



*Scene Statistics: Neural Representation of Real-world Structure in Rapid Visual Perception*

I.I.A. Groen

## Nederlandse samenvatting

behorende bij het proefschrift

### **Scene statistics: neural representation of real-world structure in rapid visual perception**

door Iris Groen

Vanaf het moment dat we onze ogen openen, begint ons brein met het verwerken van visuele informatie. De ooglenzen projecteert licht uit onze omgeving op het netvlies achter in het oog, dat het licht omzet in een neurale signaal. Het brein vertaalt dit neurale signaal vervolgens bliksemsnel in een mentaal beeld: een neurale representatie van de buitenwereld. Maar hoe werkt dit proces eigenlijk?

Deze vraag staat centraal in de visuele neurowetenschappen, een wetenschappelijke discipline die door middel van experimenten met mensen en dieren inzicht probeert te krijgen in de wonderbaarlijke vertaling van licht in neurale activiteit. Zo weten we nu welke delen van de hersenen betrokken zijn bij visuele perceptie. Er zijn hersengebieden die op simpele beeldelementen reageren zoals lijnen en kleuren, terwijl andere delen juist actief worden als bepaalde objecten zichtbaar zijn, zoals gezichten (zie Figuur 1.1A in de Inleiding van dit proefschrift). Samen zorgen deze hersengebieden er op een of andere manier voor dat alle informatie samenkomt tot een compleet mentaal beeld van onze omgeving.

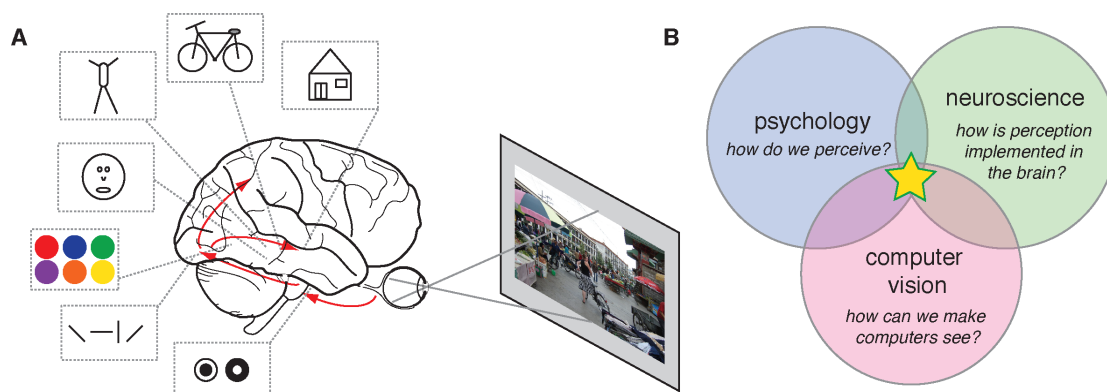
Om dit psycho-biologische proces goed te begrijpen, is echter belangrijk om niet alleen te onderzoeken *waar* in het brein activiteit plaatsvindt, maar ook *hoe* dit gebeurt. Welke taal, welke 'neurale code' gebruikt het brein om visuele informatie te representeren? Wat voor berekeningen voert het uit om de informatie te integreren tot een samenhangend, coherent beeld? Dit proefschrift gaat over een specifiek type visuele informatie, *scene statistics*, ofwel fysieke wetmatigheden in onze omgeving. De hoofdvraag is of deze informatie betrokken is bij het vormgeven van neurale representaties. De onderzochte stelling is dat het brein deze wetmatigheden oppikt en gebruikt om visuele informatie op een efficiënte manier te verwerken, zodat het snel een beslissing kan maken over de inhoud van een beeld.

#### *De uitdaging: het verklaren van supersnelle waarneming*

Al in de jaren zeventig kwamen onderzoekers erachter dat visuele perceptie in het brein enorm snel gaat. Mary Potter (1975) voerde experimenten uit om te bepalen hoeveel tijd we nodig hebben om visuele informatie te verwerken. Zij liet op een projector - er waren toen nog geen computers - een aantal foto's zien die elkaar heel snel opvolgden. De taak van de proefpersonen was om een van tevoren opgegeven beeld te zoeken in de reeks. Soms werd het beeld in kwestie eerst aan hen getoond, maar soms kregen ze alleen een verbale instructie, bijvoorbeeld "zoek twee mannen die bier drinken". Uit de resultaten van het experiment bleek dat zelfs wanneer de foto's maar in een flits te zien waren, men de beelden nog makkelijk kon vinden - ook als de instructie alleen verbaal, en dus meer abstract was. Hieruit kunnen we concluderen er maar een klein beetje tijd nodig is om de betekenis van de visuele beelden te bepalen. Nu kunnen we dit experiment zo nadoen op onze eigen telefoon: zelfs als je supersnel door je foto-albums heen klikt, herken je direct de inhoud ervan, zoals de locatie en welke mensen en objecten erop staan.

Is dat nu echt zo bijzonder, vraagt u zich nu misschien af: het zou toch heel onhandig zijn als onze perceptuele processen enorm langzaam zouden gaan? Tot op de dag van vandaag zijn we er echter nog niet geslaagd dit proces - de supersnelle waarneming van beelden - na te bootsen in een computer. Zelfs de meest geavanceerde programma's en algoritmes kunnen niet zo snel beelden analyseren als het menselijk brein. In de jaren negentig raakte men geïnteresseerd in de vraag hoe het brein dit nou precies klaarspeelt. Het onderzoek richtte zich nu niet alleen op hoeveel tijd nodig is om beelden te herkennen, maar ook op de verwerkingstijd van de hersenen, door hun activiteit te meten met een electro-encefalogram (EEG), een techniek die ook in dit proefschrift is toegepast. Hieruit bleek dat hersensignalen opgewekt door visuele beelden binnen enkele tienden van een seconde al anders gaan verlopen voor beelden met een andere inhoud: een teken dat het brein dan al een onderscheid maakt tussen de beelden. Vervolgens werd duidelijk dat het brein dit ook doet als er meerdere plaatjes tegelijk worden getoond, en zelfs ook nog als de proefpersonen tegelijkertijd een andere taak uitvoerden. De analyse van beelden verloopt dus snel en efficiënt: er is maar één blik op een plaatje, plus een korte stoot hersenactiviteit, nodig om de beelden correct te herkennen.

Wat het geheel nog mysterieuzer maakt, is dat het brein hier vooral heel goed in is als het beelden moet herkennen die afkomstig zijn uit de 'echte wereld', oftewel gewone, alledaagse foto's. Bij letters, gekleurde schijven, of andere abstracties zoals tekeningen wordt de verwerking juist langzamer! Dit strookt niet met onze intuïtie dat dergelijke plaatjes simpeler en daarom juist makkelijker te herkennen zijn. De verklaring die wetenschappers hiervoor bieden is dat de verwerking van informatie uit de echte wereld in het brein geoptimaliseerd is gedurende de evolutie. Deze speciale ontwikkeling zou bijvoorbeeld kunnen inhouden dat het brein optimaal gebruik maakt van bepaalde cues, aanwijzingen, die alleen in de echte visuele wereld te vinden zijn. Maar wat zijn die aanwijzingen dan precies, en hoe helpen ze ons brein om de inhoud van een beeld snel te analyseren?



**Figuur 1** Onderzoeksvraag en aanpak. **A)** Als een visuele stimulus wordt aangeboden, zet het brein de lichtstralen die vanaf het plaatje komen om in een neurale signaal. Vanaf het oog worden signalen verzonden (rode pijlen) via subcorticale structuren naar de visuele cortex. Subcorticale kernen reageren voornamelijk op contrast, dat wil zeggen overgangen van licht naar donker in een plaatje. Het eerste corticale gebied, in het achterste puntje van het brein, reageert op contrast met een bepaalde oriëntatie (bijvoorbeeld verticaal of horizontaal). Gebieden verderop in het brein zijn gevoelig voor hele objecten, en soms zelfs specifieke objecten zoals gezichten, huizen of lichamen. Het is echter nog onduidelijk hoe al deze informatie samenkomt in een coherente representatie van de hele visuele scène. **B)** Het in dit proefschrift beschreven onderzoek bevindt zich op het snijvlak van psychologie, neurowetenschap en *computer vision*, een tak van informatica. Door middel van psychologische gedragsexperimenten is onderzocht hoe mensen complete visuele scènes ervaren; met behulp van moderne *neuroimaging* technieken is door scènes opgeroepen neurale activiteit gemeten in het menselijk brein; en visuele informatie in scènes is gekwantificeerd met behulp van computationele modellen uit de informatica.

### *Is het brein gevoelig voor fysieke wetmatigheden?*

Uit eerder onderzoek, nu met een techniek genaamd *functional magnetic resonance imaging* (fMRI), is al gebleken dat er hersengebieden zijn die selectief actief worden als een alledaags beeld, oftewel een scène, getoond wordt in plaats van een abstract object. Deze gebieden zijn vooral gevoelig voor het feit dat scènes een ruimtelijke structuur hebben. Bovendien lijken ze ook onderscheid te maken tussen verschillende omgevingen, zoals een bos versus een strand: iets dat ook weer te maken lijkt te hebben met ruimtelijke verschillen tussen dit soort omgevingen, bijvoorbeeld dat een bos vooral uit verticale contouren bestaat terwijl stranden horizontaal uitstrekken. Vervolgens hebben heel precieze experimenten aangetoond dat het brein deze ruimtelijke informatie nóg sneller dan de categorie (bos of strand) of de inhoud (mannen die bier drinken) kan onderscheiden. Omdat ze voorafgaan aan de herkenning zelf, zouden dergelijke ruimtelijke eigenschappen van alledaagse foto's wellicht kunnen helpen om ze snel te verwerken. Maar hoe 'weet' het brein nou dat het naar een bepaald soort ruimtelijke omgeving kijkt?

Hier komen de fysieke wetmatigheden om de hoek kijken die zojuist al genoemd werden. Gemiddeld is het namelijk zo dat afbeeldingen van bossen fundamenteel andere statistische eigenschappen hebben dan die van stranden: in bossen zijn veel meer kleine, zwakke lijnelementen aanwezig, terwijl een strand bestaat uit grote vlakken met een grove lijn die de vlakken doorkruist (en een stadsscène juist weer uit meerdere vlakken, objecten en lijnen met sterke contouren). Dit is een wetmatigheid omdat een bos per definitie bestaat uit een verzameling van bomen met blaadjes en een strand of stad niet. Als het brein op een of andere manier op de hoogte is van deze statistische regeltjes, dan kan het die informatie gebruiken - als een soort trucje - om snel een inschatting te maken van het soort visuele input waar het mee te maken heeft.

Verrassend genoeg is er erg weinig onderzoek gedaan naar de vraag in hoeverre het brein inderdaad gevoelig is voor fysieke wetmatigheden. Een belangrijke reden hiervoor is dat de analyse van statistische eigenschappen van visuele beelden voornamelijk plaats heeft gevonden in een ander vakgebied dan de visuele neurowetenschappen, namelijk in *computer vision*, een tak van de informatica. Om deze vraag te kunnen beantwoorden is dus een interdisciplinaire aanpak nodig: een aanpak die een brug kan slaan tussen verschillende vakgebieden.

### *Uitgevoerde experimenten en belangrijkste bevindingen*

De centrale onderzoeksvraag van dit proefschrift is hier dan ook benaderd vanuit meerdere vakgebieden (zie Figuur 1.1B in de Inleiding). Psychologische experimenten zijn gecombineerd met EEG en fMRI om de hersenactiviteit die optreedt tijdens visuele perceptie te meten. Deze zijn vervolgens gecombineerd met uit de informatica afkomstige modellen om statistische informatie die de fysieke wetmatigheden beschrijft in scènes te kwantificeren. Om deze psychologische, neurologische en computationele metingen voor het eerst direct aan elkaar te relateren, zijn in dit proefschrift zowel nieuwe methodes ontwikkeld, als bestaande methodes toegepast op nieuwe vormen van data. Het overkoepelende doel was om te begrijpen of visuele fysieke wetmatigheden gereflecteerd worden in neurale activiteit en hoe ze bijdragen aan snelle waarneming van visuele scènes.

Het proefschrift kan grofweg worden opgesplitst in twee delen. Het eerste deel (*Hoofdstuk 2, 3 en 4*), heeft een sterk methodologische focus: deze hoofdstukken zijn met name gericht op de vraag hoe we de statistische informatie in beelden kunnen relateren aan

hersenenactiviteit en perceptuele ervaringen. Hierbij gebruikten we niet meteen echte scènes, maar begonnen we met iets eenvoudiger plaatjes waarvan we de eigenschappen beter konden controleren. Zo gaat *Hoofdstuk 2* over een EEG experiment met abstracte plaatjes (die echter dezelfde statistische eigenschappen hadden als echte scènes) om te kijken in hoeverre verschillen in die statistische eigenschappen terug te zien waren in de hersenenactiviteit en het gedrag van proefpersonen die plaatjes moesten groeperen. In *Hoofdstuk 3* gebruikten we een vergelijkbare experimentele opzet, maar nu lieten we de proefpersonen plaatjes van texturen zien – van dichtbij gefotografeerde materialen zoals zand of textiel – die ook weer verschilden in hun statistische eigenschappen.

Uit beide hoofdstukken bleek dat de hersenen sterke gevoeligheid vertoonden voor statistische eigenschappen: de patronen van activiteit kwamen direct overeen met de statistische verschillen tussen de plaatjes. Ook konden we op basis van die verschillen het categorisatiegedrag van de proefpersonen voorspellen: plaatjes met meer vergelijkbare statistiek werden als meer gelijkwaardig beschouwd. Een belangrijk resultaat uit deze experimenten was bovendien dat niet alle statistische eigenschappen de hersenen- en gedragsdata even goed konden voorspellen. Bepaalde statistieken die theoretisch gezien wel een onderscheid tussen plaatjes kunnen maken, werden in feite nauwelijks door het brein opgepikt. Hiermee wordt dus duidelijker welke berekeningen het brein daadwerkelijk wel en niet kan uitvoeren.

Op basis van deze veelbelovende resultaten verlegden we in *Hoofdstuk 4* de aandacht naar de hoofdvraag: de snelle verwerking van alledaagse scènes. We lieten proefpersonen 1600 (!) verschillende foto's zien, waarbij we steeds vroegen of ze per plaatje snel konden beslissen of het een natuurlijke of door mensen gemaakte omgeving betrof. Weer liet de hersendata zien dat het brein een sterke gevoeligheid vertoonde voor bepaalde statistische eigenschappen van de plaatjes. Het bleek dat deze statistiek vooral informatie bevatte over de mate van coherentie in de scene: of deze heel chaotisch en ongestructureerd was of juist bestond uit een sterk georganiseerde omgeving. Dit bleek ook voorspellend voor het uiteindelijke categorisatiegedrag van de proefpersonen: omdat door mensen gemaakte omgevingen meer gestructureerd zijn dan omgevingen waar de natuur zijn gang kan gaan, is dergelijke informatie nuttig voor het bepalen van het soort omgeving waar we ons in bevinden. Bovendien konden we een directe link leggen tussen deze eigenschappen, de sterkte van de hersenenactiviteit en de uiteindelijke keuze van de proefpersonen. Voor plaatjes waarvan de statistische structuur zeer duidelijk op een chaotische of juist gestructureerde omgeving wees, was een sterker hersensignaal aanwezig dan voor plaatjes met minder duidelijke statistische eigenschappen; we konden aantonen dat de plaatjes om die reden makkelijker te categoriseren waren.

In het tweede deel van dit proefschrift werd plaatsgemaakt voor een wat bredere focus, waarbij we de hersenenactiviteit niet alleen probeerden te begrijpen vanuit de fysieke wetmatigheden (het *bottom-up* perspectief) maar ook vanuit de taak context en doelstelling van de waarnemer (*top-down* perspectief). Met name *Hoofdstuk 5* draaide om de vraag in hoeverre de verwerking van de statistische eigenschappen van plaatjes automatisch was. Hoeveel van deze verwerking vindt plaats als het brein helemaal geen beslissing moet maken over de beelden, maar bezig is met een andere taak? Uit de resultaten bleek dat het eerste deel van de verwerking altijd aanwezig was, zelfs als dit niet nodig was om de categorisatie taak te volbrengen. Een tweede deel van de activiteit was echter wel taakafhankelijk. Dit

experiment liet zien dat de gevoeligheid voor statistische eigenschappen selectief verhoogd kan worden als het relevant is voor de huidige doelstelling.

Tenslotte bekeken we in *Hoofdstuk 6* in hoeverre de statistische structuur van scènes van invloed is op de efficiëntie waarmee het brein objecten in die scènes kan vinden. Dit is een taak die we regelmatig uitvoeren in ons dagelijks leven - denk bijvoorbeeld aan het terugvinden van een geparkeerde fiets voor de deur van de universiteit – die zoals eerder gezegd al vanaf de jaren 70 onderzocht wordt. Deze keer gebruikten we zowel fMRI als EEG om te laten zien dat de visuele processen die betrokken zijn bij het detecteren van objecten anders verlopen afhankelijk van de structuur van de scène. Voor drukke, chaotische scènes is meer van de zogenaamde *feedback* activiteit nodig, die de informatie in de visuele hersengebieden als het ware vasthoudt, terwijl dit niet het geval lijkt te zijn voor de meer gestructureerde scènes. De extra activiteit is waarschijnlijk nodig om het object 'uit te lichten' ten opzichte van de drukke achtergrond, terwijl voor een gestructureerde scene het object als het ware automatisch naar de voorgrond komt. Dit experiment laat dus zien dat de statistiek niet alleen de neurale representaties van ruimtelijke informatie vormgeeft, maar ook nog eens de verwerking van andere, meer gedetailleerde visuele informatie over objecten beïnvloedt.

### *Conclusie*

In dit proefschrift toon ik aan dat het menselijk brein een sterke gevoeligheid vertoont voor statistische eigenschappen van de alledaagse visuele wereld, die afgeleid zijn van fysieke wetmatigheden in die wereld. Door middel van een nieuwe interdisciplinaire aanpak heb ik onderzocht hoe die eigenschappen visuele verwerkingsprocessen beïnvloeden, dat wil zeggen neurale representaties vormgeven en perceptuele ervaringen voorspellen. Uit de resultaten blijkt dat de herkenning van texturen en scènes deels gestuurd wordt door hun statistische eigenschappen, en dat deze eigenschappen ook visuele processen betrokken bij het detecteren van objecten in scènes beïnvloeden. Samengenomen wijzen de resultaten erop dat de visuele cortex statistische informatie aan plaatjes onttrekt om de mate van coherentie in de wereld te kunnen bepalen, oftewel de aanwezigheid van georganiseerde versus chaotische structuur. Deze informatie draagt mogelijk bij aan snelle categorisatie en herkenning van de inhoud van een scènes, en kan, afhankelijk van de doelstelling van de waarnemer, leiden tot aanpassingen van andere visuele processen.