trajectories in the Necker and Cross experiments, with large effect sizes. But we did not observe the

same pattern for the Square experiment (see Fig. 3-E).

The contrast given by diverging results (Necker-LJ and Cross-LJ being different from Square-LJ) is interesting as it gives us a graduation of how likely, the same predictable motion (LJ) can make observers generates micro-pursuit. It also suggests that a coupling between the oculomotor and cognitive systems in the occurrence of micro-pursuits, which could be predicted and interpreted by a modeling framework we proposed when encountering the original observations. To go further, we propose a model, in **Appendix C**, that can describe all fixational eye movements in a single mechanism and can take into account the competition between multiple stimuli.

Discussion

Micro-pursuits

Given our definition of micro-pursuits (see section **Micro-pursuits**) which was based on an extrapolation of results available from the literature and our hypothesis of an oculomotor continuum, we have now gathered sufficient evidence to validate—at least, partially—our proposed working definition. We believe that the following characteristics are elementary building blocks in defining micro-pursuits as a class of oculomotor movements or fixational eye movements:

Tracking *or* similarity with target Probably the most prominent characteristic of micro-pursuits. When measuring similarity between the stimulus and gaze along the direction of maximum similarity using the MPC, we are able to categorize fixations as micro-pursuits, whether or not they contain micro-saccades. In addition, our proposed measure of similarity is invariant to scale, translation, and uncorrelated additive noise, compensating respectively for competition between fixation and tracking of a moving target as well as for instrumental or oculomotor drift and for acquisition noise. When the subject's gaze stays localized within the fixation (Square, all conditions), MPC indeed indicates that Square-LJ does no longer has gaze following up on the target motion contrary to Necker-LJ and Cross-LJ (see Fig. 3).

Velocity and acceleration Based on the literature review (Martins et al., 1985; Skinner et al., 2018) all velocities of our stimuli were kept below 2 deg.s^{-1}. At these velocities we have detected potential candidates of micro-pursuits, especially when the acceleration was moderate (LJ) Fig. 3 (MPC; Necker & Cross). When the acceleration was too high (RW[5]), micro-pursuits are no longer produced Fig. 3 (MPC; Necker, Cross, & Square). This advocates for the inclusion of both velocity and acceleration into the definition.

Binocularity Binocular conjugacy is an essential ingredient if micro-pursuit is to be interpreted as an expression of a central control over the oculomotor system. Our results show that micro-pursuits in our experiments appear as conjugated both at the group and at the individual level (see Fig. 4).

In contrast to the above, the following elements of our working definition are no longer retained in our final proposition for a definition of micro-pursuits:

Amplitude Given we focus solely on fixational eye movements, we found that there exists a category of movements that follow the below characteristics whilst staying under 1 deg in amplitude. However, if the oculomotor continuum holds, amplitude should no longer be a characteristic trait of (micro-)pursuit.

Duration Although initially thought to be a defining characteristic of micro-pursuits, duration is a mere operational limitation. Indeed, the oculomotor system exhibits mechanical inertia and is thus intrinsically limited in its velocity and acceleration, resulting in trajectories with long autocorrelation times. Hence, for short observation periods, one has insufficient variability to accurately estimate *similarity*, independent of the method used.

In this work we focused on a proof of micro-pursuits' existence through the results obtained from the Necker experiment as well as through results from the replication experiments (Cross or Square).

Although the above results are obtained retaining only fixations that do not contain any micro-saccades—as such being maximally conservative,—our conclusions generalize when we include fixations with micro-saccades. As far as the micro-saccades are concerned, our data (presented in Fig. 2) show main sequences that are invariant with respect to conditions and experiments. Furthermore, when looking at their marginal amplitude and peak velocity distributions, no clear differences can be observed across conditions and experiments. A similar observation can be made regarding their rate of occurrence. Also, micro-saccades within fixations that show pursuit behavior (high MPC values) show similar characteristics as those that are found in other fixations, since the MPC statistic does correlate with neither the peak velocity, nor amplitude of the micro-saccade under study. This provides evidence about the fact that slow fixational eye movements—tagged micro-pursuits in our work—can indeed be punctuated by micro-saccadic movements within a single fixation, and these do not interfere with the overall trend of the micro-pursuit movement. If the oculomotor continuum hypothesis indeed holds true, these micro-saccades could be associated with catch-up saccades. Unfortunately, due to our limited spatial resolution (video-based gaze tracker), we can not provide any further evidence for these.

Indeed, eye movement research is gradually considering an *oculomotor continuum*. For instance, it is becoming less and less credible to consider a hard separation between micro-saccades and saccades because of their common neural origins

[5] Acceleration is due to rapid changes both in the direction as in the magnitude of the velocity vector, due to additive white Gaussian perturbations of the latter.

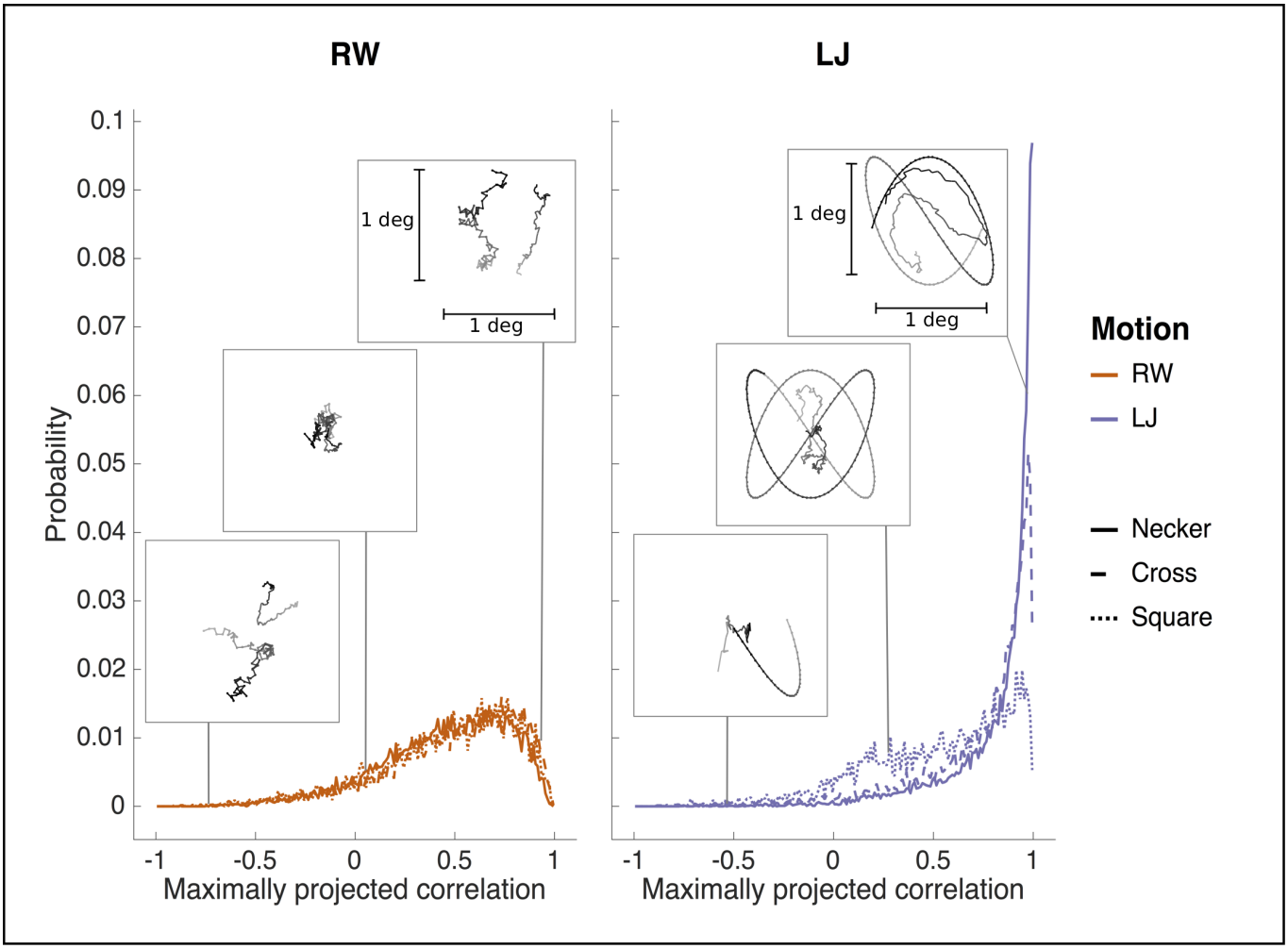


Figure 5: **Focus on MPC results.** Histogram of fixations by maximized correlation ρ^* (MPC) scores in the Necker Cube experiment. Illustrations of signals for values in some typical score intervals are presented to give a graphical intuition of the computed measure. We picked high similarity near a score of 1, no correlation near 0 and anti-correlation near -1. Dotted trajectories correspond to stimulus signals and continuous trajectories correspond to gaze signals. Temporal discourse is represented by lighter to darker samples.

in oculomotor programming (Krauzlis et al., 2017), their common properties, and mathematical models that can account for both (Sinn & Engbert, 2016). One may thus also consider that micro-pursuits share physical properties as well as neural correlates with large-amplitude smooth pursuits.

Alternative interpretations might classify fixations showing high inertia (w.r.t. fixation) as ocular drift. However, drift is considered independent from the stimulus, and hence should not showcase high values of MPC as in the Cross-LJ and Necker-LJ conditions.

A limitation of this work is that it does not explicitly contrast experimental stimuli that are known to generate pursuit versus OFR. Indeed, as presented in the introduction, OFR are reflexive eye movements generated using sudden changes of a wide-field image (Quaia et al., 2012) and should thus appear invariantly w.r.t. our experimental settings, but the lack of replication in the Square-LJ condition discredits this hypothesis.

A limitation of our similarity measure MPC resides mostly in its variance and thus the number of (temporally correlated) samples needed to accurately measure similarity. This is illustrated through Fig. 6 in **Appendix B**. While on the one hand, physical properties (amplitude, peak velocity) can be used to discriminate micro-pursuits from micro-saccades, on the other hand, functional characterization will help provide discrimination between drift, slow motor control, and micro-pursuit. Indeed, the first two may be slow fixational eye movements, but have no requirement for target tracking, like pursuit, whereas the latter does.

Finally, micro-pursuit's link to visual perception remains speculative, though interpreting our data suggests that the designation of the observed object, for perceptual report, and its associated motion (static, unpredictable or predictable)—related to the distribution of cognitive capacities between perceptual and oculomotor tasks—may lead to a tentative explanation (Spering & Montagnini, 2011).

Influence of oculomotor and perceptual tasks on target locking

In our two replication experiments, we have manipulated the task and target properties. In the Cross experiment, the task was to follow the moving object (cross) and to report its changes in illumination, while a static square was present in the background. In the Square experiment, the task was to fix a central fixation cross and report changes on the moving square object. Both have a similar relative movement of the cross with respect to the square object. In the first experiment, one can consider that participants had to focus on the cross. Whenever the latter was moving in a predictable, tractable fashion (LJ), the cross induced micro-pursuits. In the Square experiment, the competition between the perceptual and oculomotor tasks remained. Thus, one can consider the Square experiment to provide a competition between two attractors at the level of the oculomotor control, but given the reduced number of observed micro-pursuits (Fig. 3-E-F), one can interpret the competition between its attractors as unbalanced, where the fixation is more prominent than the follow-up on the moving target.

A first step towards a quantitative characterization of how a task may influence oculomotor dynamics is proposed in **Appendix C**. The proposed model is based on a competing attractor model inspired by gravitational field models. The model links the visual stimulation to perceptual objects modeled as gravitational attractors with dynamically varying masses, as such coping with the attention whereas gaze position is modeled through a unit-mass particle subject to the gravitational field evolving in time. To account for perturbations and noise, velocity is subject to additive white Gaussian noise (Langevin dynamics). By manipulating the attractor's positions, masses, and the curvature of their *energy potential*, it is possible to generate (micro-)saccades, (micro-)pursuits, fixations, and drift. This mathematical model offers a quantitative method that may be interpreted in terms of spatial attentional loads, saliency, or intention, with respect to oculomotor programming and execution. It is an extension of some models already proposed in the field of fixational eye movements modeling based on energy potential (Engbert, Mergenthaler, Sinn, & Pikovsky, 2011; Herrmann, Metzler, & Engbert, 2017) as well as modeling work on bi-stable perception and processes (Moreno-Bote, Rinzel, & Rubin, 2007; Shpiro, Moreno-Bote, Rubin, & Rinzel, 2009; Moreno-Bote, Knill, & Pouget, 2011; Moreno-Bote & Drugowitsch, 2015), to incorporate the influence of, e.g., ambiguous figures like the Necker cube.

Future work

We proposed to use a set of metrics to detect micro-pursuit, but we need further experimental work to define the limits, the functional role, and the specificity of micro-pursuit with respect to other fixational eye movements.

First, discrimination between OFR and micro-pursuit can be assessed by contrasting stimuli with a variety of targets and backgrounds, e.g., gratings (Gellman et al., 1990). One may contrast pursuit capacity between tracked motion applied to a background texture and a target in the foreground.

Second, interaction between saccade and pursuit needs to be further studied. This can be done by varying speed and predictability of the target trajectory. When increasing velocity of the target, and under the oculomotor continuum hypothesis, pursuit movements will get interleaved with catch-up saccades that compensate for the accumulated retinal error (drift). Beyond a certain speed limit, a sequence of saccades and erratic movements—similar to those observed in our random walk condition or in the proposed simulation model—should be observed, indicating that micro-pursuits falters beyond an upper bound velocity. However, we here attain the limits of our apparatus and more precise and accurate eye-tracking methods are needed to determine whether specific catch-up micro-saccades do occur in micro-pursuit, and in discriminating them from more generic micro-saccades.

Third, decreasing the predictability of the trajectories (increasing acceleration) will also lead to a transition from micro-pursuit over micro-pursuit interleaved with micro-saccades, to erratic movement. One possibility is to tune noise and inertia (mass) for a stimulus position driven by Langevin dynamics as for the particle in our model.

Furthermore, manipulating the scale of the motion could provide insight into micro-pursuit's link to large amplitude smooth pursuit characteristics, and may provide hints on its functional role.

Finally, the link between perception and oculomotor control of smooth pursuits have to be studied, e.g., by varying the relative difficulty of the task (i.e., report changes) or the difficulty of the tracking. This might help in explaining the absence of positive results with respect to smooth pursuits within the Square experiment.

Conclusion

In this work, micro-pursuits are proposed as a type of fixational eye movement occurring at small amplitude, within a fixation, as the gaze follows a target. We proposed two metrics: inertia and MPC to measure gaze displacement within a fixation and to quantify gaze-target trajectory similarity, respectively. We observed fixations in a predictable motion condition with higher gaze displacement, and more specifically, for both the Necker and Cross experiment data-sets, fixations with high gaze-stimulus similarity values under predictable target trajectories for position and velocity analyses. Binocular conjuga-

tion of the reported observations also provided evidence supporting the existence of micro-pursuit fixational eye movements. Micro-pursuit here is presented as a class of fixation, but further research is needed to identify the physical properties and distinguish it from other fixational eye movements. Moreover, this article calls for further investigation on the functional role of micro-pursuits, and how the oculomotor and perceptual systems interact during such movements.

Acknowledgments

We wish to thank Rubén Moreno-Bote, Laurent Madelain and the three anonymous reviewers for their helpful comments on earlier versions of this manuscript as they provided a fresh look for new analyses that led to key findings post-hoc. This research was partially funded by a grant from the LabEx PERSYVAL-Lab (ANR-11-LABX-0025-01).

Appendix A: notations used

In this work, we use $\boldsymbol{q}$ and $\dot{\boldsymbol{q}}$ for the two-dimensional position (in deg) and velocity vectors (in deg·s^{-1}). Subscripts R, G, and S will respectively refer to the retinal image, the gaze, and the stimulus. Over-lined notation will refer to the mean over a set of trials and a tilde to the median over a set of trials, for all metrics. Mean values will be reported with their standard deviation and median values with median absolute deviation (mad), for instance $\overline{\rho} = 2.93 \pm 0.01$ or $\widetilde{\rho} = 3.01 \pm 0.02$.

Appendix B: metrics

We use $\cdot^\top$ for the transpose operator and trace$(\cdot)$ will denote the trace operator (sum over the diagonal elements of a matrix). The identity matrix in dimension 2 will be denoted by Id_2.

Variance-covariance and inertia

Let $\boldsymbol{q}_G(t) = \left[x_G(t), y_G(t)\right]^\top$, $\boldsymbol{q}_S(t) = \left(x_S(t), y_S(t)\right)^\top$, and $\boldsymbol{q}_R = \left[x_R(t), y_R(t)\right]^\top$ be the screen Cartesian coordinates (column vectors) at time instant t of the Gaze, the stimulus, and the retinal image, respectively. Now, having samples at n discrete times $\{t_i\}_{i=1}^n$, we estimate the center of gravity of a gaze trajectory $\{\boldsymbol{q}_G(t_i)\}_{i=1}^n$ by its empirical mean $\boldsymbol{m}_G = n^{-1}\sum_{i=1}^n \boldsymbol{q}_G(t_i)$. This estimate approaches the true center of gravity if we sample sufficiently regularly and beyond twice the Nyquist frequency, conditions that are met when working with the Eyelink 1000(+), sampling at about 1000Hz for each eye.

A second-order statistic of interest is the empirical variance-covariance matrix, which gives the inertia of the gaze trajectory defined as

$$\boldsymbol{\Sigma}_G = n^{-1}\sum_{i=1}^n \boldsymbol{q}_G(t_i)\boldsymbol{q}_G^\top(t_i) - \boldsymbol{m}_G\boldsymbol{m}_G^\top$$

and analogously for the stimulus and retinal image empirical variance-covariance matrix. The inertia about its center of gravity is then given by $I_{\boldsymbol{m}_G} = n^{-1}\sum_{i=1}^n \|\boldsymbol{q}_G(t_i) - \boldsymbol{m}_G\|^2 = \mathrm{trace}\,(\boldsymbol{\Sigma}_G)$.

The inertia $I_{\boldsymbol{r}}$ of the gaze trajectory $\boldsymbol{q}_G$ with respect to any fixed point $\boldsymbol{r}$ having screen coordinates (x_r, y_r) is

$$I_{\boldsymbol{r}} = \mathrm{trace}\,(\boldsymbol{\Sigma}_G) + (\boldsymbol{m}_G - \boldsymbol{r})^\top(\boldsymbol{m}_G - \boldsymbol{r}) \quad .$$

Maximally projected correlations

Taking now the simultaneously recorded gaze $\{\boldsymbol{q}_G(t_i)\}_{i=1}^n$ and stimulus $\{\boldsymbol{q}_S(t_i)\}_{i=1}^n$ signals, and their respective empirical variance-covariance matrices $\boldsymbol{\Sigma}_G$ and $\boldsymbol{\Sigma}_S$. Denote the inter-covariance matrix by

$$\boldsymbol{\Sigma}_{GS} \doteq n^{-1}\sum_{i=1}^n \left(\boldsymbol{q}_G(t_i) - \boldsymbol{m}_G\right)\left(\boldsymbol{q}_S(t_i) - \boldsymbol{m}_S\right)^\top = \boldsymbol{\Sigma}_{SG}^\top \quad .$$

This matrix is particularly useful when considering the inertia of gaze with respect to the time-changing coordinates of the stimulus. Indeed, after some manipulations, we obtain:

$$I_{GS} = n^{-1}\sum_{i=1}^n \|\boldsymbol{q}_G(t_i) - \boldsymbol{q}_S(t_i)\|^2$$
$$I_{GS} = \mathrm{trace}\,(\boldsymbol{\Sigma}_G + \boldsymbol{\Sigma}_S - \boldsymbol{\Sigma}_{GS} - \boldsymbol{\Sigma}_{SG}) + \|\boldsymbol{m}_G - \boldsymbol{m}_S\|^2 \quad .$$

Unfortunately, the inertia does not account for differences in scale, nor for coordinate translation, two characteristics that are typical aspects for pursuits and for which we require an invariance[6].

Noise robustness & signal size dependency

Fig. 6 shows results of simulations operated on a Lissajous signal degraded by noise on the position of the stimulus at different signal to noise ratios (SNR) and for different signal sizes. For signals with more than 167 samples, the behavior of MPC scores over SNR remains stable and shows quasi-unchanged dynamics.

Appendix C: model

Models come in a variety of forms, depending on the mathematical framework used to formalize and compute their mechanics. Two main families can be differentiated: descriptive statistical and generative mechanistic models. Here, we focus on the latter. The motivation is the following: generative models can produce simulated and synthetic results that can be compared to observed empirical data. The model can then be studied and decomposed such that each internal force can be characterized, and their functional role in creating the analogous behavior can be investigated. All together, models remain key to understand a phenomenon and make predictions for empirical and experimental work. We focused here on fixational eye movements in

[6] Indeed, we suppose the stimulus will always be at a constant phase with respect to the gaze, either lacking behind in phase (catching up on the stimulus) or ahead of phase (prediction), the scale difference is our main objective, showing that the stimulus trajectory is reproduced at a smaller scale and, finally, the coordinate translation shows a systematic bias in the trajectories.

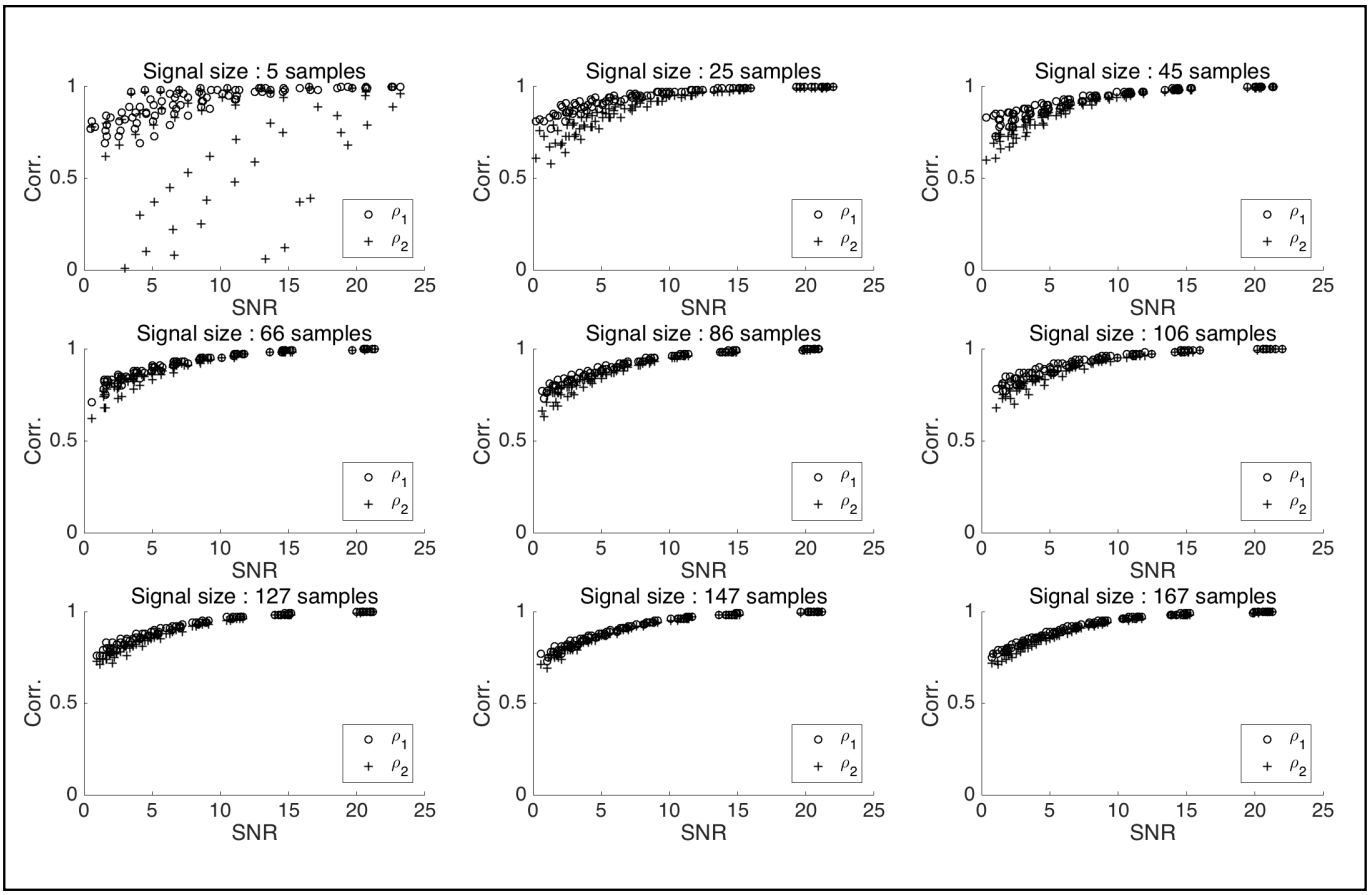


Figure 6: Behavior of MPC scores over signal to noise ratios (SNR) in simulated similarity computations with a Lissajous base signal from LJ, with varying signal sample sizes.

an attempt to explain and understand the data observed and reported in the article. Generative eye movement models use different approaches including, for instance, probabilistic models (Tatler, Hayhoe, Land, & Ballard, 2011; Gide, Karam, et al., 2017), accumulation process models (Orquin & Loose, 2013), or energy potential models (Engbert et al., 2011). Here we focus on the latter approach.

Recently, Engbert and colleagues (Engbert et al., 2011) proposed a generative model that could reproduce the statistical properties of fixational eye movements stationary displacement, namely the short term persistence and long term antipersistence of drift and tremors. They used a self-avoiding walk (Freund & Grassberger, 1992) in a discretized quadratic energy potential: at each iteration, the gaze, represented by a particle in the energy potential landscape, can either go left, right, up or down. The walker will choose the slot with the lowest energy. Once a step is made, the slot of the previous iteration is set to a high energy value, and the entire energy landscape follows a linear relaxation law. Hence, fixational eye movements bottom-up dynamics can be reproduced. Furthermore, the model also proposed to integrate micro-saccade generation by a threshold rule: when the particle is surrounded by options with energy higher than the threshold, it jumps to the global minimum of the energy landscape. Here, the authors provide an accumulation process linked to a global integration of the oculomotor field.

The integrated fixational eye movements model described above is a key foundation to bridge the oculomotor modeling communities and accounts for multiple fixational eye movement phenomena (e.g., drift displacement, micro-saccade, spatial orientation biases). However, it did not posses a mechanism to account for micro-pursuit, as these are hardly studied and reported. The observation of micro-pursuits presented in the article implies that the dynamics of the gaze within a fixation can be affected and attracted by motion of a perceptual object in or nearby the foveal field. Therefore, we propose modeling approach, gravitational fixational eye movements (GraFEM), inspired by gravitational energy field theory to model motion of eye movements and derived from the work on integrating fixational eye movements in energy potential models (Engbert et al., 2011).

Gravitational potential energy field modeling

Integrated and generative fixational eye movements models make use of energy potentials to generate self-avoiding walks, constrain the walks and replicate oculomotor biases (Engbert et al., 2011). In fact, the latter is used to constrain the pseudo random walk's spatial horizon. Furthermore, it can be considered as an attractor of the energy landscape. Thus, the use of the particle in an energy potential framework can be adjusted to provide biases of the stimulus on the fixational eye movement generation. Combining attractors in the energy fields, that increase the probabilities of having the gaze at some spatial coordinates, and adding stochasticity to the movement of the particle can provide a simple mechanism for fixational eye movement generation.

The attractors' properties can be manipulated over time to affect the energy field and thus dynamics of the fixational eye movements generated. The energy field that is mapped to the visual field can be populated by an arbitrary number of n attractors of varying strength (see Fig. 7-A). Inspired by the formalism of gravitational fields, one can generate fields with the following equations. Let Φ_i represent the field generated by the i^{th} attractor given by:

$$\Phi_i(\boldsymbol{q},t) = -\frac{1}{\|\boldsymbol{q}(t)-\boldsymbol{a}_i(t)\|^{2\beta_i(t)}+\delta_i(t)} \tag{4}$$

with $\boldsymbol{q}$ and $\boldsymbol{a}_i$ corresponding to the spatial x-y coordinates (at time t) of the observer's gaze position and the i^{th} attractor, respectively. The potential landscape can be fine tuned according to assumptions on attractive attributes of the stimulus and the tasks. First, it is necessary to set how many attractors are present and give them spatial coordinates in the plane over time. Secondly, it is possible to handle the mass of those attractors and their subsequent force of attraction and distortion of the field by tuning two parameters; δ for the depth of the well and β for the concavity of its slope. Summation and normalization of the field allow for the fusion of the multiple attractors.

$$\Phi(\boldsymbol{q},t) = \sum_{i=1}^{n}\Phi_i(\boldsymbol{q},t) \tag{5}$$

A logarithmic attenuation is added to allow the possibilities of exploring high energy areas of the visual/foveal field, giving the energy E:

$$E = -\ln(-\Phi) \tag{6}$$

Memory of attractor motion (Fig. 7-B) are modeled by adding a moving average (MA) process (Hannan, 2009) on the field at a given time t:

$$E_{FEM}(\boldsymbol{q},t) = E(\boldsymbol{q},t) + \sum_{k=1}^{K}\frac{\lambda}{k+1}E(\boldsymbol{q},t-k\Delta t) \tag{7}$$

where K is the temporal parameter limiting how far in time will the fields be summed over and with λ the relaxation rate parameter and Δt is the temporal step size. It is also possible to set the impact of memory and anticipation through parameters that define the iteration window over which the field is deformed using traces of the attractor in the past of a given current iteration and the rate λ at which the deformation affects for a given lag.

A particle of position ($\boldsymbol{q}$) with negligible mass (or with very high friction) is dropped in the field and is disturbed by noisy force, in order to generate and simulate gaze dynamics. Therefore, given the fundamental relation for dynamics, where the accelerating second order component is neglected, the gaze particle's motion is derived by the Langevin equation (Langevin, 1908), in which $m\ddot{\boldsymbol{q}}$ is equal to the sum of forces applied to the particle, and can be rewritten as follows:

$$m\ddot{\boldsymbol{q}} = -\gamma\dot{\boldsymbol{q}} - \nabla E_{FEM}(\boldsymbol{q},t) + \xi(t) \tag{8}$$

with m the negligible mass, γ the friction and where ξ is an external force, here an oculomotor noise (η) applied to the gaze, such that $\eta(t) = \frac{\xi(t)}{\gamma}$. With the assumption of low mass and after normalization[7], such that $E_{FEM} = \frac{E_{FEM}}{\gamma}$, the dynamics can be expressed as:

$$\dot{\boldsymbol{q}} = -\nabla E_{FEM}(\boldsymbol{q},t) + \eta(t) \tag{9}$$

The evolution of the gaze particle's dynamics can be computed by making the problem a discrete one using the Euler-Maruyama method (Kloeden & Platen, 2013), for instance.

Model simulations: what are the parameters corresponding to ocular events & interpretation?

Fixations of 1.5 seconds, with a discretization Euler-Maruyama step $\Delta t = 1$ ms equal to the time step, were simulated using the GraFEM model with two attractors, $\boldsymbol{a}_{\text{cross}}$ corresponding to the attractor of a fixation cross at the center and $\boldsymbol{a}_{\text{stim}}$, the attractor representing the stimulus, with a Lissajous motion: $\boldsymbol{a}_{\text{stim}} = (\sin(2t),\sin(3t))$. Only the slope and depth parameters were manipulated: $\beta_{\text{stim}} \in [0;50]$ and $\delta_{\text{stim}} \in [0;1200]$. All other parameters were kept constant with the other attractor position at $\boldsymbol{a}_{\text{cross}} = (0,0)$ with $\beta_{\text{cross}} = 1$ and $\delta_{\text{cross}} = 100$, the relaxation rate parameter $\lambda = 0.9$, the memory temporal limit $K = 5$ and noise $\xi \sim \mathcal{U}[-0.5;0.5]$. These simulated fixations were then analyzed using the measures presented in this article, namely, inertia, MPC and micro-saccade detection using the Engbert-Kliegl (EK) algorithm based on relative velocity thresholds (Engbert & Kliegl, 2003). Fig. 9-A shows that higher inertia follows a diagonal region along the $\{\beta_{\text{stim}},\delta_{\text{stim}}\}$ space. When looking at Fig. 9-B, one can see that the same area in the parameter space has systematically high MPC. Finally, the EK algorithm was applied (without the binocularity criterion) to measure detected micro-saccades, and summed over the time of a fixation. The results (Fig. 9-C) show that micro-saccades are detected when concavity is high due to a larger β_{stim} parameter.

[7] Note that in the next equation, we use E_{FEM} with the same notation as above, which is not exact writing though it simplifies reading. We refer to a normalized term by γ in the next equation.

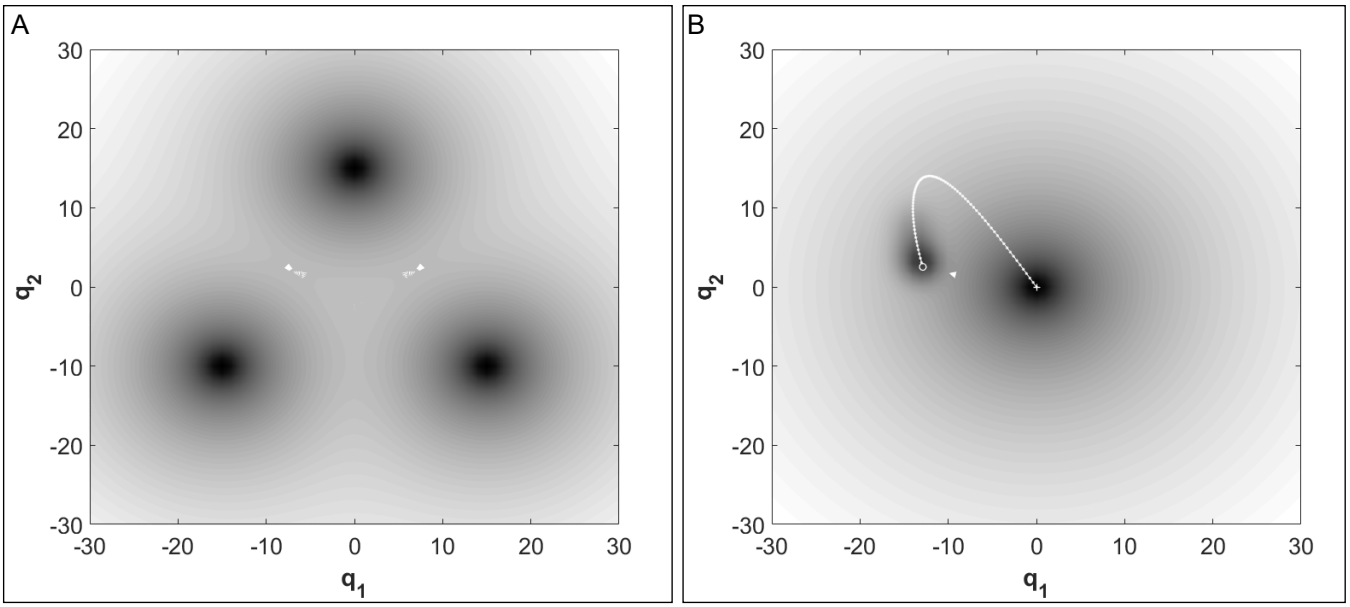


Figure 7: Examples of energy landscape surface plots setup using the gravitational fixational eye movements (GraFEM) model for: **A** shows 3 attractors ($n = 3$) with all attractors i having no motion and the following parameters: $\beta = 2$; $\delta = 1$; $\tau = 5$; $\lambda = 0.9$, and **B** shows 2 attractors ($n = 2$) with all attractors i having the following parameters: $\beta_1 = 2$; $\beta_2 = 4$; $\delta_1 = \delta_2 = 1$; $\tau = 15$; $\lambda = 0.9$ and attractor motion computed with the following arbitrary sinusoidal motion: $\boldsymbol{a}_1(t) = [0,0]$; $\boldsymbol{a}_2(t) = \boldsymbol{a}_2(t=0) + [-5\sin(2t), 5\sin(3t)]$ on the $75^t h$ iterations. The motion of $\boldsymbol{a}_2$ is shown in white. Though the model has many parameters, those manipulated in this work's results are exclusively the depth δ (or mass) of the attractors and the slope β by affecting the concavity of the attractors' field.

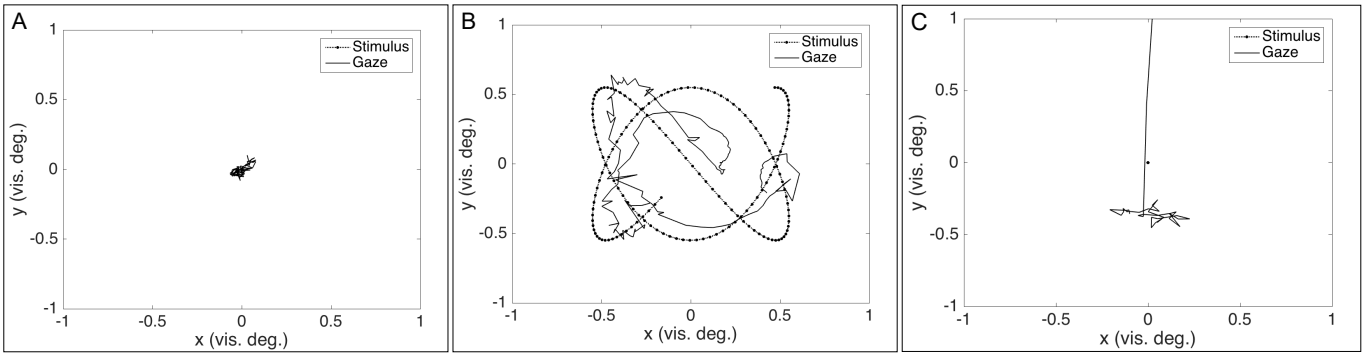


Figure 8: **Simulation examples generated with the GraFEM model.** Simulations of fixations of 3.5 seconds with Euler-Muruyama time steps of $\Delta t = 13$ ms, with variable fixation dynamics generated through by manipulating of δ_{stim} and β_{stim} parameters. Constant parameters of the model were: number of attractors $n = 2$, with one for the fixation cross $\boldsymbol{a}_{\text{cross}} = (0,0)$ and another for the motion of the stimulus following a Lissajous trajectories with the same parameters as in the three experiments (Necker, Square, Cross); $\boldsymbol{a}_{\text{stim}}(t) = \big(\sin(2t), \sin(3t)\big)$. The relaxation rate parameter $\lambda = 0.9$, memory temporal limit parameter $K = 5$ and noise $\xi \sim \mathcal{U}[-0.5; 0.5]$ were used. **A** shows a simulated fixation with stable fixation dynamics with $\delta_{\text{cross}} = 100$; $\delta_{\text{stim}} = 100$; $\beta_{\text{cross}} = 1$; $\beta_{\text{stim}} = 1$. **B** shows a simulated fixation with micro-pursuit dynamics with $\delta_{\text{cross}} = 100$; $\delta_{\text{stim}} = 25$; $\beta_{\text{cross}} = 1$; $\beta_{\text{stim}} = 1$. **C** shows a simulated fixation with micro-saccade dynamics with $\delta_{\text{cross}} = 100$; $\delta_{\text{stim}} = 25$; $\beta_{\text{cross}} = 1$; $\beta_{\text{stim}} = 12$ and detected using the EK algorithm for micro-saccade detection.

Discussion and perspectives: attractor, oculomotor and perceptual multi-stability

The simulation results presented above show the following three points. First, fixations' dynamics can be modeled including a variety of fixational eye movements such as drift, tremors, micro-saccades and micro-pursuits. Second, attractor dynamics can be intuitively manipulated by two parameters that control their slope and depth, hence imposing, by gravity, faster or

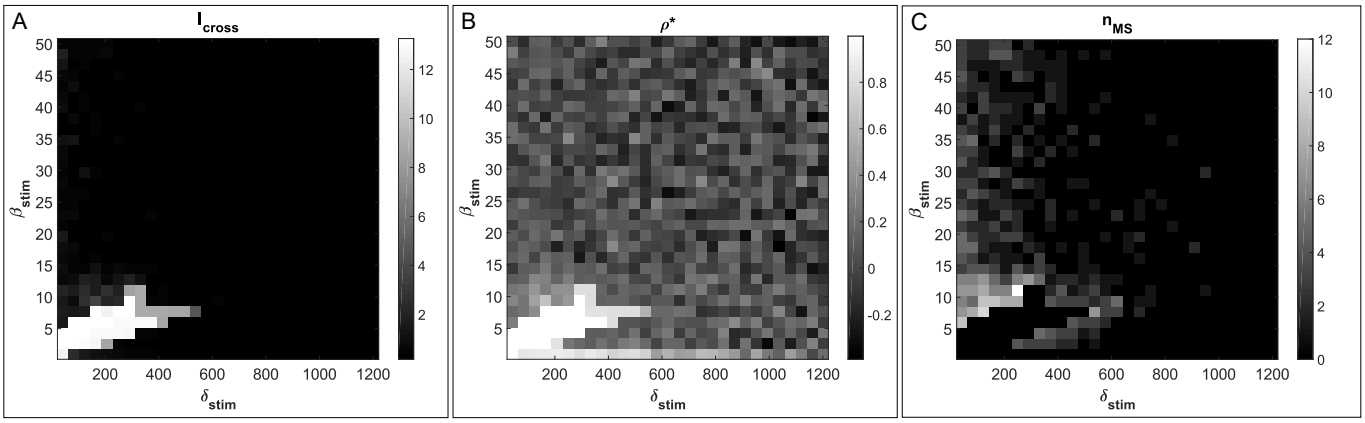


Figure 9: **Simulations and analyses of the GRAFEM model.** Simulations of fixations of 1.5 seconds with Euler-Muruyama time steps of $\Delta t = 1$ ms, with variable fixation dynamics generated through the variation of $\delta_{\text{stim}} \in [0;1200]$ and $\beta_{\text{stim}} \in [0;50]$ parameters. Constant parameters of the model were: number of attractors $n = 2$, with one for the fixation cross ($\boldsymbol{a}_{\text{cross}} = [0,0]$) and another for the motion of the stimulus following Lissajous trajectories with the same parameters as in the Necker cube experiment: $\boldsymbol{a}_{\text{stim}}(t) = \big(\sin(2t), \sin(3t)\big)$. The relaxation rate parameter $\lambda = 0.9$, memory temporal limit parameter $K = 5$ and noise $\xi \sim \mathcal{U}[-0.5;0.5]$ were used. **A** shows the behavior of inertia over the parameter space of the GraFEM model. **B** shows the behavior of the similarity between stimulus and simulated fixation motion using the MPC ρ_1. **C** shows the number of micro-saccades detected by the EK micro-saccade detection algorithm.

slower dynamics on the gaze-particle. Third, generalization to more complex stimuli or tasks can be maneuvered by such a model as attractors can be multiplied, if necessary. However, this work remains preliminary and calls for further investigation. Such perspectives are discussed in the following paragraphs.

Model interpretation for eye movements

The GraFEM model proposed in this paper is capable of generating micro-saccade, drift and tremor fixational eye movements (see Fig. 8) as classified in the literature (Martinez-Conde et al., 2004) as well as the micro-pursuits presented and detected in the article as reported in Fig. 9 & Fig. 10. By using classified data (observations), the parameters of the model that allow the generation of these fixational eye movements could be inferred, and insights on the mechanics of micro-saccades, micro-pursuit, drift and tremor generation and their interaction can be studied. The diagrams Fig. 10 already give a useful and overall understanding of the model, with respect to the manipulated parameters, but the work on parameter inference should be addressed in a near future in more details.

Given the observed data and the proposed model to account for it, questions and perspectives can be redefined with a novel angle for interpretation of fixational eye movements. Inversion and a full analysis of a model, like GraFEM, with multiple free parameters is a complex task out of the scope of this thesis but should be tackled and reported in a near future.

The model presented here gives a mathematical framework in which eye movement phenomena can be generated and interpreted. Attractors are interesting as tools to explain and interpret cognitive and physiological behaviors as they allow an intuitive understanding of the evolution of dynamical systems (T. Watanabe, Masuda, Megumi, Kanai, & Rees, 2014; Kelso, 2012). Furthermore, complex learning systems—i.e., neural networks—are known to develop such properties as the parameters of their processes tend to learn the statistics of the environment by creating attractors in the parameter space (Moreno-Bote et al., 2007; Shpiro et al., 2009; Moreno-Bote et al., 2011; Moreno-Bote & Drugowitsch, 2015).

With this modeling framework, the fixational eye movements classification of the literature can be described and interpreted in terms of attractor spatio-temporal dynamics (Fig. 9 & 10).

A stable fixation (Fig. 8-A) in the GraFEM model corresponds to a stabilization of an attractor with the energy landscape having little change. The gaze-particle is stuck and only the noise affecting its position may lead to small random movements of the eyes, as in other generative fixational eye movement models (Engbert et al., 2011; Herrmann et al., 2017). In these models, constraints to the energy field of the fixation are used in an analogous fashion to reflect the higher probabilities of having fixational eye movements in horizontal and vertical directions. A fixation attractor can thus be predicted by the task or the stimulus controlled by the experiments, and its parameters can be inferred by *a priori* information and data. Hence, the model gives predictive capabilities that can be tested and requires assimilation of data to constrain its range of possibilities.

Micro-saccades (Fig. 8-C) correspond to sudden changes in the energy depth of attractors, with a new one emerging or deepening while the attractor of fixation has suddenly disappeared. They are likely to emerge as the gaze-particle rushes down a gra-

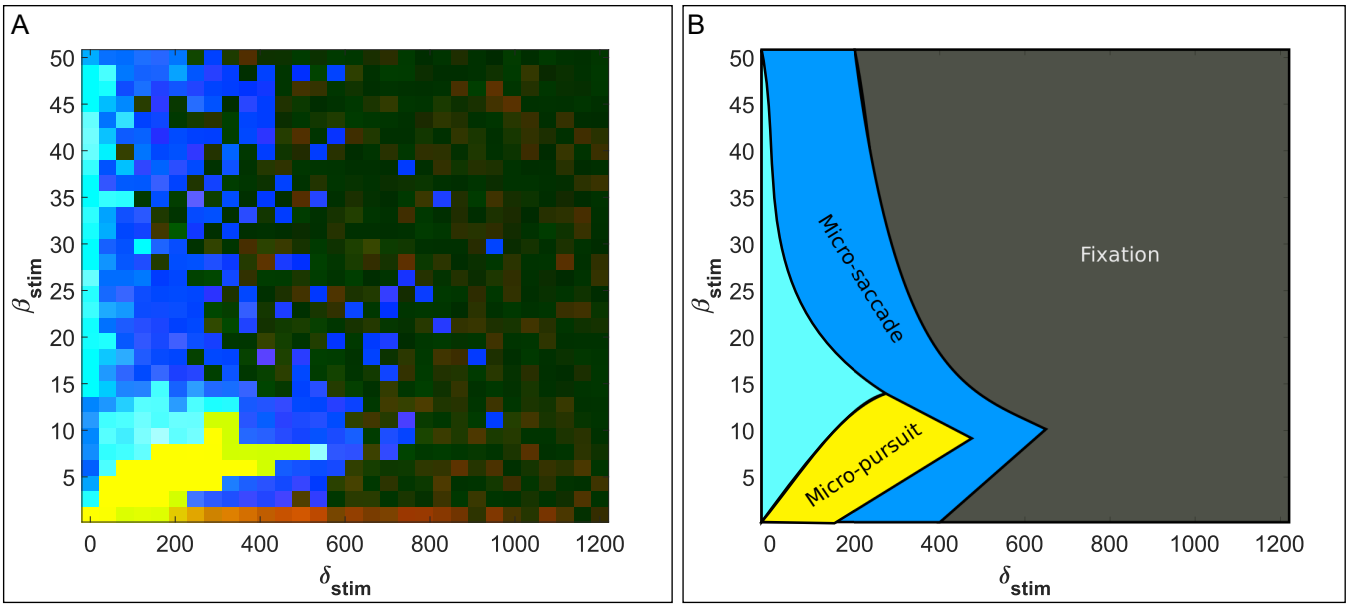


Figure 10: **GraFEM oculomotor interpretation. A** is a merger of measures applied on GraFEM synthetic data. Simulated fixations' inertia, MPC and detected micro-saccades results are assigned to the red, green and blue components of this RGB matrix, respectively. The yellow space provides β and δ values that generate micro-pursuit, the blue pixels show a micro-saccade generation transition space, the dark area shows stable fixation parameters while the red pixels could be interpreted as slow control. **B** is a schematic interpretation of oculomotor dynamics generated in the parameter space of the GraFEM model with manipulation of the δ_{stim} and β_{stim} parameters of the stimulus attractor, while keeping all other parameters constant. Micro-pursuit can be generate in a restrained subspace (yellow) while micro-saccades are detected in the surrounding space (dark blue). When δ_{stim} values are low, a transition area (light blue) exist where micro-saccades and similarity are high, but inertia is not. Finally, stable fixation sub-space occupy the rest (gray).

dient to the center of an attractor, giving it sufficient velocity. The depth and slope of the attractor can be manipulated (following the dynamics described in Fig. 10), thus making it possible to infer, based on observed velocities and amplitudes, the saliency of that attractor. The GraFEM model does not use an explicit and separated mechanism for micro-saccade generation—as the model presented in (Engbert et al., 2011)—though it is not incompatible.

Drifts correspond to a stability of the gaze-particle with respect to the attractor by which it is transported. However, the attractor might itself slowly drift away in the visual space (independently from the target motion) or alternatively, the shape of the well might get larger (by manipulating the parameter β), allowing for the noisy gaze-particle to explore further. These are two hypotheses that could be tested, in future work, by inferring the model parameters given sufficient data. These fixational eye movements are known to help reduce visual redundancy and extract features in complex visual stimuli (Kuang, Poletti, Victor, & Rucci, 2012) but are mostly considered to be consequences of the eye muscles and their neural control properties. Therefore, they have mostly been considered as independent processes from the visual stimulus presented.

The micro-pursuits detected and described in the article could be interpreted as a form of stimulus related drift, as its signal dynamics place it in similar ranges, and is capture by the proposed metric; namely MPC. Consequently, this argues in favor of our proposition that drifts are composed of two categories — stimulus independent and dependent — and micro-pursuits logically fall within visually dependent ocular drifts. This dependency can be interpreted as the interference of bottom-up salient elements interrupting the top-down task of fixation. Micro-pursuits (Fig. 8-B) are therefore close to drifts in the energy landscape dynamics.

Model mechanics

The model sets the gaze as a particle in an evolving gravitational energy potential field. When the system has no dynamics added to the potentials' landscapes, the particle will fall into its nearest local minimum. In this implementation, at each iteration—here a discrete time step using Equation (9),—the first derivative is computed to update the position of the particle in the plane, corresponding to the screen. Noise is then added to the deterministic dynamics and can drive fixational oculomotor decision-making with respect to attractors if its amplitude is sufficiently large (Shpiro et al., 2009; Moreno-Bote et al., 2007). This mechanism is similar to bi-stable energy potential models,

though it extends on the dimensions of the system. In a set of simulations reported in Fig. 9, we show that through two continuous parameters applied to a target attractor, it is possible to generate and interpret oculomotor dynamics observed in fixational eye movements. However, here, there is no prior requiring the existence of different systems for each class of movements observed (Liversedge et al., 2011). fixational eye movement dynamics can be reproduced through a unique mechanism as shown by the simulated examples in Fig. 8.

Top-down intention processes can be tested and simulated, given the context of a task, by applying changes in the model's β and δ parameters. Fig. 10 can be used as a road map of the oculomotor dynamics and regimes expected, depending on parameter values. Moreover, bottom-up saliency or attentional effects can also be taken into account. This can be done with simpler assumptions, such as the ones presented here for the task used in the article, but can be more complex if using natural scene tasks, for instance. An interesting and practical perspective in this context lays in investigating how salience models, which derive probability distribution based on the statistics of images or videos, can be integrated such that only attractors are fed into a GraFEM oculomotor execution system.

How would this be implemented in the brain?

Anatomically, oculomotor programming has been shown to be highly correlated and linked to a network of areas involving neural activity in the superior colliculus (SC), the frontal eye field (FEF) and the lateral intra-parietal (LIP) cortex (Hafed, Goffart, & Krauzlis, 2009; Krauzlis, 2004, 2005; Krauzlis et al., 2017; Astrand, Ibos, Duhamel, & Hamed, 2015; Peel, Hafed, Dash, Lomber, & Corneil, 2016; Taouali, Goffart, Alexandre, & Rougier, 2015). There are inter-individual differences in anatomy and behavior for fixational eye movements measuring and observed dynamics. For instance, it has been shown that not only oculomotor behavior between trained and untrained participants vary a lot, but that drift accounts for more fixation correction motion than micro-saccades (Cherici, Kuang, Poletti, & Rucci, 2012). The observations of micro-pursuits presented in the article suggest that the dynamics of the gaze within a fixation can be affected and attracted by motion of an object in or nearby the foveal field.

However, rather than having an attractor with a pseudo-random displacement, its motion follows a deterministic and predictable trajectory, that can be computed and estimated by the oculomotor system. Moreover, that attractor is, given our observations so far, only related to a target motion. This could, for instance, be implemented in the brain by the means of an efference copy (Astrand et al., 2015), though this idea remains speculative and further modeling and neuro-physiological research is needed. The low energy attributed to a decoded and perceived object moving across space encourages the oculomotor system to track it as it tries to minimize the energy of the gaze-particle. Finally, tremors are generated and explained by the noise given to the particle over all fixational eye movement events.

This model complements the eye movement field of research with the possibility to program intentions, salience, and their effects on the gaze dynamics by simply using attractors and setting out their dynamics in terms of motion on the visual field, depth and memory. For instance, the model can predict the different dynamics reported based on the eccentricity of an attractor corresponding to an afterimage, as observed in (Heywood & Churcher, 1972). Thus, one can use the model to generate statistical predictions of eye movement dynamics. Given an understanding of the visual attention or saliency effects of their stimulus and take into account all the associated intentions to the tasks that participants are required to be operated during a trial, it is possible to use this modeling to generate quantitative predictions on the oculomotor dynamics. Moreover, the generative properties makes it possible to work on simulated data and extract dynamics' statistics in terms of eye movements, and this is possible using the traditional algorithms for eye movements classification. Inversely, obtaining the parameters of the model that replicate the dynamics of observations could help understand better the internal processes that drive eye movements.

Perspectives: towards oculomotor multi-stability.

A key aspect of this family of models is that it showcases *multi-stability* regarding their attractors. This phenomenon can emerge in many complex biological systems and is present in many cognitive processes (Schwartz, Grimault, Hupé, Moore, & Pressnitzer, 2012). It is linked to coordination dynamics between sub-systems which have varying levels of coupling, leading to mono-stable, multi-stable or meta-stable dynamics (Kelso, 2012). The consequent interpretation is that the oculomotor system could have multi-stable dynamics with respect to visual attractors. In this case, the oculomotor dynamics are likely driven by noisy signals (J. Braun & Mattia, 2010) representing other interfering systems, such as perception, attention, intention, and other cognitive systems. This framework connects to the growing body of studies linking perceptual decisions and multi-stable system dynamics. It also creates a link for motor systems to studies of noise as a component that helps a perceptual system operate through stochastic resonance[8] (Gammaitoni, Hänggi, Jung, & Marchesoni, 1998; Patel & Kosko, 2005; Kim, Grabowecky, & Suzuki, 2006).

References

Anderson, T. (2003). *An introduction to multivariate statistical analysis.* Hoboken, New-Jersey, USA: John Wiley & Sons.

Astrand, E., Ibos, G., Duhamel, J.-R., & Hamed, S. B. (2015). Differential dynamics of spatial attention, position, and color coding within the parietofrontal network. *Journal of Neuroscience, 35*(7), 3174–3189.

Bahill, A. T., Clark, M. R., & Stark, L. (1975). The main sequence,

[8] Stochastic resonance refers to phenomena in which a system is able to detect a weak signal because noise boosts it, by providing the energy needed for the signal's frequencies to resonate mutually.

a tool for studying human eye movements. *Mathematical Biosciences*, *24*(3-4), 191–204.

Behrens, F., MacKeben, M., & Schröder-Preikschat, W. (2010). An improved algorithm for automatic detection of saccades in eye movement data and for calculating saccade parameters. *Behavior research methods*, *42*(3), 701–708.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial vision*, *10*(4), 433–436.

Braun, D. I., Pracejus, L., & Gegenfurtner, K. R. (2006). Motion aftereffect elicits smooth pursuit eye movements. *Journal of Vision*, *6*(7), 1–1.

Braun, J., & Mattia, M. (2010). Attractors and noise: twin drivers of decisions and multistability. *Neuroimage*, *52*(3), 740–751.

Chen, C.-Y., & Hafed, Z. M. (2013). Postmicrosaccadic enhancement of slow eye movements. *Journal of Neuroscience*, *33*(12), 5375–5386.

Cherici, C., Kuang, X., Poletti, M., & Rucci, M. (2012). Precision of sustained fixation in trained and untrained observers. *Journal of Vision*, *12*(6), 31–31.

Choe, K. W., Blake, R., & Lee, S.-H. (2016). Pupil size dynamics during fixation impact the accuracy and precision of video-based gaze estimation. *Vision research*, *118*, 48–59.

Cornsweet, T. N. (1956). Determination of the stimuli for involuntary drifts and saccadic eye movements. *JOSA*, *46*(11), 987–993.

Cunitz, R. J. (1970). Relationship between slow drift and smooth pursuit eye movements.

De Brouwer, S., Yuksel, D., Blohm, G., Missal, M., & Lefèvre, P. (2002). What triggers catch-up saccades during visual tracking? *Journal of Neurophysiology*, *87*(3), 1646–1650.

Ditchburn, R., & Ginsborg, B. (1953). Involuntary eye movements during fixation. *The Journal of physiology*, *119*(1), 1–17.

Dodge, R. (1907). An experimental study of visual fixation. *The Psychological Review: Monograph Supplements*, *8*(4), i.

Engbert, R., & Kliegl, R. (2003). Microsaccades uncover the orientation of covert attention. *Vision research*, *43*(9), 1035–1045.

Engbert, R., & Kliegl, R. (2004). Microsaccades keep the eyes' balance during fixation. *Psychological science*, *15*(6), 431–431.

Engbert, R., & Mergenthaler, K. (2006). Microsaccades are triggered by low retinal image slip. *Proceedings of the National Academy of Sciences*, *103*(18), 7192–7197.

Engbert, R., Mergenthaler, K., Sinn, P., & Pikovsky, A. (2011). An integrated model of fixational eye movements and microsaccades. *Proceedings of the National Academy of Sciences*, 201102730.

Epelboim, J., & Kowler, E. (1993). Slow control with eccentric targets: evidence against a position-corrective model. *Vision research*, *33*(3), 361–380.

Fang, Y., Gill, C., Poletti, M., & Rucci, M. (2018). Monocular microsaccades: Do they really occur? *Journal of vision*, *18*(3), 18–18.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using g* power 3.1: Tests for correlation and regression analyses. *Behavior research methods*, *41*(4), 1149–1160.

Freund, H., & Grassberger, P. (1992). The red queen's walk. *Physica A: Statistical Mechanics and its Applications*, *190*(3-4), 218–237.

Gammaitoni, L., Hänggi, P., Jung, P., & Marchesoni, F. (1998). Stochastic resonance. *Reviews of modern physics*, *70*(1), 223.

Gellman, R., Carl, J., & Miles, F. (1990). Short latency ocular-following responses in man. *Vis Neurosci*, *5*(2), 107–122.

Gide, M. S., Karam, L. J., et al. (2017). Computational visual attention models. *Foundations and Trends® in Signal Processing*, *10*(4), 347–427.

Hafed, Z. M., Goffart, L., & Krauzlis, R. J. (2009). A neural mechanism for microsaccade generation in the primate superior colliculus. *science*, *323*(5916), 940–943.

Hannan, E. J. (2009). *Multiple time series* (Vol. 38). John Wiley & Sons.

Heinen, S. J., & Watamaniuk, S. N. (1998). Spatial integration in human smooth pursuit. *Vision research*, *38*(23), 3785–3794.

Herrmann, C. J., Metzler, R., & Engbert, R. (2017). A self-avoiding walk with neural delays as a model of fixational eye movements. *Scientific Reports*, *7*(1), 12958.

Heywood, S., & Churcher, J. (1971). Eye movements and the afterimage—i. tracking the afterimage. *Vision Research*, *11*(10), 1163–1168.

Heywood, S., & Churcher, J. (1972). Eye movements and the after-image—ii the effect of foveal and non-foveal afterimages on saccadic behaviour. *Vision research*, *12*(5), 1033–1043.

Hicheur, H., Zozor, S., Campagne, A., & Chauvin, A. (2013). Microsaccades are modulated by both attentional demands of a visual discrimination task and background noise. *Journal of vision*, *13*(13), 18–18.

Hothorn, T., Hornik, K., Van De Wiel, M. A., & Zeileis, A. (2006). A lego system for conditional inference. *The American Statistician*, *60*(3), 257–263.

Ilg, U. J. (1997). Slow eye movements. *Progress in neurobiology*, *53*(3), 293–329.

Kassambara, A. (2020). rstatix: Pipe-friendly framework for basic statistical tests [Computer software manual]. Available from https://cran.r-project.org/web/packages/rstatix/index.html (R package version 0.5.0)

Kelso, J. S. (2012). Multistability and metastability: understanding dynamic coordination in the brain. *Phil. Trans. R. Soc. B*, *367*(1591), 906–918.

Kim, Y.-J., Grabowecky, M., & Suzuki, S. (2006). Stochastic resonance in binocular rivalry. *Vision research*, *46*(3), 392–406.

Kloeden, P. E., & Platen, E. (2013). *Numerical solution of stochas-*

tic differential equations (Vol. 23). Springer Science & Business Media.

Ko, H.-k., Poletti, M., & Rucci, M. (2010). Microsaccades precisely relocate gaze in a high visual acuity task. *Nature neuroscience*, *13*(12), 1549.

Ko, H.-k., Snodderly, D. M., & Poletti, M. (2016). Eye movements between saccades: Measuring ocular drift and tremor. *Vision research*, *122*, 93–104.

Komogortsev, O. V., & Karpov, A. (2013). Automated classification and scoring of smooth pursuit eye movements in the presence of fixations and saccades. *Behavior research methods*, *45*(1), 203–215.

Kowler, E. (2011). Eye movements: The past 25 years. *Vision research*, *51*(13), 1457–1483.

Kowler, E., Martins, A. J., & Pavel, M. (1984). The effect of expectations on slow oculomotor control—iv. anticipatory smooth eye movements depend on prior target motions. *Vision research*, *24*(3), 197–210.

Kowler, E., & Steinman, R. M. (1979a). The effect of expectations on slow oculomotor control—ii. single target displacements. *Vision research*, *19*(6), 633–646.

Kowler, E., & Steinman, R. M. (1979b). The effect of expectations on slow oculomotor control—ii. single target displacements. *Vision research*, *19*(6), 633–646.

Kowler, E., & Steinman, R. M. (1981). The effect of expectations on slow oculomotor control—iii. guessing unpredictable target displacements. *Vision research*, *21*(2), 191–203.

Krauzlis, R. J. (2004). Recasting the smooth pursuit eye movement system. *Journal of neurophysiology*, *91*(2), 591–603.

Krauzlis, R. J. (2005). The control of voluntary eye movements: new perspectives. *The Neuroscientist*, *11*(2), 124–137.

Krauzlis, R. J., Goffart, L., & Hafed, Z. M. (2017). Neuronal control of fixation and fixational eye movements. *Phil. Trans. R. Soc. B*, *372*(1718), 20160205.

Kuang, X., Poletti, M., Victor, J. D., & Rucci, M. (2012). Temporal encoding of spatial information during active visual fixation. *Current Biology*, *22*(6), 510–514.

Langevin, P. (1908). Sur la théorie du mouvement brownien. *Compt. Rendus de l'académie des sciences de Paris*, *146*, 530–533.

Liversedge, S., Gilchrist, I., & Everling, S. (2011). The oxford handbook of eye movements. In (p. 115-132). Oxford University Press.

Madelain, L., & Krauzlis, R. J. (2003). Pursuit of the ineffable: perceptual and motor reversals during the tracking of apparent motion. *Journal of Vision*, *3*(11), 1–1.

Martinez-Conde, S., Macknik, S. L., & Hubel, D. H. (2004). The role of fixational eye movements in visual perception. *Nature Reviews Neuroscience*, *5*(3), 229.

Martins, A. J., Kowler, E., & Palmer, C. (1985). Smooth pursuit of small-amplitude sinusoidal motion. *JOSA A*, *2*(2), 234–242.

Masson, G. S., & Stone, L. S. (2002). From following edges to pursuing objects. *Journal of neurophysiology*, *88*(5), 2869–2873.

Michalski, A., Kossut, M., & Żernicki, B. (1977). The ocular following reflex elicited from the retinal periphery in the cat. *Vision research*, *17*(6), 731–736.

Mihali, A., Opheusden, B. van, & Ma, W. J. (2017). Bayesian microsaccade detection. *Journal of vision*, *17*(1), 13–13.

Miles, F., Kawano, K., & Optican, L. (1986). Short-latency ocular following responses of monkey. i. dependence on temporospatial properties of visual input. *Journal of neurophysiology*, *56*(5), 1321–1354.

Moreno-Bote, R., & Drugowitsch, J. (2015). Causal inference and explaining away in a spiking network. *Scientific reports*, *5*, 17531.

Moreno-Bote, R., Knill, D. C., & Pouget, A. (2011). Bayesian sampling in visual perception. *Proceedings of the National Academy of Sciences*, *108*(30), 12491–12496.

Moreno-Bote, R., Rinzel, J., & Rubin, N. (2007). Noise-induced alternations in an attractor network model of perceptual bistability. *Journal of neurophysiology*, *98*(3), 1125–1139.

Muirhead, R. J. (2009). *Aspects of multivariate statistical theory* (Vol. 197). John Wiley & Sons.

Murphy, B. J., Kowler, E., & Steinman, R. M. (1975). Slow oculomotor control in the presence of moving backgrounds. *Vision research*, *15*(11), 1263–1268.

Nachmias, J. (1961). Determiners of the drift of the eye during monocular fixation. *Josa*, *51*(7), 761–766.

Nyström, M., & Holmqvist, K. (2010). An adaptive algorithm for fixation, saccade, and glissade detection in eyetracking data. *Behavior research methods*, *42*(1), 188–204.

Orquin, J. L., & Loose, S. M. (2013). Attention and choice: A review on eye movements in decision making. *Acta psychologica*, *144*(1), 190–206.

Otero-Millan, J., Macknik, S. L., Langston, R. E., & Martinez-Conde, S. (2013). An oculomotor continuum from exploration to fixation. *Proceedings of the National Academy of Sciences*, *110*(15), 6175–6180.

Palmer, S. E. (1999). *Vision science: Photons to phenomenology*. MIT press.

Patel, A., & Kosko, B. (2005). Stochastic resonance in noisy spiking retinal and sensory neuron models. *Neural Networks*, *18*(5-6), 467–478.

Peel, T. R., Hafed, Z. M., Dash, S., Lomber, S. G., & Corneil, B. D. (2016). A causal role for the cortical frontal eye fields in microsaccade deployment. *PLoS biology*, *14*(8), e1002531.

Poletti, M., Listorti, C., & Rucci, M. (2010). Stability of the visual world during eye drift. *Journal of Neuroscience*, *30*(33), 11143–11150.

Poletti, M., & Rucci, M. (2016). A compact field guide to the study of microsaccades: Challenges and functions. *Vision research*, *118*, 83–97.

Quaia, C., Sheliga, B. M., FitzGibbon, E. J., & Optican, L. M. (2012). Ocular following in humans: Spatial properties. *Journal of Vision*, *12*(4), 13–13.

Ratliff, F., & Riggs, L. A. (1950). Involuntary motions of the eye

during monocular fixation. *Journal of experimental psychology*, *40*(6), 687.

Rolfs, M. (2009). Microsaccades: small steps on a long way. *Vision research*, *49*(20), 2415–2441.

Rolfs, M., Kliegl, R., & Engbert, R. (2008). Toward a model of microsaccade generation: The case of microsaccadic inhibition. *Journal of vision*, *8*(11), 5–5.

Schwartz, J.-L., Grimault, N., Hupé, J.-M., Moore, B. C., & Pressnitzer, D. (2012). *Multistability in perception: binding sensory modalities, an overview.* The Royal Society.

Shpiro, A., Moreno-Bote, R., Rubin, N., & Rinzel, J. (2009). Balance between noise and adaptation in competition models of perceptual bistability. *Journal of computational neuroscience*, *27*(1), 37.

Sinn, P., & Engbert, R. (2016). Small saccades versus microsaccades: Experimental distinction and model-based unification. *Vision research*, *118*, 132–143.

Skinner, J., Buonocore, A., & Hafed, Z. M. (2018). Transfer function of the rhesus macaque oculomotor system for small-amplitude slow motion trajectories. *Journal of neurophysiology*, *121*(2), 513–529.

Spering, M., & Montagnini, A. (2011). Do we track what we see? common versus independent processing for motion perception and smooth pursuit eye movements: A review. *Vision research*, *51*(8), 836–852.

Stone, L. S., Beutter, B. R., & Lorenceau, J. (2000). Visual motion integration for perception and pursuit. *Perception*, *29*(7), 771–787.

Taouali, W., Goffart, L., Alexandre, F., & Rougier, N. P. (2015). A parsimonious computational model of visual target position encoding in the superior colliculus. *Biological cybernetics*, *109*(4-5), 549–559.

Tatler, B. W., Hayhoe, M. M., Land, M. F., & Ballard, D. H. (2011). Eye guidance in natural vision: Reinterpreting salience. *Journal of vision*, *11*(5), 5–5.

Tomczak, M., & Tomczak, E. (2014). The need to report effect size estimates revisited. an overview of some recommended measures of effect size.

Watanabe, M., Okada, K.-i., Hamasaki, Y., Funamoto, M., Kobayashi, Y., MacAskill, M., et al. (2019). Ocular drift reflects volitional action preparation. *European Journal of Neuroscience*, *50*(2), 1892–1910.

Watanabe, T., Masuda, N., Megumi, F., Kanai, R., & Rees, G. (2014). Energy landscape and dynamics of brain activity during human bistable perception. *Nature communications*, *5*, 4765.

Wyatt, H. J. (2010). The human pupil and the use of video-based eyetrackers. *Vision research*, *50*(19), 1982–1988.

Xivry, J.-J. Orban de, & Lefevre, P. (2007). Saccades and pursuit: two outcomes of a single sensorimotor process. *The Journal of physiology*, *584*(1), 11–23.

Yarbus, A. L. (1967). Eye movements during perception of complex objects. In *Eye movements and vision* (pp. 171–211). Springer.