AFRL-AFOSR-VA-TR-2020-0108

Stochastic Resonance and Perceptual Decision Making Under Inattention

Hakwan Lau
UNIVERSITY OF CALIFORNIA LOS ANGELES

01/13/2020
Final Report

DISTRIBUTION A: Distribution approved for public release.

Air Force Research Laboratory
AF Office Of Scientific Research (AFOSR)/ RTA2
Arlington, Virginia 22203
Air Force Materiel Command

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*: 14-01-2020

2. REPORT TYPE: Final Report

3. DATES COVERED *(From - To)*: April 2015-October 2019

4. TITLE AND SUBTITLE
Stochastic Resonance and Perceptual Decision Making Inattention

5a. CONTRACT NUMBER

5b. GRANT NUMBER
FA9550-15-1-0110

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Hakwan Lau

5d. PROJECT NUMBER
FOS20140035

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of California, Los Angeles
Department of Psychology
2191 Franz Hall
Los Angeles, CA 90095

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Office of Scientific Research
875 North Randolph Street, Room 3112
Arlington, VA 22203-1954
United States

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Distribution A – Approved for Public Release

13. SUPPLEMENTARY NOTES

14. ABSTRACT
In this study, we empirically investigated to what extent the phenomenon of stochastic resonance under lack of attention generalizes to naturalistic stimuli and different contexts, and may work under different mechanisms (Specific Objectives 1 & 2). We also tried to understand the neurobiological mechanism behind this intriguing phenomenon, by means of computational modeling (Specific Objectives 3). We found that people tend to confirm what they expect whenever they don't attend and that participants' subjective sense of the visual surround is inflated. We also developed a simple leaky competing accumulator neural network model incorporating a known property of sensory neurons: tuned normalization. We demonstrate that this biologically plausible model can account for several counterintuitive findings reported in the literature, where confidence and decision accuracy were shown to dissociate -- and that the differential contribution a neuron makes to decisions versus confidence judgments based on its normalization tuning is vital to capturing some of these effects.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**	Unclassified	25	James Lawton
					19b. TELEPHONE NUMBER *(include area code)* 718-938-3780

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

Stochastic Resonance and Perceptual Decision Making Under Inattention

Principal Investigator: Hakwan Lau

Co-Investigators: Brian Odegaard, Megan Peters and JD Knotts

FA9550-15-1-0110

January 14, 2020

Contents

Project Overview

The project titled "Stochastic Resonance and Perceptual Decision Making Under Inattention," supported through Grant FA9550-15-1-011 from the Air Force Office of Scientific Research, is comprised of two different studies led by Principal Investigator (PI) Hakwan Lau, Professor of Psychology at the University of California, Los Angeles, concerning perceptual decision-making. The first study investigates the capacity of the visual system to detect objects in naturalistic scenes. The second study recorded eye gaze with an infrared camera to conduct investigations with subjects attending directly at fixation (overt attention) or attending to objects in the visual periphery (covert attention) in experiments using naturalistic stimuli. Finally, we modeled behavioral data with large-scale neuronal network models. This report will provide a comprehensive summary of the significant work accomplished for the project.

1 Background and Introduction

Physiological and behavioral studies suggest that visual information processing in the unattended periphery is typically poor, characterized by limited spatial resolution and low sensitivity for processing various features (Carrasco, 2011, Corbetta and Shulman 2002, Petersen and Posner 2012, Treisman and Kanwisher 1998, Lu and Dosher 2004, Dosher and Lu 2000). Yet in everyday life one gets a sense that subjective vision is relatively uniform across the visual field; we do not experience a striking lack of details in the unattended periphery (Lau and Rosenthal 2011, Cohen and Dennett 2011). We have recently developed a psychophysical paradigm to behaviorally characterize this puzzling phenomenon (Rahnev et al., 2011). Essentially, under lack of attention, we are over-confident as to what we see, and the brain uses a liberal criterion for detecting objects.

We have proposed a formal model (Rahnev et al., 2011) to explain this phenomenon, and to account for the psychophysics data (Rahnev et al., 2011). In this detection theoretic (Green and Swets, 1966) model, attention reduces the trial-by-trial variability of the internal perceptual response in the brain, in addition to boosting the gain of the response (i.e., the total amount of signal). An additional assumption, based on previous empirical findings (Gorea and Sagi, 2000), is that human observers use a single unified (or unique) criterion for detection of both the attended and unattended signal (Figure 1). Because the

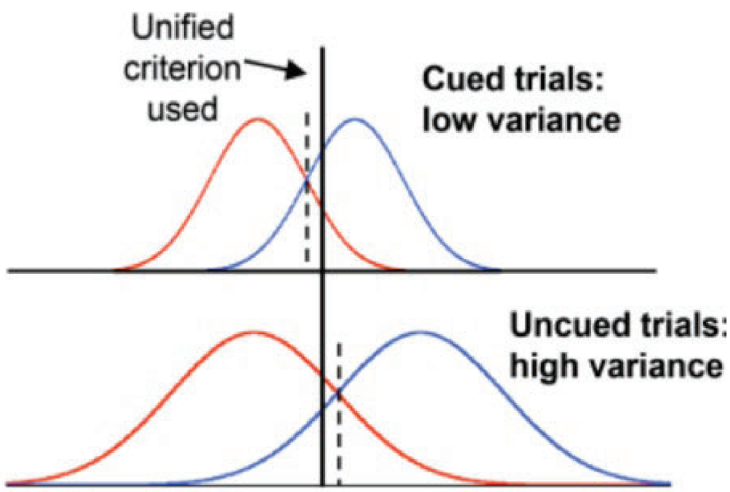


Figure 1. Attended (top) versus unattended (bottom) detection. Shown here are probability distributions with internal response strength on the horizontal axis, in a situation when sensitivity d’ is matched between the conditions, cued (attended) and uncued (unattended). The red and blue curves represent the situations when the target is not presented (i.e. baseline noise), and when the target is presented, respectively. The distributions are wider in the unattended case because it is assumed that the variability is higher. That the mean of the blue curve is further shifted to the right in the unattended is to reflect the fact that a higher contrast stimulus was needed to match sensitivity d’ between attended and unattended. Using the same detection criterion (solid black line) for both attended and unattended leads to a more liberal detection strategy in the periphery. I.e. there are more true hits as well as false alarms.

distribution of the internal perceptual response is more spread out for the unattended stimuli, the single unified criterion means that detection is compromised for both the attended and unattended targets, leading to conservative detection for the attended stimuli and liberal detection for the unattended stimuli. This occurs because the "optimal" decision criterion is different in the attended and unattended case; optimality as in defined as maximizing overall accuracy rate. In this experiment where the target and non-target are equally likely, the optimal criterion in each condition is at the midpoint between the two distributions, i.e. they should be different between the attended and the unattended conditions. However, human observers seem to use a single criterion that is applied to both conditions. This results in a suboptimality for both the attended (cued) and the unattended (uncued).

The same principles for the model described above are also applicable to discrimination tasks. As is customary in detection theoretical modeling (Macmillan and Creelman, 1987), discrimination can be treated as equivalent or similar to detection. In detection a criterion is applied to determine target from noise in detection, while in discrimination the criterion is applied to determine one target from another. In addition to asking subjects to discriminate between the two stimulus alternatives (a left tilted or right tilted grating pattern), Rahnev et al (2011) also asked subjects to provide confidence/visibility ratings associated with their responses. In detection theoretic modeling (Macmillan and Creelman, 1987), confidence/visibility ratings can be thought of as generated by placing additional criteria on the side of the main decision criterion (see Figure 2). By imposing the same set of criteria for both the attended and unattended conditions, the model accounted for the observed overconfidence in the unattended (high variance) condition.

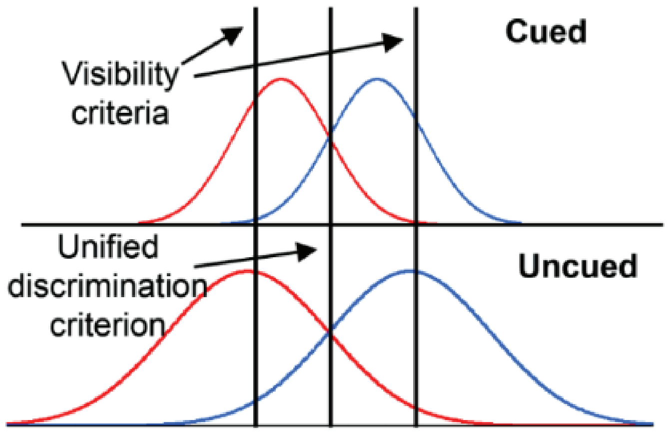


Figure 2. Discrimination and confidence/visibility ratings within the signal detection framework. As in detection, subject set a central criterion to discriminate between two distributions of internal responses, driven by two the stimulus alternatives respectively. To give a confidence or visibility rating response, one can place two other "flanking" criteria, such that responses getting into the extreme region will be considered higher confidence or high visibility. To have more levels of confidence or visibility ratings one can set more criteria. As in Figure 1, the unattended case has wider distributions because internal variability of the perceptual response is assumed to be higher. When the same set of decision criteria are applied to both the attended and the unattended, the attended to lead to higher confidence/visibility ratings: because the wider distributions means more trials will get into the high ratings zone.

This intriguing phenomenon potentially represents a simple and straightforward case of stochastic resonance (McDonnell and Ward 2011, McDonnell and Abbott 2009, Schwarzkopf et al 2011) By "stochastic resonance" we mean it in the broad sense, i.e. a situation in which the presence of noise/variability improves signal detection, as measured here by hit rate ("Target" trials passing the criterion). In some terminology this can be referred to as "stochastic facilitation" (McDonnell and Ward 2011), and does not imply frequency-specific modulation of signal-to-noise ratio. We speculate that, given the relatively poor signal-to-noise ratio for unattended stimulus perception and the fact that detection criteria may not be flexibly controlled for different stimuli presented simultaneously (Gorea and Sagi, 2000), the relatively high variability in unattended stimuli perception may in fact be helpful in promoting detection success. That is, the high variability inherent in unattended processing helps to push the signal across a somewhat rigid threshold. This may paradoxically be a good strategy for the brain to encode unattended stimuli with high noise/variability, given its limited signal-to-noise ratio.

Whereas stochastic resonance has been a controversial concept in biological systems (McDonnell and Abbott 2009, McDonnell and Ward 2011), here we have a behaviorally relevant case that is relatively tractable with current methodology. With this, our project has explored the important issue of how the human brain may capitalize on its intrinsically noisy processing to achieve statistical optimality, given the ubiquitous task of detecting objects in the environment and deriving decision confidence appropriately in perception.

Specific Objectives

The overall purpose of this project is two-fold: first is to empirically investigate to what extent the phenomenon of stochastic resonance under lack of attention generalizes to naturalistic stimuli and different contexts, and may work under different mechanisms (Specific Objectives 1 & 2). The second, which is the more ambitious goal, is to try to understand the neurobiological mechanism behind this intriguing phenomenon, by means of computational modeling (Specific Objectives 3)

This project has three specific objectives:

Specific Objective 1: Naturalistic Stimuli

In this study we employed naturalistic stimuli taken from short video clips of actual battlefield-like scenes. We use state-of-the-art computer graphics technology to manipulate visual stimuli to insert visual “targets” which subjects have to detect or discriminate. The experiments are intended to be similar in structure as those in Rahnev et al (2011), and the main manipulation is to make the setting more naturalistic and ecologically valid.

Specific Objective 2: Overt Attention (Central vs. Peripheral Vision)

In everyday setting, when we deploy our attention to particular location, two things happen. First we typically direct our eye gaze towards the target, and second we also mentally focus on the target. Psychologists have long distinguished these two forms of attentional selection. Under overt change in eye gaze (i.e. overt attention), essentially the difference between the attended objects and the unattended objects is that the former falls onto the fovea of the retina and the latter falls onto the periphery. In the previous experiments in Rahnev et al (2011) eye movements were carefully controlled, as we focused on the effects of covert attention (i.e. shifting of mental focus in the absence of eye movements) in isolation. Here we tested whether overt attention may have similar effects. That is, if we compare foveal versus peripheral vision, and see if under peripheral vision subjects may show liberal detection bias and over-confidence in discrimination, just as they did in Rahnev et al (2011) under the lack of covert attention.

Because this experiment required the careful control of eye movements we ran this in the laboratory where eye gaze is recorded with an infrared camera. However, as in the experiments in Specific Objectives 1 we also used naturalistic stimuli (i.e. battlefield like scenes).

Specific Objective 3: Computational Modeling

We modeled behavioral data with large-scale neuronal network models. We capitalized on the existing knowledge about the biophysical properties of single neurons (Dayan and Abbott, 2001), and used computer simulations to evaluate the behavior of a large-scale network as informed by realistic neuronal connection profiles. Similar large-scale networks for perceptual decisions have been proposed (Dayan and Abbot 2001, Ma et al 2006), and they provided a theoretical platform for the present work. Our strategy was to construct multiple versions of network models with different architecture, and use information-theoretic model comparison methods (Dayan and Abbot 2001) to decide which model can most parsimoniously account for the observed behavioral data.

2 Summary of Major Accomplishments

2.1 Year One Accomplishments

During Year One (April 15, 2015 – April 14, 2016), our research team began work on preparing experiment equipment and developing an experiment interface. The research team also conducted several pilot studies, administered a questionnaire and designed a biologically plausible recurrent leaky competing accumulator network model. Postdocs Brian Odegaard and Megan Peters presented preliminary findings at conferences.

Specific Objective 1: Naturalistic Stimuli

Research Development

- Established online platform.
- Administered a questionnaire to address whether there may be individual differences in the degree of phenomenological inflation in the periphery.
- Recruited 93 subjects.

Experimental Accomplishments

- Conducted several pilot studies with artificial stimuli to determine subject reliability using these online methods. Obtained preliminary results.
- Investigated the differences between central and peripheral vision in summary statistical processing by implementing a task in which observers had to judge whether the average orientation of a group of lines was to the right or left of vertical.

Research Dissemination

- Postdoc Brian Odegaard attended the Computational and Systems Neuroscience (COSYNE) Workshop in Snowbird Utah (February 29 – March 1, 2016). Odegaard organized a workshop entitled "Recent Innovations in Attention Research: Linking Models, Mechanisms, and Behavior," which discussed the above and other research findings, ideas, and motifs from both behavioral and computational investigations of attention as well as relevant neurophysiological findings.

Specific Objective 2: Overt Attention (Central vs. Peripheral Vision)

Research Development

- Purchased an eye tracker to use for this series of experiments: the REDn Eye Tracking System from SMI (Binocular 60Hz REDn Camera Unit).
- Worked on assembling the eye tracker device and addressing technical considerations to allow for online data collection of both behavioral and physiological variables in our experiments.
- Worked with the software developer kit for the SMI unit to create different experiments that interface with MATLAB and can potentially be combined with a new EEG system to be purchased in May 2016.

Specific Objective 3: Computational Modeling

Research Development

- Designed a biologically plausible recurrent leaky competing accumulator network model to perform a perceptual decision-making task and rate confidence based on the stimulus' detectability.
- Validated the model with multi-unit recordings from Rhesus macaques.
- Reevaluated a previously published dataset to assess performance differences between a 'detectability' heuristic model and the optimal Bayesian ideal observer (which ignores stimulus detectability).

- Presented and validated the Confidence as Detectability (CaD) model: a normative Bayesian framework accompanied by a recurrent neural network implementation that utilizes tuned normalization using a series of simulations.

Research Dissemination

- Postdoc Megan Peters presented findings through "Separable calculations underlie decisions and confidence judgments: Tuned normalization, detectability, and confidence in perceptual decision-making" at the COSYNE Workshop in March 2016 and "A Neuronal Network Model of Perceptual Confidence Supports the Empirical Link Between Consciousness and Metacognition" at the ASSC Annual Meeting in June 2016.

2.2 Year Two Accomplishments

During Year Two (April 15, 2016 – April 14, 2017), our research team developed and extended an online experimental paradigm, integrated the SMI REDn Eye Tracking System, collected data and launched several online experiments. Our research team also extended the computational modeling work. We presented our work at a symposium and published a commentary.

Specific Objective 1: Naturalistic Stimuli

Research Development

- Created an experimental paradigm using an internet testing platform to investigate the capacities of the visual system to both *detect* and *discriminate* objects in naturalistic scenes.
- Integrated the SMI REDn Eye Tracking System (purchased during Year One of this grant) with this task, which enabled us to not only track subjects' eye position at every moment while they are driving in the simulator, but also to control stimulus presentation based on where the eyes fixate at a particular instance.

Experimental Accomplishments

- Collected data in several different experiments that evaluate how both eye position and attention influence detection and discrimination capacities in the periphery.
- Launched several online experiments for larger numbers of subjects (N > 30) to investigate the visual system's capacity for naturalistic object detection and identification, in order to better understand the degree of perceptual detail that is represented in the visual periphery.
- Conducted several online tasks that evaluate how effectively visual observers can *detect* changes in the periphery, as well as *discriminate* objects that have changed.

Research Dissemination

- Research team presented results by leading a symposium at the Vision Sciences Society in St. Petersburg, Florida, in May 2017.
- Research team published a brief commentary on the importance of a study which implemented a binoculary rivalry paradigm of subjective awareness (Giles et al., 2016).

Specific Objective 2: Overt Attention (Central vs. Peripheral Vision)

Research Development

- Developed an extension of the online paradigm described in Specific Objective 1, where we include a cue (i.e., a line) that denotes the position that needs to be attended to *before* the first presentation of images are shown, as well as a second line (presented after the images) which denotes the actual location that will be probed.

Experimental Accomplishments

- Experimentally manipulated observers' focus of attention to examine how attention influences detection and discrimination capacities in these tasks.

Specific Objective 3: Computational Modeling

Research Development

- Extended our work using the biologically plausible recurrent leaky competing accumulator network model to perform a perceptual decision-making task and rate confidence based on the stimulus' detectability.
- Demonstrated that the biologically plausible mechanism can reproduce even counterintuitive behaviors reported in the literature.
- Used electrophysiological recordings to show that that tuned normalization exists in monkey superior colliculus.

2.3 Year Three Accomplishments

During Year Three (April 15, 2017 – April 14, 2018), our research team conducted additional experiments and used results from the prior year to refine some of these experiments. We also submitted a paper for review and began drafting a manuscript for eventual journal submission.

Specific Objective 1: Naturalistic Stimuli

Experimental Accomplishments

- Conducted two experiments to investigate how subjects detected colorful stimuli in the unattended periphery in a naturalistic environment.

Specific Objective 2: Overt Attention (Central vs. Peripheral Vision)

Research Development

- Used findings from Specific Objective 1 to effectively isolate covert attention, to study how this process functions when covert attention is correctly allocated to different locations (valid condition) and incorrectly allocated to different locations (invalid condition).

Experimental Accomplishments

- Conducted two additional experiments to investigate whether metacognitive impairments and decisional biases would emerge in tasks which exploit two well-established phenomena in the visual surround: crowding and summary statistics.

Research Dissemination

- The research team submitted a paper on these experiments to a journal.
- The research team started preparing an additional paper in reference to the work done on this Specific Objective.

Specific Objective 3: Computational Modeling

Research Development

- Refined the model, reduced the number of free parameters to constrain data fitting and optimized the code in preparation for sharing it upon publication.

2.4 Year Four Accomplishments

During Year Four (April 15, 2018 – April 14, 2019), no new experiments were conducted. The research team published two papers, a review and revised and resubmitted another paper. We also presented our findings at a conference.

Specific Objective 1: Naturalistic Stimuli

Research Dissemination

- Published two papers in peer-reviewed journals.
- Presented at the Vision Sciences Society Annual Meeting in May 18-23, 2018.

Specific Objective 2: Overt Attention (Central vs. Peripheral Vision)

Research Development

- Reviewed the "Rich versus Sparse" debate about how to explain the seemingly rich nature of visual phenomenology while accounting for impoverished perception in the periphery.

Research Dissemination

- Published the review in *Current Opinion in Psychology*.

Specific Objective 3: Computational Modeling

Research Development

- Refined the model, reduced the number of free parameters to constrain data fitting and optimized the code in preparation for sharing it upon publication.

Research Dissemination

- Revised a paper on the computational model to reflect more recent findings and resubmitted the paper.

3 Experimental Facilities & Study Procedures

3.1 Specific Objective 1: Naturalistic Stimuli

We used a modern game engine to create a simulated driving environment in which participants (as drivers) had to make judgments about the colors of pedestrians' clothing in the periphery. In our first experiment, on each trial, we asked observers whether an individual wearing a shirt with a specific color (yellow) had been presented at a specific location. In our second experiment, on each trial, we asked observers whether an individual wearing a shirt with a particular color had been presented at a specific location, but in this experiment, we varied the target color randomly from trial to trial. We integrated the SMI REDn Eye Tracking System (purchased during Year One of this grant) with this task, which enables us to not only track subjects' eye position at every moment while they are driving in the simulator, but also to control stimulus presentation based on where the eyes fixate at a particular instance.

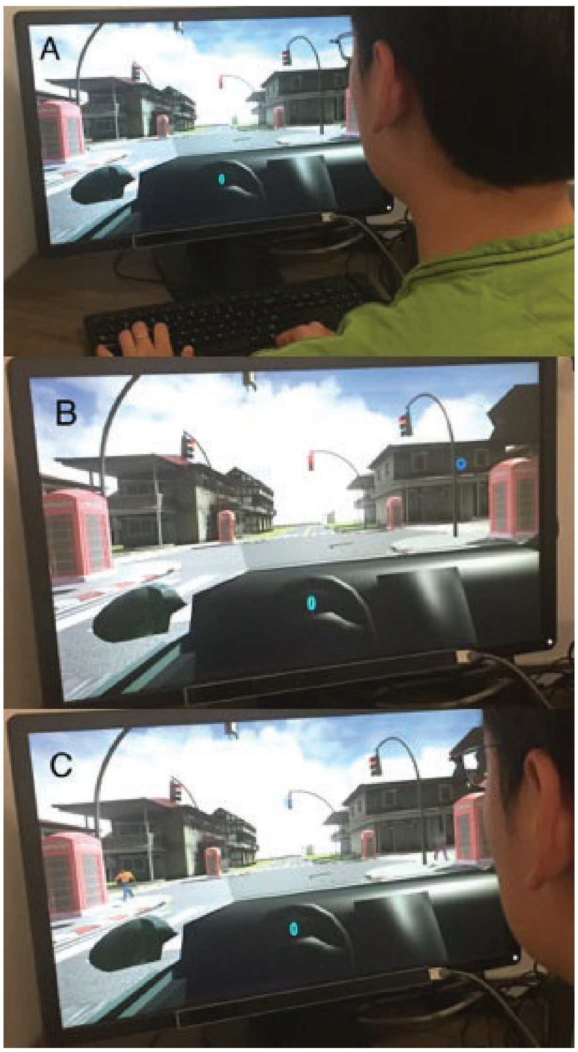


Figure 3. *(a) Basic settings of experiment. (b) Tracking participant's eye movement. Blue circle on screen is gaze point. Stimuli will not present when participant is not looking at traffic light. (c) Stimuli present when participant is looking at traffic light.*

The basic paradigm is shown below (Figure 4a, from Sligte et al., 2010), and involves an initial presentation of images in the periphery, a delay period, and a second presentation of images. Observers have to respond whether or not a change occurred for any of the items, and on some trials, select which item was originally presented from an "identification display" (Figure 4a). Critically, on some trials, a "retro-cue" occurs to label certain positions that subjects are required to remember. By evaluating the timing of the cue, we can track the fidelity of peripheral object representations in subjects' working memory. By integrating this existing task with psychophysical measures that quantify overall performance (i.e., *d'* (Green & Swets, 1966)) as well as how effectively confidence judgments distinguish between correct and incorrect answers (i.e., *meta-d'* (Maniscalco & Lau, 2012)), we can quantify both perceptual sensitivity and metacognitive sensitivity for peripheral object representations in this task.

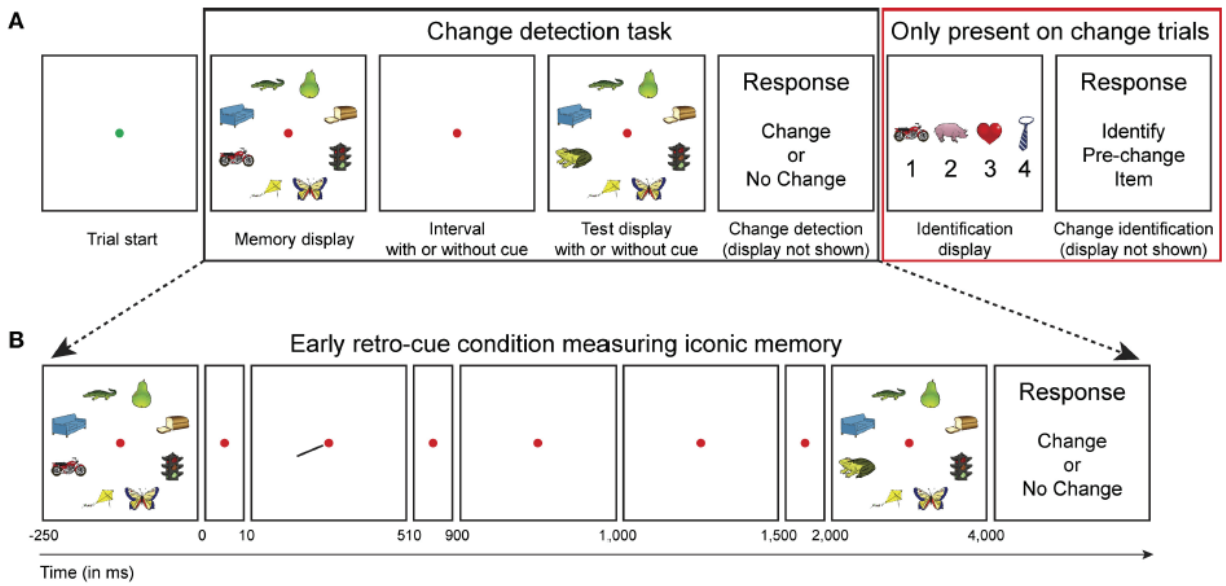


Figure 4. *Our basic online object detection/identification paradigm, based on Sligte et al., 2010.*

3.2 Specific Objective 2: Overt Attention (Central vs. Peripheral Vision)

In different versions of the experiments described in Specific Objective 1, we experimentally manipulate observers' focus of attention to examine how attention influences detection and discrimination capacities in these tasks. Overall, measuring how these mechanisms operate in our naturalistic driving setting will allow us to determine how attention may differentially impact the capacities for object detection and discrimination in the visual periphery. Similarly, we have also developed an extension of the online paradigm described in Specific Objective 1, which allows us to assess the effects of how *valid* (i.e., correct) attentional allocation (when the first and second lines point to the same location) facilitates object detection and discrimination capacities, as well as how *invalid* (i.e., incorrect, or when the first and second lines point to different locations) attentional allocation impairs performance.

Using eyetracking, we were able to ensure that subjects were fixating on the center of the screen in our driving task, while individuals were presented in the periphery. This ensured that we could effectively isolate covert attention, to study how this process functions when covert attention is correctly allocated to different locations (valid condition) and incorrectly allocated to different locations (invalid condition). It remains possible that even stronger differences in criteria may be found when naturalistic stimuli are presented in the center (overt attention) vs. the periphery (covert attention), as has been done in previous studies using artificial stimuli (Solovey et al., 2015).

To better understand what it means when people say they see the target more often, and to better understand how this phenomenon relates to central (i.e., overt attention) and peripheral (i.e., covert attention) presentation, we conducted two additional experiments during Year 3 to investigate whether metacognitive impairments and decisional biases would emerge in tasks which exploit two well-established phenomena in the visual surround: crowding and summary statistics.

3.3 *Specific Objective 3: Computational Modeling*

We used a standard desktop computer for this Specific Objective. We proposed that confidence is central to conscious phenomenology: it turns out that empirically, in humans as well as in monkeys, confidence in perceptual decisions critically reflects the subjective saliency and detectability of the stimulus. This finding explains and supports the theoretical validity behind the long tradition in psychophysics of using confidence ratings as a proxy for measuring subjective awareness in perception. It also calls into question the Bayesian Brain Hypothesis, according to which confidence is an optimal readout of perceptual reliability (e.g., Pouget et al., *Nat Neuro*, 2016).

Motivated by previous empirical results, we designed a biologically plausible recurrent leaky competing accumulator network model to perform a perceptual decisionmaking task and rate confidence based on the stimulus' detectability. The basic idea is that decisions themselves should weigh all evidence favoring possible stimulus alternatives (e.g., "Is this a dog, cat, or monkey?"), but the primary factor in confidence ratings is the detectability of the stimulus choice you made (i.e., "I chose cat. How much 'cat evidence' is there to suggest I am not hallucinating?"). This model reproduces a number of studies' findings in psychophysics (e.g., Rahnev et al., *Nat Neuro*, 2011; Maniscalco, Peters, & Lau, *Att Percep Psychophys*, 2016) and neuroimaging (Cortese et al., submitted). We also validated the model with multiunit recordings from Rhesus macaques. Finally, we applied this logic in reevaluating a previously published dataset (Peters & Lau, eLife, 2015), and show that a 'detectability' heuristic model performs better than the optimal Bayesian ideal observer (which ignores stimulus detectability). We proposed a single, normative framework to account for these findings, and additionally provided a recurrent neural network implementation.

Finally, we applied this logic in reevaluating a previously published dataset (Peters & Lau, eLife, 2015), and show that a 'detectability' heuristic model performs better than the optimal Bayesian ideal observer (which ignores stimulus detectability). We proposed a single, normative framework to account for these findings, and additionally provided a recurrent neural network implementation.

It was recently proposed that the absolute magnitude of responsecongruent evidence (i.e., stimulus energy/detectability favoring the chosen stimulus alternative) is disproportionately weighted in confidence judgments(Zylberberg et al., Front Int Neuro, 2012). We also recently showed that confidence judgments can be well described both by an ideal observer and a Bayesian observer that takes into account stimulus detectability (Peters & Lau, eLife, 2015). Following this work, we present the Confidence as Detectability (CaD) model: a normative Bayesian framework accompanied by a recurrent neural network implementation that utilizes tuned normalization (Lee & Maunsell, PLoS ONE, 2009; Ni et al., Neuron, 2012) to represent the degree to which a neuron codes for balance of evidence versus stimulus energy magnitude, which contribute differentially to perceptual decisions and confidence judgments. In a series of simulations, we showed that this model can explain all of the abovementioned findings with a single set of parameters. Importantly, using the "detectability" of a stimulus in judging confidence may not be suboptimal in the real environment, when the task is to judge not only which of innumerable stimulus alternatives is most likely to be present, but also whether a stimulus is present at all.

We hypothesized that the detectability of a stimulus constitutes an important component of confidence judgments, in addition to balance of evidence favoring the stimulus alternatives. We first implemented this decision strategy with Bayesian hierarchical inference: rather than marginalize over stimulus energy, as is the traditional approach, our Bayesian observer first makes a guess s^ at the most i,j probable stimulus intensity i to have generated the current data d for each stimulus alternative j, via s^ i,j = argmaxi p(d|Si,j) with p(d|Si,j)~N(si,j ,Σ) . The observer then makes its discrimination decision among the stimulus alternatives via

$$\hat{S}_{chosen} = argmax_j\, p(\hat{S}_{i,j}|d),\ \text{with}\ p(\hat{S}_{i,j}|d) = \frac{p(d|\hat{S}_{i,j})p(\hat{S}_{i,j})}{p(d)}\ \text{and}\ p(d|\hat{S}_{i,j}) \sim N(\hat{s}_{i,j}, \Sigma) \qquad (1)$$

Confidence is defined as the probability that the selected stimulus alternative is correct, $p(\hat{S}_{chosen}|d)$. This observer captures human behavior well(Peters & Lau, eLife, 2015), showing that detectability is important to confidence judgments.

How might this be implemented at the neural level? We proposed that the absolute stimulus energy and the balance of evidence favoring the various stimulus alternatives are respectively represented by neurons that possess lesser or greater degrees of divisive normalization (i.e., tuned normalization [Lee & Maunsell, PLoS ONE, 2009; Ni et al., Neuron, 2012]). Thus, extending previous work (Usher & McClelland, Psych Rev, 2001; Wang, Curr Op Neurobio, 2012) in this model the evolution over time of a neuron's firing rate in response to a stimulus is

$$dx_i = [\rho_i - (\lambda_i - \alpha_i)x_i - \beta_i \sum_{i' \neq i} \gamma_{i'} x_{i'}] + \xi_i \, , \; x_i \rightarrow max(x_i, 0) \tag{2}$$

with ρi representing the feedforward input to accumulator unit i, λi the leakage or decay rate of the unit's excitation, αi the scaling factor for its recurrent selfexcitation, and ξi an additive noise term. γi and βi are the factors of interest, representing the amount of tuned normalization (i.e., lateral inhibition) unit i receives and contributes, respectively: when βi is 1, the unit is strongly inhibited by neurons tuned to opposing stimuli; when βi is 0, the unit is not inhibited by opposingly tuned units; γi represents a scaling factor for the strength of inhibitory influence from unit i to other units. Note that Eq. 2 represents a simplified form: only two units possessing opposing stimulus preferences are present, and time scales and steps are set to 1 for convenience.

We define the perceptual decision according to which units are firing the most,

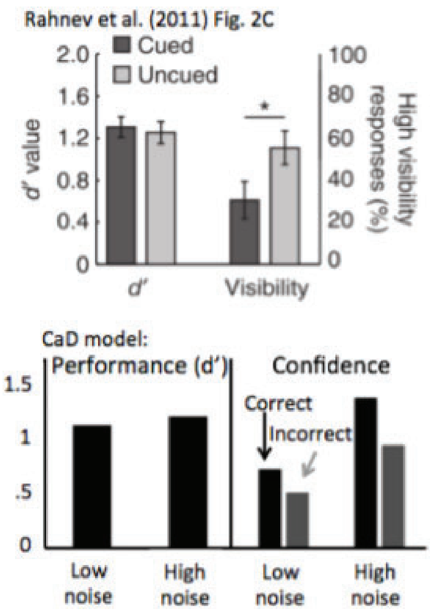
$$D = argmax_i f(x_i) \, , \; f(x_i) = V x_{i,\beta_i > \underline{\beta}} + (1 - V) x_{i,\beta_i < \underline{\beta}} \tag{3}$$

and define confidence judgments similarly, as

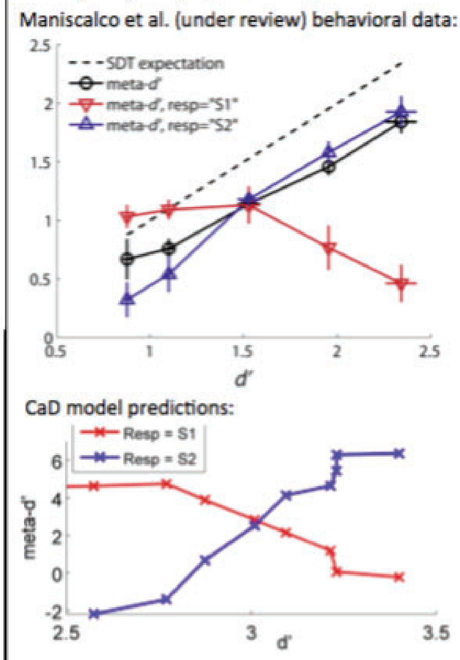
$$C = g(x_i) = W x_{i,\beta_i > \underline{\beta}} + (1 - W) x_{i,\beta) i < \underline{\beta}} \tag{4}$$

with *V* and *W* representing the relative contributions from more or lessnormalized units, respectively. We initialize the network with parameters drawn from the literature (Usher & McClelland, Psych Rev, 2001) and reasonable perturbations thereof(Lee & Maunsell, PLoS ONE, 2009; Ni et al., Neuron, 2012), and test its predictions for decision and confidence judgment behavior with varying stimulus inputs. The recurrent network implementation of the CaD model explains the following findings:

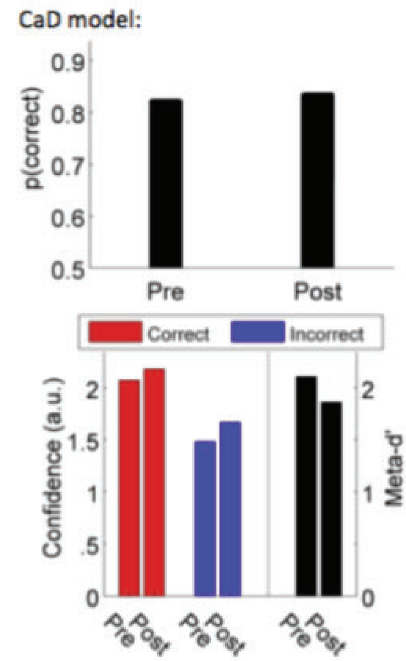
(a) More “seen” trials when stimuli are stronger but unattended (and therefore represented with less precision), despite matched discrimination performance (Rahnev et al., Nat Neuro, 2011).



(b) Different metacognitive sensitivities for each response alternative when evidence for each is systematically unbalanced (Maniscalco, Peters, & Lau, *Att Percep Psychophys*, 2016).



(c) Decreased metacognitive sensitivity (meta-d’) when decoded neurofeedback (DecNef [Shibata et al., *Science*, 2011]) changes confidence ratings without changing accuracy (Cortese et al., submitted).



4 Results & Discussion

4.1 Specific Objective 1: Naturalistic Stimuli

We proposed using an internet testing platform to investigate the capacities of the visual system to both *detect* and *discriminate* objects in naturalistic scenes. To initially quantify subject reliability for judging visual stimuli across different parts of the visual field, we investigated the differences between central and peripheral vision in summary statistical processing. We hypothesized that the visual periphery may possess a greater capacity for accurate summary computations than central vision in some instances, and that this superior ability relates to why we think we see more than we actually do outside the fovea. We implemented a task in which observers had to judge whether the average orientation of a group of lines was to the right or left of vertical.

Our results showed that with coherent stimuli, sensitivity for average orientation discrimination was better in the periphery. However, as the task difficulty increased, peripheral superiority quickly diminished relative to performance at the center. Subjects also exhibited higher levels of metacognitive efficiency (indexed by *meta-d'/d'*) for central compared to peripheral vision, indicating distorted awareness of this ability in the periphery. Finally, to address whether there may be individual differences in the degree of phenomenological inflation in the periphery, we administered a questionnaire with our relatively large sample (N=93). We found little support in the psychophysical data for evidence of direct correlations with subject reports of peripheral phenomenology. These results indicate that while *overestimation* of perceptual abilities influences peripheral phenomenology in specific cases, *underestimation* of some unexpected capacities exists, too, and these may be general phenomena existing in most subjects.

In our first experiment, on each trial, we asked observers whether an individual wearing a shirt with a specific color (yellow) had been presented at a specific location. Results showed that observers exhibited higher numbers of false alarms (i.e., saying a yellow shirt was present, even when it was not) in unattended locations in the visual periphery, compared to locations that were fully or partially attended. This provided evidence that participants adopted more liberal criteria for making detection judgments when the target was unattended and presented in periphery. This tendency to use liberal detection criteria in unattended or peripheral locations has been shown previously in artificial settings (Rahnev et al. 2011; Solovey, Graney, and Lau 2015). Here, we confirmed the hypothesis that this detection bias extends to more naturalistic stimuli and tasks.

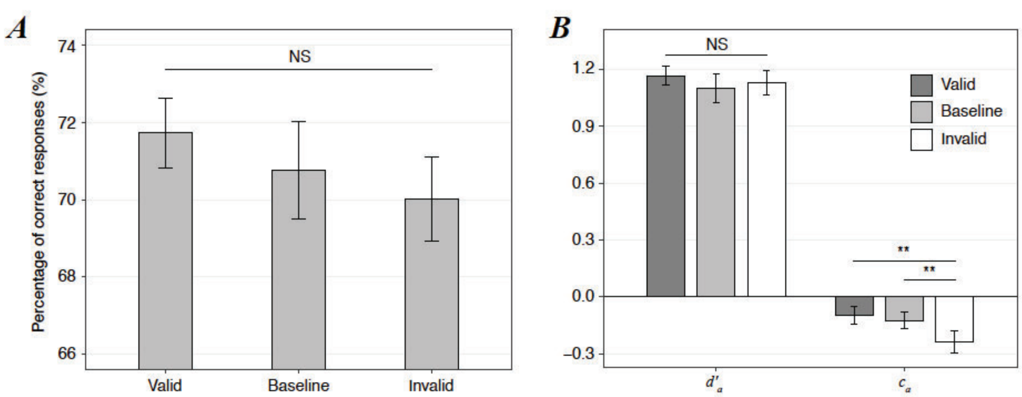


Figure 5. Results from investigating detection capacities in a simulated driving task. (A) The percentage of correct responses across attention conditions. While subjects exhibited the best performance for valid attention trials and the lowest performance for invalid trials, the conditions were not significantly different from one another. Bars represent averages, error bars represent SEM. (B) d_a and c_a across attention conditions. d_a values were quite consistent across attention conditions. c_a under inattention (i.e., invalid trials) was significantly lower than c_a in the valid or baseline conditions, providing evidence for a liberal detection bias. Bars represent the average values across subjects; error bars represent the standard error of the mean. **p < .01, NS: not significant.

In our second experiment, on each trial, we asked observers whether an individual wearing a shirt with a particular color had been presented at a specific location, but in this experiment, we varied the target color randomly from trial to trial. Results showed that subjects used relatively conservative perceptual criteria (i.e., were relatively reluctant to say a color was present) when making detection-related judgments in this experiment, regardless of the amount of attention that was allocated to a given location. This finding indicates that our results in the first experiment were not due to a sheer confirmation bias at the decisional level. The signal detection theory model which predicts liberal detection criterion in the periphery applies to situations when there is an *a priori* well defined stimulus dimension on which subject can place the criterion to do the detection. If the feature to be detected can only be known after the stimulus is gone, there is no method one could use to put the criterion in the same place for both the cued and uncued *over many trials* (which is the assumption in the model by Rahnev et al., 2011, and Solovey et al., 2015). As such, we did not predict the inflation effect in experiment 2, and this is exactly what we obtained.

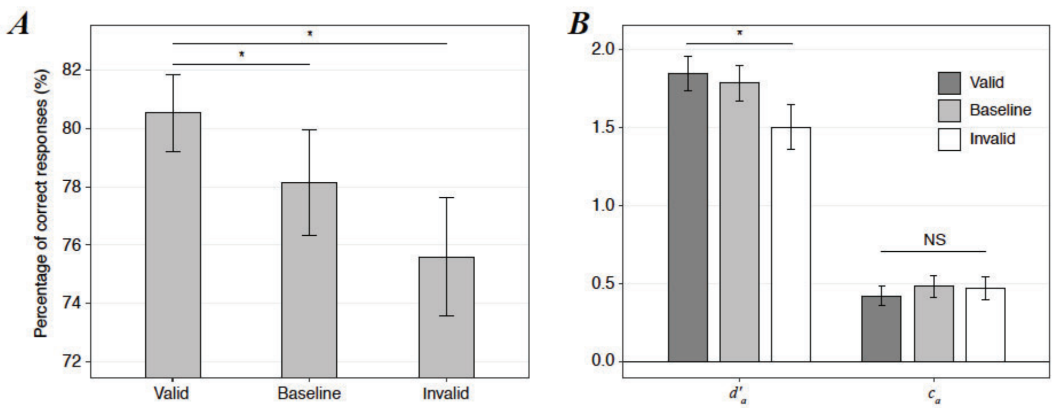


Figure 6. Behavioral results for Experiment 2. (A) The percentage of correct responses across attention conditions. The percentage of correct responses for valid trials was significantly larger than the percentage correct for baseline and invalid trials. (B) d_a and c_a across attention conditions. Similar to the effect of attention on percentage of correct responses, d_a was significantly greater in the valid attention condition compared to the invalid condition. c_a did not significantly differ across attention conditions, which indicated that participants used similar internal criteria to make perceptual judgments in all attention conditions. *$p < .05$, NS: not significant.

Together, these results support the idea that in everyday visual experience, there is subjective inflation of experienced detail in the periphery, which may happen at the decisional level. A paper on this study has been reviewed and is being revised for resubmission.

In terms of practical implications, if people tend to confirm what they expect whenever they do not attend, this raises an important issue regarding driving safety. Many people are optimistic and tend to expect things to be positive (Sharot 2011; Sharot et al. 2012, 2007), so they may tend to mistakenly detect hazards that present in unattended periphery as no danger, while what is actually happening is they just don't see the hazards. Similar results have been found in previous studies where drivers showed conservative criteria in hazard detection regardless of driving experience (Ventsislavova et al. 2016; Wallis and Horswill 2007). Hazard detection is a special detection task, because the penalty for a miss and a false alarm is different: a miss may cause a crash, while a false alarm may only lead to unnecessary brakes; thus, the criterion in hazard detection task is closely related with driving safety. In our research, the task is to detect a pedestrian when the vehicle has stopped at a stoplight. However, in real driving, most of the hazards are presented while driving, which is quite different from our task. Future research should more systematically address how inattention affects hazard detection judgment while driving. Specifically, while the impact of distractions (e.g., due to phone calls and text messaging) on driving performance is well known (Haigney, Taylor, and Westerman 2000; Rumschlag et al. 2015), the impact on specific aspects of the perceptual decision making process remains relatively unexplored.

4.2 Specific Objective 2: Overt Attention (Central vs. Peripheral Vision)

We proposed laboratory testing to record eye gaze with an infrared camera, in order to conduct investigations with subjects attending directly at fixation (overt attention), or attending to objects in the visual periphery (covert attention), in experiments using naturalistic stimuli. One important question is what it means when people say they see the target more often. Traditionally, it is thought that much of peripheral vision is 'filled in' via top-down mechanisms (Komatsu 2006). However, it has also been reported that people sometimes trust unreliable, filled-in percepts more than precepts based on external input (Ehinger et al. 2017), suggesting that filling-in may not be the complete mechanism to explain peripheral phenomenology, and that decisional or metacognitive mechanisms are also involved. In our study, perhaps the findings of detection bias in the unattended periphery can also be interpreted as congruent with this account involving mechanisms at the decisional or metacognitive level. Importantly, just because the effect is to be thought of at the 'decisional' level does not mean that this is unrelated to perceptual phenomenology; criterion effects can also reflect subjective percept (Phillips 2016; Witt et al. 2015). This interpretation is in line with previous findings that people tend to overestimate their ability to detect changes in change blindness experiments (Levin et al. 2000), and in a sense, people are not subjectively aware of their poor acuity and color perception in the periphery (Cohen, Dennett, and Kanwisher 2016).

To better understand what it means when people say they see the target more often, and to better understand how this phenomenon relates to central (i.e., overt attention) and peripheral (i.e., covert attention) presentation, we conducted two additional experiments to investigate whether metacognitive impairments and decisional biases would emerge in tasks which exploit two well-established phenomena in the visual surround: crowding and summary statistics. In our first experiment using crowded stimuli, observers showed relative deficits in metacognitive measures (i.e., confidence judgments) (Maniscalco and Lau 2014; Fleming and Lau 2014) for crowded compared to single stimuli. This metacognitive deficit was primarily driven by overconfidence in incorrect responses, which is striking given that subjects did not perform the primary discrimination task very well under crowding; the overconfidence is highly unwarranted.

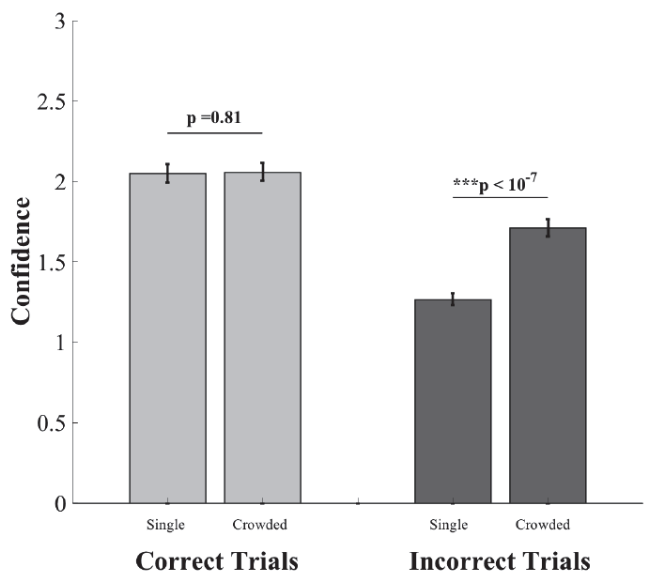


Figure 7. Crowding & metacognition experiment: average confidence for correct and incorrect trials. The light gray bars indicate the average confidence for correct trials for the single and crowded conditions. As can be seen in the figure, the difference between these conditions is not significant. The dark gray bars indicate the average confidence for incorrect trials. A clear difference between the single and crowded conditions is evident, and participants are significantly more confident in the incorrect crowded trials compared to the incorrect single trials.

In our second experiment using a summary statistical stimulus (groups of oriented lines), observers exhibited liberal detection criteria and high numbers of false alarms, showing that decisional biases extend to more complex stimuli than has been previously shown. Both of these findings provide experimental evidence that, far from perceiving the visual periphery with a high degree of fidelity (Block 2011; Haun et al. 2017; Vandenbroucke et al. 2014), our subjective sense of the visual surround is inflated, corroborating our finding in the study described above.

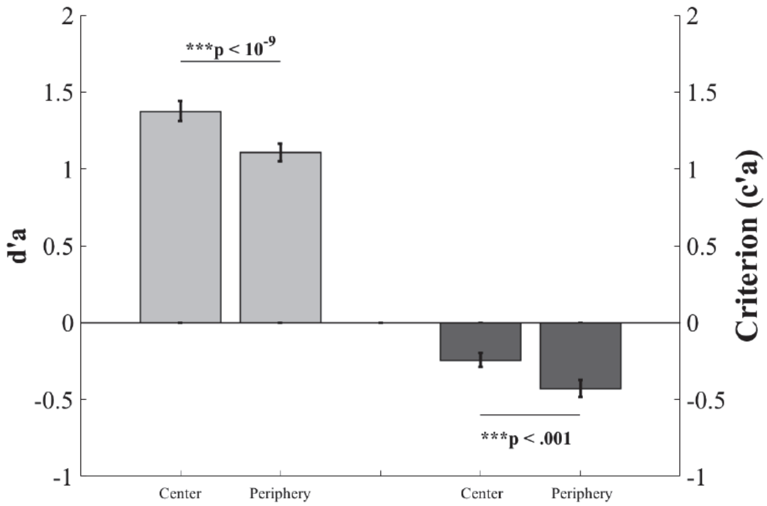


Figure 8. Sensitivity and bias for detecting congruently-oriented groups of lines. Shown here are results from an experiment where participants were asked to detect whether a group of lines with congruent orientations were presented, in either a central or peripheral location. As shown by the light gray panels, using the measure d_a (which corrects potential unequal variance in detection tasks), participants were more sensitive in detecting the congruent patch of lines at the central location compared to a peripheral location, and yet they used a more liberal criterion c_a in the periphery for indicating that a patch of lines was present. Notice that although sensitivity was not perfectly matched between center and periphery, usually we expect subjects to be relatively conservative for weaker detection, based on the Neyman-Pearson objective. Therefore, the results are striking in that it went opposite to that expectation.

4.3 Specific Objective 3: Computational Modeling

We proposed to model behavioral data with large-scale neuronal network models. During Year One, motivated by previous empirical results, we designed a biologically plausible, recurrent, leaky competing-accumulator network model to perform a perceptual decision-making task and rate confidence based on the stimulus' detectability. During Years Two, Three, and Four, we extended this work by refining the model, reducing the number of free parameters to constrain data fitting, and optimizing the code in preparation for sharing it upon publication.

This neural network model incorporates a known property of sensory neurons: tuned normalization. The key idea of the model is that each accumulator neuron's normalization 'tuning' dictates its contribution to perceptual decisions versus confidence judgments. Specifically, we reasoned that highly normalized evidence accumulation neurons encode the balance of evidence for various perceptual interpretations (e.g., net evidence for leftwards or rightwards motion direction), and thus are ideally suited for making discrimination judgments. By contrast, less normalized evidence accumulation neurons encode evidence in favor of one perceptual interpretation (e.g., leftward motion) while ignoring evidence for alternative interpretations (e.g., rightward motion), and thus are ideally suited for implementing decision-congruent confidence computations.

<u>Model Architecture</u>

Intuitively, this network's architecture can be summarized as follows. Accumulator units (with self-excitation and leak) tuned to varying stimulus alternatives accumulate momentary stimulus evidence and inhibit other units that have opposing tuning preferences. Units differ in the degree to which they are inhibited: the more normalized units are the ones that receive stronger inhibition. A discrimination decision is made when a linear combination of accumulator unit activity for a given tuning preference reaches a threshold level of evidence, and confidence is evaluated by reference to a linear combination of accumulator unit activity for the chosen stimulus alternative. Specifically, extending previous work, at each timestep we simulate the change in firing rate dx for each accumulation unit x with stimulus tuning preference i and normalization tuning level k (where i and k range from minimum values of 1 to maximum values of I and K, respectively) as:

$$dx_{ik} = B_0 + [S_i + \varepsilon_{add} + \varepsilon_{mult}] - (\lambda - \rho)x_{ik} - \beta_k \sum_j D_{ij} \left(\frac{1}{K}\sum_l x_{jl}\right)$$

Our primary innovations in the past year focused on refining the lateral inhibition component $-\beta_k \sum_j D_{ij}(\frac{1}{k}\sum_l x_{jl})$ in our model.

This term for lateral inhibition can be decomposed into two components:

1. Whole-population activation of inhibitory interneurons: $InhInt = (\frac{1}{k}\sum_l x_{jl})$

Every accumulator unit with tuning preference *j* and normalization tuning level *l* activates an inhibitory interneuron with the same tuning preference *j*, $InhInt_j$, to a degree proportional to its current firing rate x. (Subscripts *j* and *l* are used here rather than *i* and *k* because the units subscripted with *j* and *l* are summed over the whole population of neurons in determining the effect of the population on a single unit with particular tuning preference *i* and normalization tuning level *k*, i.e. in determining dx_{ik}). The summed activation across all normalization tuning levels *l* is divided by the overall number of tuning levels implemented in the simulation K. Thus, $InhInt_j$ is effectively an average of the activity of all accumulator units with tuning preference *j* computed across all normalization tuning levels *l*.

2. Inhibitory interneuron inhibition of single units: $Inh\_x_{ik} = -\beta_k \sum_j D_{ij}\ InhInt_j$

Each inhibitory interneuron in turn inhibits each accumulator unit. The degree of inhibition depends on the dissimilarity in their respective tuning preferences *j* and *i*, according to the equation:

$D_{ij} = 1 - (1/2 \cos(2\pi(i-j)/I) + 1/2)$

where I is the number of tuning preferences implemented in the simulation. D_{ij} is thus a sinusoidal function varying between 0 and 1 that represents the dissimilarity in tuning preference between the accumulator unit with tuning preference *i* and the inhibitory interneuron with tuning preference *j*.[18] Tuning preference dissimilarity is maximal when $|(i - j) / I| = 1/2$ (i.e. when the accumulator unit and inhibitory interneuron have diametrically opposing tuning preferences, such as “motion left” vs “motion right”) and minimal when $i = j$ (i.e. when they have the same tuning preference).

Inhibitory interneuron inhibition of accumulator units is further modulated by β_k. The magnitude of β_k is inversely proportional to normalization tuning level *k*, from a maximum value of 1 at $k = 1$ (full normalization tuning) to a minimum value of 0 at $k = K$ (no normalization):

$\beta_k = 1 - (k-1 / K-1)$

When $\beta_k = 0$, a unit receives no lateral inhibition and is thus not normalized. When $\beta_k = 1$, a unit is “fully normalized” in the sense that it receives momentary inhibition from the average activity of all units with opposite tuning preference (for which $D_{ij} = 1$). Intermediate levels of β_k reflect intermediate levels of normalization.

These components are shown in Figure 1:

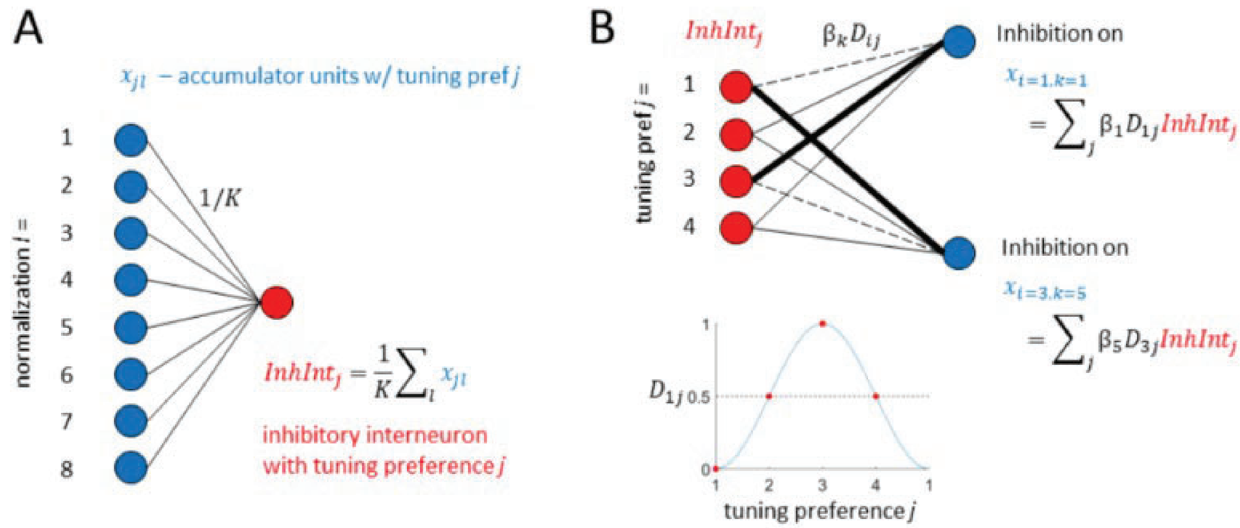


Figure 9. **Schematic of lateral inhibition in the model.** (A) *Inhibitory interneuron units.* There are accumulator units with tuning preference *j* at each level of normalization tuning *l*. Activity in these units is integrated into a single inhibitory interneuron unit with the same tuning preference, *InhIntj*, where the weighting factor on each unit is set to a constant 1/K, where K is the number of normalization tuning levels (8 in this example). (B) *Inhibition of accumulator units according to stimulus tuning preference dissimilarity.* Inhibitory interneuron pools *InhIntj* inhibit accumulator units *xik* as a sinusoidal function of the similarity in tuning preferences *i* and *j*, *Dij*. For instance, for a set of I = 4 circularly arranged tuning preferences (corresponding e.g. to motion directions left, up, right, down), an accumulator unit with tuning preference *i* = 1 receives weakest inhibition from inhibitory interneuron pools with the same tuning preference *j* = 1, and strongest inhibition from inhibitory interneuron pools with the opposite tuning preference *j* = 3. Here we illustrate an example using I = 4, although in the actual simulations we set I = 2 for simplicity, corresponding to diametrically opposed stimulus properties, e.g. motion directions left vs. right. The overall inhibition strength βk depends on normalization tuning level *k*, with more normalized units receiving stronger inhibition.

Crucially, the weighting of accumulator units in these linear combinations depends on their level of normalization tuning, and this weighting differs for discrimination decisions and confidence ratings. More normalized units are weighted more heavily for discrimination decisions, since they effectively encode the balance of evidence for one stimulus alternative versus the others by virtue of their normalization. By contrast, less normalized units are weighted more heavily for confidence ratings, since they effectively encode a more faithful representation of the raw magnitude of evidence supporting each decision alternative, regardless of evidence supporting other possible decisions.

This development in the model allows for more a more parsimonious account of current behavioral findings in the literature, including such as findings from Zylberberg (2016), shown in Figure 8:

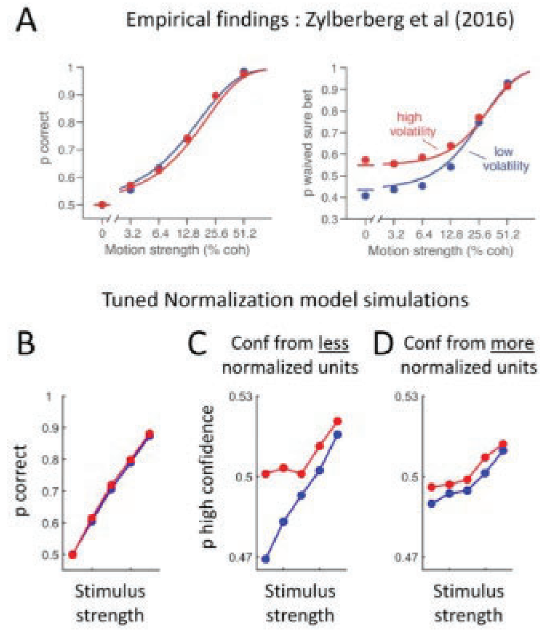


***Figure 10*. The Tuned Normalization model predicts the effect reported by Zylberberg and colleagues.**[19] (a) The authors showed that increased stimulus volatility leads to similar objective performance but increased confidence ratings, especially at low objective performance levels. This figure is reproduced from Zylberberg (2016). (b, c) The Tuned Normalization model qualitatively reproduces these effects. (d) The effect of stimulus volatility on confidence was nearly abolished in control simulations in which more normalized units are weighted more heavily for computing confidence, suggesting that the result in (b) critically depends on confidence being computed primarily from less normalized units. Cohen's d for confidence effects in the main model simulations was ~1.5 - 6 times greater than in the control simulations across all levels of stimulus strength.

We demonstrated that normalization tuning provides a biologically plausible mechanism for implementing confidence computations that demonstrate an overreliance on decision-congruent information. Our findings lead to testable hypotheses about the role of tuned normalization in a neuron's contribution to a decision versus a confidence judgment: activity of more normalized units should reflect an observer's objective decisions more than confidence judgments, while the opposite should be true for less normalized neurons. The present findings thus pave the way for noninvasive neuroscience techniques, such as spatially coarser functional MRI in humans, to clarify the role of normalization tuning in perceptual and cognitive decisions and metacognitive evaluations of these choices.

5 Conclusions & Future Work

We investigated how subjects detected colorful stimuli in the unattended periphery in a simulated driving task. If people tend to confirm what they expect whenever they don't attend, this raises an important issue regarding driving safety. Hazard detection is a special detection task, because the penalty for a miss and a false alarm is different: a miss may cause a crash, while a false alarm may only lead to unnecessary brakes; thus, the criterion in hazard detection task is closely related with driving safety. In our research, the task is to detect pedestrian when the vehicle has stopped. However, in real driving, most of the hazards are presented while driving, which is quite different from our task. Future research should more systematically address how inattention affects hazard detection judgment while driving.

We also considered how peripheral visual perception may demonstrate inflation, whereby subjective judgements in this region of space are marked by two behavioral characteristics: metacognitive impairments in how effectively confidence judgements track the correctness of responses in experimental tasks, and decisional biases in observers' tendencies to assume stimuli are more likely to be presented in the periphery than what actually occurs. Our experimental findings showed that our subjective sense of the visual surround is inflated. These findings raise an important question: what may be a mechanistic explanation for inflation? Future investigations should aim to systematically investigate how exogenous attention and endogenous attention may alter the characteristics of inflation that we observed.

We developed a simple leaky competing accumulator neural network model incorporating a known property of sensory neurons: tuned normalization. The key idea of the model is that each accumulator neuron's normalization 'tuning' dictates its contribution to perceptual decisions versus confidence judgments. We demonstrate that this biologically plausible model can account for several counterintuitive findings reported in the literature, where confidence and decision accuracy were shown to dissociate -- and that the differential contribution a neuron makes to decisions versus confidence judgments based on its normalization tuning is vital to capturing some of these effects. We confirmed the model's biological substrate using electrophysiological recordings in monkeys. These results challenge the dominant model, suggesting that the brain instead capitalizes on the diversity of available machinery (i.e., neuronal resources) to implement heuristic -- not optimal -- strategies to compute subjective confidence.

6 Supported Students & Personnel

Megan Peters
Megan Peters was a postdoctoral scholar in Psychology between the years of 2014-2017 at the University of California, Los Angeles before obtaining an Assistant Professor position at the University of California, Riverside in the Department of Bioengineering. Her research focuses on how the brain represents and uses uncertainty, and performs adaptive computations based on incomplete information. She uses neuroimaging, computational modeling, machine learning and neural stimulation techniques to study these topics.

Brian Odegaard
Brian Odegaard was a postdoctoral scholar in Psychology between the years of 2015-2019 at the University of California, Los Angeles before obtaining an Assistant Professor position at the University of Florida in the Department of Psychology. His research focuses on mid- and high-level vision, attention and consciousness, multisensory integration, brain plasticity and sensory recalibration, perceptual metacognition and sensory processing in clinical populations. He uses computational (Bayesian) modeling and machine learning to study these topics.

JD Knotts
JD Knotts is a PhD student in Psychology at the University of California, Los Angeles. His research focuses on are visual perception, visual metacognition, neural correlates of consciousness, multisensory binding and casual inference in perception. He uses Bayesian statistical modeling of cognition to study these topics.

7 Projection Publications

1. Li, M. K., Lau, H. & Odegaard, B. An investigation of detection biases in the unattended periphery during simulated driving. *Atten. Percept.Psychophys*. 80, 1325–1332 (2018).https://doi.org/10.3758/s13414-018-1554-3
2. Odegaard, B., Chang, M. Y., Lau, H. & Cheung, S.-H. Inflation versus filling-in: why we feel we see more than we actually do in peripheralvision. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 373, (2018).https://doi.org/10.1098/rstb.2017.0345
3. Odegaard, B., Li, M. & Lau, H. Subjective inflation in the unattended periphery in a naturalistic environment. *J. Vis.* (2018).http://dx.doi.org/10.1167/18.10.530
4. Knotts, J. D., Odegaard, B., Lau, H. & Rosenthal, D. Subjective inflation: phenomenology's get-rich-quick scheme. *Curr Opin Psychol* 29, 49–55(2018).https://doi.org/10.1016/j.copsyc.2018.11.006
5. Maniscalco, B., Odegaard, B., Grimaldi, P. Cho, S.H., Basso, Michele A., Lau, H. and Peters, M.A.K. Tuned normalization in perceptual decision-making circuits can explain seemingly suboptimal confidence behavior. *Journal of Vision*. (forthcoming). https://doi.org/10.1101/558858

8 References

1. Block, Ned. 2011. "Perceptual Consciousness Overflows Cognitive Access." *Trends in Cognitive Sciences* 15 (12): 567–75.
2. Carrasco, M. "Visual Attention." *Vision research* 51, no. 13 (2011): 1484-525.
3. Cohen, Michael A., Daniel C. Dennett, and Nancy Kanwisher. 2016. "What Is the Bandwidth of Perceptual Experience?" *Trends in Cognitive Sciences* 20 (5): 324–35.
4. Rahnev, D., B. Maniscalco, T. Graves, E. Huang, F. P. de Lange, and H. Lau. "Attention Induces Conservative Subjective Biases in Visual Perception." *Nature Neuroscience* 14 (2011): 1513-15.
5. Giles, N., Lau, H., Odegaard, B. (2016). What Type of Awareness Does Binocular Rivalry Assess? *Trends in Cognitive Science*, 20(10), 719–720. doi: 10.1016/j.tics.2016.08.010.
6. Green, D. M., and J. A. Swets. Signal Detection Theory and Psychophysics. Vol. 1974: Wiley New York, 1966.
7. Gorea, A., and D. Sagi. "Failure to Handle More Than One Internal Representation in Visual Detection Tasks." *Proceedings of the National Academy of Sciences* 97, no. 22 (2000): 12380.
8. McDonnell, M. D., and D. Abbott. "What Is Stochastic Resonance? Definitions, Misconceptions, Debates, and Its Relevance to Biology." *PLoS computational biology* 5, no. 5 (2009): e1000348.
9. Schwarzkopf, D. S., J. Silvanto, and G. Rees. "Stochastic Resonance Effects Reveal the Neural Mechanisms of Transcranial Magnetic Stimulation." *The Journal of Neuroscience* 31, no. 9 (2011): 3143-47.
10. Banks, M. S., A. B. Sekuler, and S. J. Anderson. "Peripheral Spatial Vision: Limits Imposed by Optics, Photoreceptors, and Receptor Pooling." *JOSA A* 8, no. 11 (1991): 1775-87.
11. Rosenholtz, R., J. Huang, and K. Ehinger. "Frontiers: Rethinking the Role of Top-Down Attention in Vision: Effects Attributable to a Lossy Representation in Peripheral Vision." *Frontiers in Consciousness Research* 3, no. 13 (2012).
12. Ratcliff, R., and G. McKoon. "The Diffusion Decision Model: Theory and Data for Two-Choice Decision Tasks." *Neural computation* 20, no. 4 (2008): 873-922.
13. Dayan, P., and L. Abbott. "Theoretical Neuroscience, 2001." MIT Press, Cambridge MA, Chapt.
14. Ma, W. J., J. M. Beck, P. E. Latham, and A. Pouget. "Bayesian Inference with Probabilistic Population Codes." *Nature neuroscience* 9, no. 11 (2006): 1432-38.
15. Ma, W. J., J. M. Beck, and A. Pouget. "Spiking Networks for Bayesian Inference and Choice." *Current opinion in neurobiology* 18, no. 2 (2008): 217-22.
16. Mitchell, J. F., K. A. Sundberg, and J. H. Reynolds. "Spatial Attention Decorrelates Intrinsic Activity Fluctuations in Macaque Area V4." *Neuron* 63, no. 6 (2009): 879-88.
17. Cohen, M. R., and J. H. R. Maunsell. "Attention Improves Performance Primarily by Reducing Interneuronal Correlations." *Nature neuroscience* 12, no. 12 (2009): 1594-600.
18. Zohary, E., M. N. Shadlen, and W. T. Newsome. "Correlated Neuronal Discharge Rate and Its Implications for Psychophysical Performance." *Nature* 370, no. 6485 (1994): 140-43.
19. Azzopardi, P., and A. Cowey. "Preferential Representation of the Fovea in the Primary Visual Cortex." *Nature* 361, no. 6414 (1993): 719-21.
20. Knill, D. C., and A. Pouget. "The Bayesian Brain: The Role of Uncertainty in Neural Coding and Computation." *TRENDS in Neurosciences* 27, no. 12 (2004): 712-19.
21. Kahneman, D. "A Perspective on Judgment and Choice: Mapping Bounded Rationality." *American psychologist* 58, no. 9 (2003): 697.
22. Reynolds, J. H., and D. J. Heeger. "The Normalization Model of Attention." *Neuron* 61, no. 2 (2009): 168-85.
23. Carandini, M., and D. J. Heeger. "Normalization as a Canonical Neural Computation." *Nature Reviews Neuroscience* 13, no. 1 (2011): 51-62.

24. Li, M. K., Lau, H. & Odegaard, B. An investigation of detection biases in the unattended periphery during simulated driving. *Atten. Percept. Psychophys.* 80, 1325–1332 (2018).
25. Odegaard, B., Chang, M. Y., Lau, H. & Cheung, S.-H. Inflation versus filling-in: why we feel we see more than we actually do in peripheral vision. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 373, (2018).
26. Odegaard, B., Li, M. & Lau, H. Subjective inflation in the unattended periphery in a naturalistic environment. *J.Vis.* (2018).
27. Kouider, S., de Gardelle, V., Sackur, J. & Dupoux, E. How rich is consciousness? The partial awareness hypothesis. *Trends Cogn. Sci.* 14, 301–307 (2010).
28. Kouider, S. & Dupoux, E. Partial Awareness Creates the 'Illusion' of Subliminal Semantic Priming. *Psychol. Sci.* 15, 75–81 (2004).
29. Cohen, M. A., Dennett, D. C. & Kanwisher, N. What is the Bandwidth of Perceptual Experience? *Trends Cogn. Sci.* 20, 324–335 (2016).
30. McConkie, G. W. & Rayner, K. The span of the effective stimulus during a fixation in reading. *Percept. Psychophys.* 17, 578–586 (1975).
31. Balas, B. & Sinha, P. 'Filling-in' colour in natural scenes. *Vis. cogn.* 15, 765–778 (2007).
32. Suárez-Pinilla, M., Seth, A. K. & Roseboom, W. The Illusion of Uniformity Does Not Depend on the Primary Visual Cortex: Evidence From Sensory Adaptation. *i-Perception* 9, 204166951880050 (2018).
33. Block, N. On a confusion about a function of consciousness. *Behav. Brain Sci*. (1995).
34. Block, N. Consciousness, accessibility, and the mesh between psychology and neuroscience. *Behav. Brain Sci*. 30, 481–99; discussion 499–548 (2007).
35. Lamme, V. A. F. Why visual attention and awareness are different. *Trends Cogn. Sci.* 7, 12–18 (2003).
36. Lamme, V. A. F. How neuroscience will change our view on consciousness. *Cogn. Neurosci.* 1, 204–220 (2010).
37. Azzopardi, P. & Cowey, A. Preferential representation of the fovea in the primary visual cortex. *Nature* 361, 719–721 (1993).
38. Strasburger, H., Rentschler, I. & Jüttner, M. Peripheral vision and pattern recognition: a review. *J. Vis.* 11, 13 (2011).
39. Knotts, J. D., Odegaard, B., Lau, H. & Rosenthal, D. Subjective inflation: phenomenology's get-rich-quick scheme. *Curr Opin Psychol* 29, 49–55 (2018).
40. Usher, M. & McClelland, J. L. The time course of perceptual choice: the leaky, competing accumulator model. *Psychol. Rev.* 108, 550–592 (2001).
41. Deneve, S., Latham, P. E. & Pouget, A. Reading population codes: a neural implementation of ideal observers. *Nat. Neurosci.* 2, 740–745 (1999).
42. Zylberberg, A., Fetsch, C. R. & Shadlen, M. N. The influence of evidence volatility on choice, reaction time and confidence in a perceptual decision. *Elife* 5, (2016).
43. Ehinger, Benedikt V., Katja Häusser, José P. Ossandón, and Peter König. 2017. "Humans Treat Unreliable Filled-in Percepts as More Real than Veridical Ones." *eLife* 6 (May). https://doi.org/10.7554/eLife.21761.
44. Fleming, Stephen M., and Hakwan C. Lau. 2014. "How to Measure Metacognition." *Frontiers in Human Neuroscience 8* (July): 443.
45. Haigney, D. E., R. G. Taylor, and S. J. Westerman. 2000. "Concurrent Mobile (cellular) Phone Use and Driving Performance: Task Demand Characteristics and Compensatory Processes." Transportation Research. Part F, *Traffic Psychology and Behaviour* 3 (3): 113–21.
46. Haun, Andrew M., Giulio Tononi, Christof Koch, and Naotsugu Tsuchiya. 2017. "Are We Underestimating the Richness of Visual Experience?" Neuroscience of Consciousness 2017 (1). Oxford University Press. https://doi.org/10.1093/nc/niw023.
47. Komatsu, Hidehiko. 2006. "The Neural Mechanisms of Perceptual Filling-In." *Nature Reviews. Neuroscience* 7 (3): 220–31.

48. Levin, Daniel T., Nausheen Momen, Sarah B. Drivdahl, and Daniel J. Simons. 2000. "Change Blindness Blindness: The Metacognitive Error of Overestimating Change-Detection Ability." *Visual Cognition* 7 (1-3). Routledge: 397–412.
49. Maniscalco, Brian, and Hakwan Lau. 2014. "Signal Detection Theory Analysis of Type 1 and Type 2 Data: Meta-D′, Response-Specific Meta-D′, and the Unequal Variance SDT Model." *In The Cognitive Neuroscience of Metacognition*, 25–66. Springer, Berlin, Heidelberg.
50. Phillips, Ian. 2016. "Naive Realism and the Science of (Some) Illusions." *Philosophical Topics* 44 (2): 353–80.
51. Rahnev, Dobromir, Brian Maniscalco, Tashina Graves, Elliott Huang, Floris P. de Lange, and Hakwan Lau. 2011. "Attention Induces Conservative Subjective Biases in Visual Perception." *Nature Neuroscience* 14 (12). Nature Research: 1513–15.
52. Rumschlag, Gordon, Theresa Palumbo, Amber Martin, Doreen Head, Rajiv George, and Randall L. Commissaris. 2015. "The Effects of Texting on Driving Performance in a Driving Simulator: The Influence of Driver Age." Accident; *Analysis and Prevention* 74 (January): 145–49.
53. Sharot, Tali. 2011. "The Optimism Bias." *Current Biology*: CB 21 (23): R941–45.
54. Sharot, Tali, Marc Guitart-Masip, Christoph W. Korn, Rumana Chowdhury, and Raymond J. Dolan. 2012. "How Dopamine Enhances an Optimism Bias in Humans." *Current Biology*: CB 22 (16): 1477–81.
55. Sharot, Tali, Alison M. Riccardi, Candace M. Raio, and Elizabeth A. Phelps. 2007. "Neural Mechanisms Mediating Optimism Bias." *Nature* 450 (7166): 102–5.
56. Solovey, Guillermo, Guy Gerard Graney, and Hakwan Lau. 2015. "A Decisional Account of Subjective Inflation of Visual Perception at the Periphery." *Attention, Perception & Psychophysics* 77 (1): 258–71.
57. Usher, Marius, and James L. McClelland. 2001. "The Time Course of Perceptual Choice: The Leaky, Competing Accumulator Model." Psychological Review 108 (3). The Macmillan Company: 550–92.
58. Vandenbroucke, Annelinde R. E., Ilja G. Sligte, Adam B. Barrett, Anil K. Seth, Johannes J. Fahrenfort, and Victor A. F. Lamme. 2014. "Accurate Metacognition for Visual Sensory Memory Representations." *Psychological Science* 25 (4): 861–73.
59. Ventsislavova, Petya, Andres Gugliotta, Elsa Peña-Suarez, Pedro Garcia-Fernandez, Eduardo Eisman, David Crundall, and Candida Castro. 2016. "What Happens When Drivers Face Hazards on the Road?" *Accident; Analysis and Prevention* 91 (June): 43–54.
60. Wallis, Thomas S. A., and Mark S. Horswill. 2007. "Using Fuzzy Signal Detection Theory to Determine Why Experienced and Trained Drivers Respond Faster than Novices in a Hazard Perception Test." *Accident; Analysis and Prevention* 39 (6): 1177–85.
61. Witt, Jessica K., J. Eric T. Taylor, Mila Sugovic, and John T. Wixted. 2015. "Signal Detection Measures Cannot Distinguish Perceptual Biases from Response Biases." *Perception* 44 (3): 289–300.

Appendix

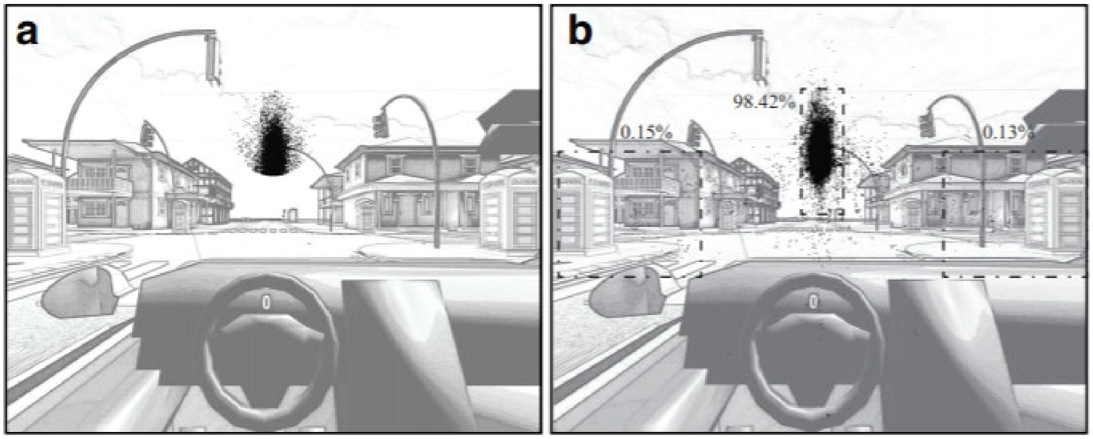


Figure 1. Participants' fixations in Experiment 1 in Li, Lau and Odegaard (2018). a Participants' fixations at 17 ms (i.e., time of one frame) before stimuli (the two male pedestrians) were presented, from every trial in Experiment 1. Each black point represents a fixation from one of the two eyes, and fixation points from both eyes are plotted in the figure. Almost all fixations are on or near the stoplight in every trial. b Participants' fixations during stimulus presentation (300 ms total) in every trial in Experiment 1.

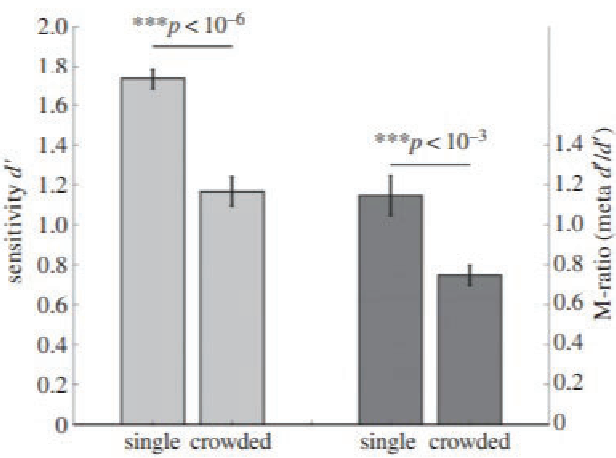


Figure 2. Perceptual sensitivity and metacognitive efficiency in an orientation discrimination task from Odegaard, Chang, Lau and Cheung (2018). Shown here are the results from 23 participants in Experiment 1. As shown by the light grey bars, participants were much less effective at discriminating the orientation of a tilted grating when it was surrounded by other gratings (the 'crowded' condition), compared to when it was presented alone (the 'single' condition). d0 is the standard detection theoretic measure of sensitivity. The dark grey bars show a measure of the metacognitive efficiency (the M-ratio; meta-d0 /d0) in both conditions, which indicates how effectively confidence ratings could distinguish between correct and incorrect judgements. As can be seen in the figure, metacognitive efficiency was impaired in the crowded condition compared to the single condition.

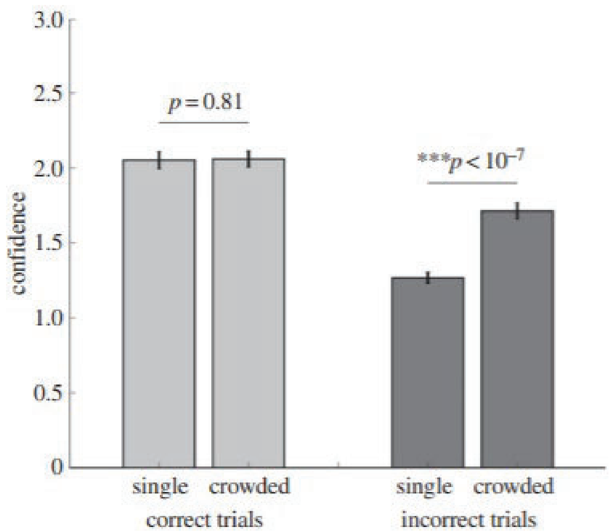


Figure 3. Average confidence for correct and incorrect trials task from Odegaard, Chang, Lau and Cheung (2018). The light grey bars indicate the average confidence for correct trials for the single and crowded conditions. As can be seen in the figure, the difference between these conditions is not significant. The dark grey bars indicate the average confidence for incorrect trials. A clear difference between the single and crowded conditions is evident, and participants are significantly more confident in the incorrect crowded trials compared to the incorrect single trials.