

UC Davis

UC Davis Electronic Theses and Dissertations

Title

The Time Course of Conscious Perception: A Computational Cognitive Neuroscience Analysis

Permalink

<https://escholarship.org/uc/item/0kj4n9qz>

Author

Krist, Lara

Publication Date

2022

Peer reviewed|Thesis/dissertation

The Time Course of Conscious Perception: A Computational Cognitive Neuroscience Analysis

By

LARA KRISST
DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Psychology

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Steven J. Luck, Chair

Andrew Yonelinas

John Henderson

Committee in Charge

2022

Acknowledgments

These are the giants upon which I stand. Thank you Kenneth Solepak for not allowing me to slip through the cracks in the third grade – life may have been very different. Thank you to Stephan Johnson for engaging me in philosophy and encouraging my continued education. Thank you to the UC Berkeley Department of Philosophy for demanding academic excellence and shaping me into a scholar. Specifically, I would like to thank John Searle, Barry Stroud, and John Campbell for imparting a lifetime of philosophical knowledge with such intellectual generosity – your classes were nothing short of mesmerizing. Thank you to George Lakoff for the generous support. Thank you to my former advisor, Ezequiel Morsella, for kind mentorship and ongoing guidance. Thank you to my mentor and advisor, Steve Luck, whose dedicated support and guidance have immeasurably impacted my career. I am left with gratitude for his invaluable feedback over the years which has always served to sharpen my thinking and fine-tune my approach to the scientific process.

I would like to thank my committee members, Zoe Drayson, Ron Mangun, Andrew Yonelinas, and John Henderson for their support and feedback. Thank you to my mentor Eve Isham and my colleague Angela Nazarian. A special thanks for the support of the Luck Lab senior staff past and present, Felix Bacigalupo, Gi-Yeul Bae, Andrew Stewart, Orestis Papaioannou, John Kiat, Brett Bahle, Carlos Carrasco, Kurt Winsler, David Railton Garrett, and a special thanks to our wizard-problem-solver lab manager, Aaron Matthew Simmons. Thank you to my research assistants Alyse Lisette Lodigiani, Kushaal Rao, and Peter Lawrence Jespersen for their dedication. Thank you to members of the Consciousness & Cognition Journal Club for many years of stimulating discussions, and thank you to SSNAP fellows, Jennifer Lee, Trey Boone, and Gerardo Viera for the inspiring discourse during the collaborative process.

None of this would be possible without the steadfast support of my mother, Dr. Ilze Krisst. It cannot be overstated how her presence in my life has been a pillar of strength, inspiration, reason, honesty, decency and love which has kept my life on course. Thank you for the support of the Krisst family, Abe, Rima, Danette, Devin, Baden, Veide, Immanuel, Anne, and my father, Dr. Raymond Krisst.

I am fortunate to have such remarkable friends. Thank you to Rachel Leamy, Karen Creech, Jami Bailey, Nicole Leveck, and the entire San Francisco crew whose ongoing emotional support and assistance have been foundational during this process. Thank you to my Davis friends Laura Delgado, Jennifer Miller, Stephanie White, Noelle Reis, Lauren Fausto, Valya Mogilner, JJ Lee, and Peter Green for their constant friendship and support. A special thanks also to Rebecca Millsop whose intellect, kindness, and friendship at UC Berkeley influenced my path in ways it is hard to quantify.

Thank you to all the friends near and far who have encouraged and supported me along the way. Lastly, thank you to those who have tended to and warmed my heart, you know who you are.

Abstract

To further understand the neural basis of conscious experience, we employ cutting edge machine learning techniques to home in on neural activity that correlates with specific *conscious content* or specific information that becomes accessible to awareness (e.g., the experience of pain, the color blue). In the context of differing views regarding the functional localization and neural underpinnings of the neural correlates of consciousness, we narrow the focus of disagreement to testable theoretical predictions concerning the *timing* of conscious visual representational content. Our findings are consistent with theories that propose a late electrophysiological correlate of visual awareness.

I. Introduction

Over the last several decades new methodological and modeling techniques have contributed to progress in isolating the notoriously elusive *neural correlates of consciousness* (NCC). NCC is a term coined by Christof Koch and is defined as the “minimal set of neuronal events and mechanisms jointly sufficient for a specific conscious percept” (Crick & Koch, 1990). When investigating the neural basis of consciousness, it is important to make the common-sense distinction between two different aspects of consciousness: 1) *conscious states* – referring to the *level* of consciousness or wakefulness (ranging from coma to sleep to the waking state), and 2) *conscious contents* – referring to specific (pieces of) information that become accessible to awareness. In this dissertation, we focus on the latter characterization of the NCC. For example, neurons that respond to vertical stimuli if and only if a monkey reports awareness of a vertical orientation during a binocular rivalry task is one such case of isolating the representational contents of the neural systems that correspond with awareness (Logothetis & Schall, 1989). Here we embark on the nontrivial task of determining how neural activity in humans differs for the informational content that are aware of versus the informational content that we are not aware of.

Recently, EEG and MEG experiments have employed multivariate decoding techniques to examine putative NCCs (King, Pescetelli, & Dehaene, 2016; Salti et al., 2015). The combination of these techniques provides a powerful tool for investigating the temporal dynamics of representational content in the visual processing stream as it pertains to the content of conscious experience. To further understand the nature of conscious visual perception, Chapter I uses psychophysics and EEG decoding to track the evolution of visual representations over time, focusing on the idea that our conscious perception of orientation is largely invariant over wide variations in luminance contrast, even though neural activity in early visual cortex is highly

contrast-dependent. If the conscious perception of orientation is not influenced by contrast, but a given neural signal is, then that signal does not seem like a plausible NCC.

Experiment 1 of Chapter I uses psychophysics to quantify the representational precision of low- and high-contrast oriented stimuli. *Experiment 2* examines the neural underpinnings of contrast-invariant orientation perception with EEG decoding, focusing on how contrast invariance evolves over time. Since previous research has shown a correspondence between layers of a deep convolutional neural network (CNN) and the human visual processing stream, we used representational similarity analysis link the EEG data to a CNN, making it possible to more closely examine the progression of contrast invariance. Furthermore, since similarities have been found between units in a CNN network and single neurons for orientation tuning, *Experiment 3* examines the potential mechanism of contrast invariance within the CNN architecture itself.

To further understand the time course of conscious perception, Chapter II uses ERP decoding methods to decode perceived and nonperceived orientation information during binocular-rivalry. We compare decoding accuracy of contents reported in the perceptually dominant eye (conscious information content) with content in the non-dominant eye (unconscious information content). We test two major theories of consciousness that make differing predictions about *when* awareness occurs in the visual processing hierarchy. Global Workspace Theory predicts a difference in conscious and unconscious processing would occur at ~300 ms after stimulus onset, alternatively, the view that an earlier electrophysiological correlate of visual awareness reflects conscious perception (e.g., Visual Awareness Negativity) predicts that the difference in conscious and unconscious processing occurs at ~200 ms.

II.

Chapter I: ERP Decoding of Low-and-High-Contrast Orientations: Implications for Contrast-Independent Perception

Abstract

Even though neurons in early visual areas respond in a contrast-dependent manner, we can consciously perceive the form of a stimulus even at relatively low contrast. To test the hypothesis that the precision of these high-level conscious representations remains relatively constant at low levels of contrast, we conducted two experiments. The first was a psychophysical experiment to demonstrate that at low contrast people could have a precise representation of the orientation of a teardrop shape. Teardrop-shaped stimuli were presented as dark stimuli on a grey background. The teardrop ranged from black (80.5% contrast) to only slightly darker than the background (1% contrast). Participants were instructed to remember the orientation of the teardrop and to reproduce it after a 1500 ms delay. We found that as the contrast level was reduced, the precision was unaffected, although at the lowest contrast we saw an increase in the guess rate. We then performed an ERP decoding experiment where we examined the ability to decode the orientation of the teardrop from the scalp distribution of the ERP at each moment in time following the onset of the stimulus. This task was similar to the behavioral task, but only two contrast levels were used. This experiment allowed us to test the hypothesis that early sensory activity would be reduced and therefore undecodable at low contrast, whereas later, higher-level (presumably conscious) processing would be decodable in both high and low contrast conditions. Results were consistent with our hypothesis, showing that during the early sensory period (~50-150 ms) decoding accuracy was much greater for high contrast than low contrast. By 200 ms, decoding accuracy was nearly as great for low contrast stimuli as for high contrast stimuli. We also used representational similarity analysis to examine the link between the neural signal at each time point and the

individual layers of a deep convolutional neural network (CNN) that models the human ventral processing stream. The hypothesis was that activation in early layers of the model would show representational similarity with the EEG data in the high contrast condition compared with the low contrast condition. The results trended in that direction but were not significant. A third experiment examined the CNN itself, finding that the responses of single units were less contrast-dependent in the higher layers of the model. These findings suggest that early visual areas are not encoding the conscious perception of orientation, but awareness instead requires extended processing over time to extract a largely contrast independent perception of orientation.

Introduction

Decades of research on the ventral object processing pathway has revealed a cascade of neurons coding increasingly complex features. Early areas of this hierarchy code basic visual features such as luminance, color and orientation, whereas higher level areas code categories, objects and abstract forms (Malach et al., 2002; Riesenhuber & Poggio, 1999). A key computational goal of this pathway is to code the shape of an object independently of its other features, such as its location, direction of motion, and color. For example, neurons in V1 are highly sensitive to an object's retinotopic location, whereas neurons in inferotemporal cortex have large receptive fields that allow them to respond in a largely location-independent manner (Gross & Mishkin, 1977). Although the ability to code the shape of an object independently of its location is well understood, the ability to code shape independently of contrast is not.

Figure 1 illustrates that increasing the contrast of a stimulus leads to a multiplicative increase in the responses of orientation-selective cells in early visual cortex. That is, the effect of contrast is small for nonpreferred orientations and large for preferred orientations. As a result, the firing rate of a given neuron is ambiguous: different combinations

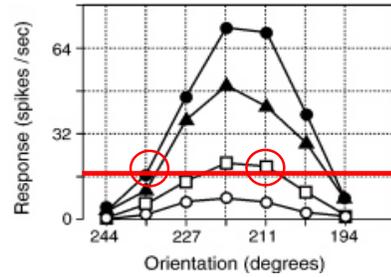


Figure 1. Orientation tuning curves of a neuron from single unit recordings in macaques measuring a stimulus that varied in contrast. Empty circles are 10% contrast and filled circles are 80% contrast.

of contrast and orientation lead to the same firing rate. More specifically, Figure 1 shows the orientation tuning curves of a neuron (from single unit recordings in macaques), which was measured using a stimulus that varied in contrast. The empty circles are 10% contrast and the filled circles are 80% contrast. If we cut a line across the Y axis, we see that there is some

CONTRAST

CONTRAST

Figure 2. Each letter is identifiable whether presented at low or high contrast, while separately perceiving that any given letter looks gray (low contrast) or black (high contrast).

ambiguity as to which orientation is present, because you can get exactly the same firing rate for a nonpreferred orientation at high contrast and a preferred orientation at low contrast (red circles). In our everyday subjective experience though, our conscious perception separates the contrast of a stimulus from its shape (as long as the contrast is modestly above threshold). For example, as illustrated in Figure 2, when we see a stimulus such as the word “CONTRAST” at different contrasts, we can clearly identify each letter whether it is presented at low or high contrast, while separately perceiving that any given letter looks gray (low contrast) or black (high contrast). In other words, whereas contrast is multiplicatively combined with other features (like orientation) in early visual areas (i.e., for any given neuron it is difficult to know what is happening), the information across the whole population is obviously there in order for us to have the conscious perception of both shape and contrast. So presumably there is a transition process which takes the information at the population level and then enables

us to perceive shapes largely independently of contrast, where we can identify both the letter identities and whether they look grey or black. In this way, conscious perception is able to separate the shape of an object from its contrast – it is subjectively contrast invariant.

In the present study, we examined this ‘subjective contrast invariance’ behaviorally by presenting teardrop-shaped stimuli as dark stimuli on a grey background (Figure 3). The teardrop stimulus was presented on a grey background at one of four luminance contrast values. A black teardrop at 80.5% contrast was the darkest stimulus, while the other three

stimuli were only slightly darker than the background at 2%, 1%, and 0.03% contrast. The teardrop was presented randomly at one of sixteen discrete orientations, and participants were instructed to remember the orientation of the teardrop and to reproduce it after a 1500 ms delay.

We hypothesized that if the perception of the shape of an object is independent of the contrast, assuming it is supraliminal, subjects are able to have a fully precise representation of the stimuli even when measured in working memory. We further tested this hypothesis using ERP decoding, which allowed us to examine the neural activity during the processing of the stimulus, instead of long after the stimulus is gone.

Some research has already explored the transition from contrast-dependent to contrast-independent coding in the primate visual system (Alitto & Usrey, 2004; Anderson, 2000; Avidan et al., 2002; Cheng et al., 1994; Reynolds & Chelazzi, 2004; Rolls & Baylis, 1986; Skottun et al., 1987). For example, single-cell recordings in monkeys found that the responses of face-selective

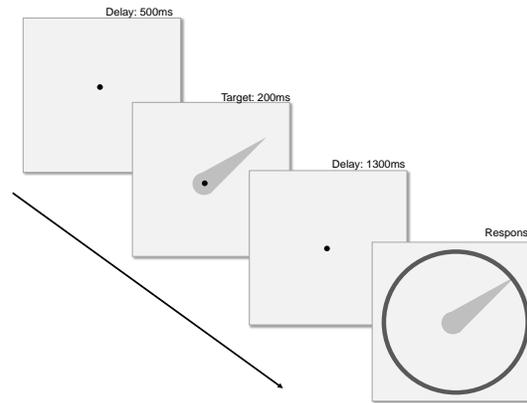


Figure 3. Schematic of task. A teardrop stimulus was presented on a grey background at one of four luminance contrast values: 80.5%, 2%, 1%, and 0.03% contrast. Participants were instructed to remember the orientation of the teardrop and to reproduce it after a 300 ms delay.

neurons in the superior temporal sulcus were relatively invariant with respect to the contrast of the face (Reynolds & Chelazzi, 2004; Rolls & Baylis, 1986). Human fMRI research has also found that neural responses to objects and faces became increasingly contrast invariant at higher levels of the ventral stream (Avidan et al., 2002). Research and applications in image processing and computer vision have also endeavored to find a solution to the problem of contrast invariance (Monasse, 1999; Monasse & Guichard, 2000; Ohayon et al., 2012; Xia et al., 2010).

In the present study, we took advantage of powerful pattern classification methods which make it possible to decode detailed, content-specific visual representations using scalp EEG recordings. Unlike fMRI, where the BOLD signal is a sluggish and secondary consequence of neural activity, EEG has the advantage of directly reflecting the extracellular voltage fields produced by neurotransmission in real time (Buzsáki et al., 2012). Recent EEG decoding studies have successfully decoded low-level features along continuous dimensions (e.g., orientation, motion), and high level categorical representations (e.g., faces, words; Bae & Luck, 2018). These powerful multivariate analysis techniques make it possible to track and decode the evolution of mental representations along the visual processing hierarchy with precise temporal resolution. In the present study, we took advantage of these decoding methods to examine the evolution of contrast-dependent coding of orientation and later, contrast-independent coding over time. We hypothesized that the initial brain response would be contrast-dependent, but later in the processing hierarchy, later in time, there would be evidence for a contrast-independent representation.

Previous research has shown a hierarchical correspondence between the layers of deep convolutional neural network networks (CNN) and the human visual processing stream (Cichy et al., 2016; Mohsenzadeh et al., 2020). Specifically, these studies showed that early layers of the

network map onto low level visual areas whereas later layers of the network map onto higher-level visual areas. Therefore, to further examine the flow of information in the visual processing hierarchy, or the progression of contrast invariance in an artificial model, we linked the EEG data to a CNN (Alexnet) using *representational similarity analysis*. We tested the hypothesis that activation in early layers of the model would show greater representational similarity with the EEG data in the high contrast condition than in the low contrast condition, an effect that would disappear in latter layers. Furthermore, because similarities have been found between units in a CNN network and single neurons for orientation tuning, we examined the potential mechanism of contrast invariance in this highly simplified CNN architecture by feeding our teardrop stimuli at varying contrast levels into the CNN and examining how the activation values of the units varied as a function of contrast (much as classical electrophysiological research examined how single-unit activity varied as a function of contrast).

Experiment 1: Psychophysics

Participants

University of California, Davis students ($n = 18$, *median age* = 21, *male* = 7, *female* = 11) received monetary compensation for their participation. All participants reported normal or corrected-to-normal vision and no neurological disorders. The experimental protocol was approved by the University of California-Davis Institutional Review Board, and informed consent was obtained before each session.

Stimuli & Procedure

The stimuli and task are illustrated in Figure 3, and were adapted from Bae and Luck (2018). All stimuli were generated in Matlab (MathWorks) using PsychToolbox (Brainard, 1997;

Pelli, 1997). Stimuli were presented on a 60 Hz LCD monitor (Dell U2412M) at a viewing distance of 100 cm. A black fixation dot was presented throughout the trial, except during the response interval. Each trial began with the fixation dot (500 ms), followed by a teardrop shaped stimulus (200 ms; 2.17° long, 0.8° wide) centered on the fixation dot. The teardrop offset was followed by a blank delay period (1300 ms). Sixteen discrete teardrop orientations were randomly presented with equal probability (0°, 22.5°, 45°, 67.5°, 90°, 112.5°, 135°, 157.5°, 180°, 202.5°, 225°, 247.5°, 270°, 292.5°, 315°, and 337.5°). Participants were instructed to remember the orientation of the teardrop over the delay period. After the delay period, a response ring was presented (radius 2.17°) and once participants moved the mouse, a sample teardrop appeared in the center of the ring. The orientation of the sample teardrop during the response period corresponded to the current position of the mouse. Participants used the mouse to rotate the tip of the teardrop to a position which corresponded to the orientation of the target teardrop, and they clicked the mouse to finalize their response. The response was followed by an intertrial interval (1000 ms). Participants completed 16 practice trials before the critical trials. Critical trials consisted of 640 trials (40 trials x 16 orientations) in random order. In Experiment 1, the teardrop stimulus was presented on a grey background (25 cd/m²) at one of four luminance contrast values. A black teardrop at 80.5% contrast (2.7 cd/m²) was the darkest stimulus, while the other three stimuli were only slightly darker than the background at 2% (23.6 cd/m²), 1% (24 cd/m²), and 0.03% (24.5 cd/m²) contrast. Contrast is given in Michelson units, $CM = (L_{max} - L_{min}) / (L_{max} + L_{min})$.

Results and Discussion

Experiment 1 Behavioral Data. Participants' mean time to report the orientation of the teardrop (via the mouse button) was 1191 ms (single trial range was 334-3053 ms; subtracting the fastest and slowest 2.5% of trials). Performance was assessed by response error, i.e., the angular difference between the target orientation and the reported orientation. We fit a mixture model, which characterizes the response errors as a mixture of 1) a von Mises distribution centered at 0 error (perfect memory of orientation), and 2) a uniform distribution (random guesses). This yields a precision parameter (κ) from the von Mises distribution, and a guess parameter (pmem), denoting the probability of a random guess, from the uniform distribution.

Probability distributions of response errors collapsed across all participants as a function of contrast are summarized in Figure 4a. Precision (κ) was largely constant across contrast levels, but the probability of a random guess increased (leading to a lower pmem) at the lowest contrast (Figure 4b).

A one-way ANOVA with a factor of contrast (80.5%, 2%, 1%, 0.03%) was performed separately for κ and pmem . A significant effect of contrast was observed for the guess rate parameter (pmem), $F(1,18) = 5.36, p = 0.02$, but not for the precision parameter (κ), $F(1,18) = 0.01, p = .095$. Bonferroni post hoc tests revealed that the guess rate for the lowest contrast value ($M = 0.30, SD = .06$) was significantly different from the other three, respectively.

These results show that, at least for the teardrop shapes used in the present study, reducing the contrast does not impact the precision of the representation of an object's orientation (a key aspect of its shape). Instead, reducing the contrast to very low levels led to an all-or-none failure, producing random guesses. These results are consistent with the hypothesis that shape processing is largely contrast-independent until the contrast gets so low that the shape is not consciously perceived at all.

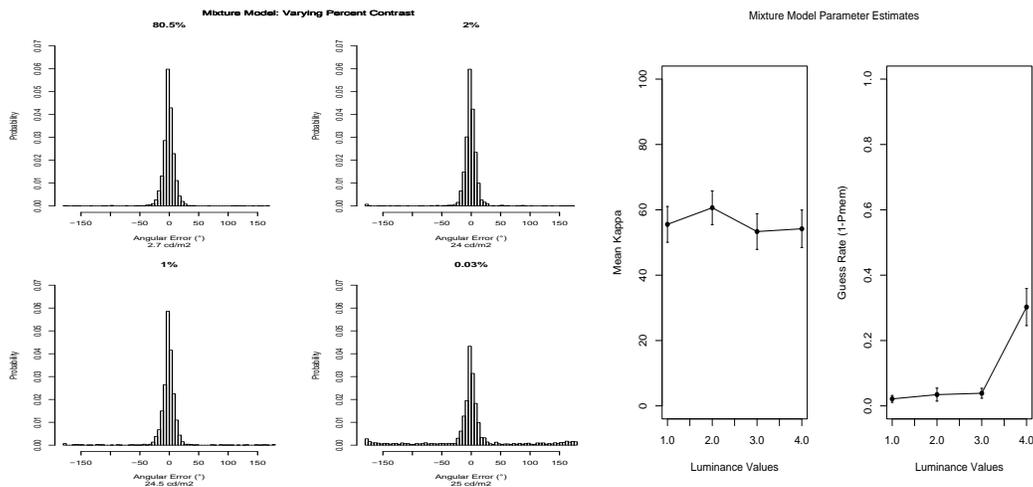


Figure 4a. Probability distribution of response errors collapsed across all participants at four different luminance contrast values. Teardrop-shaped stimuli were presented on a dark stimuli on a grey background. One stimulus was black, 80.49% contrast, while the other three stimuli were only slightly darker than the background, 2%, 1%, 0.03% contrast.

Figure 4b. Parameter estimates of a two component mixture model. Mean kappa (left) represents precision of the von-Mises distribution, while the guess rate, or 1- Pmem parameter, represents the all-or-none access to awareness (right).

Experiment 2: EEG

General Method

Participants

University of California, Davis students ($n = 22$, median age = 22, male = 9, female = 13) received monetary compensation for their participation. All participants reported normal or corrected-to-normal vision and no neurological disorders. The experimental protocol was approved by the University of California-Davis Institutional Review Board, and informed consent was obtained before each session.

Stimuli & Procedure

The task and stimuli in Experiment 2 were identical to those in Experiment 1 except for two changes designed to facilitate ERP analyses: 1) only two contrast values were use (80.5% = 2.7 cd/m² and 1% = 24 cd/m²), and 2) participants were probed to respond on a random 10% of

trials instead of reporting on every trial. Both of these changes were designed to allow more trials to be collected for each condition. In addition, the number of trials for each combination of contrast and orientation was increased (40 trials for each of the 2 x 16 combinations).

EEG Recording and Preprocessing

Continuous EEG was recorded from 32 electrodes using a Brain Products ActiCHamp system (Brain Products GmbH). Recordings were obtained from the following scalp electrode sites: FP1, FP2, F3, F4, F7, F8, C3, C4, P3, P4, P5, P6, P7, P8, P9, P10, PO3, PO4, PO7, PO8, O1, O2, Fz, Cz, Pz, POz, and Oz. Electrodes were also placed on the right and left mastoid to be used as references, and additional electrodes were placed adjacent to the right and left eye and beneath the right eye to record vertical and horizontal electro-oculogram (EOG) signals. These signals were filtered online with a cascaded integrator-comb antialiasing filter (half-power cutoff at 130 Hz) and digitized at 500 Hz. Impedances were $<10\text{ k}\Omega$ for the scalp electrodes and $<20\text{ k}\Omega$ for the EOG electrodes.

Offline analysis was performed using EEGLAB (Delorme & Makeig, 2004) and ERPLAB (Lopez-Calderon & Luck, 2014). All channels were re-referenced offline to the average of the mastoids. Data were band-pass filtered from 0.1 to 80 Hz and resampled at 250 Hz. EEG data during the intertrial intervals were removed. Independent component analysis (ICA) was used to correct for eye blinks and eye movements. Previous research has shown ICA eliminates ocular artifacts so that these artifacts do not influence the decoding analyses (Bae & Luck, 2018). To extract the ERPs, stimulus-locked epochs were extracted from -500 to 1500 ms relative to stimulus onset and baseline corrected from -500 ms to 0 ms. The epochs for each combination of contrast and orientation were then averaged together.

Analyses

Below is a brief description of analyses detailed in Bae and Luck, 2018. Additional information on the analysis technique and procedure can be found in this previous published manuscript.

Decoding Analysis

For decoding purposes, each epoch was low-pass filtered at 20 Hz and resampled at 100 Hz (one data point per 10 ms) prior to decoding. This resulted in a 4 dimensional matrix for each participant: timepoints (100) x orientations (16) x trials (~40) x electrode sites (27). Support vector machines (SVMs) and error-correcting output codes (ECOCs; Dietterich et al., 1995) were used to classify orientation based on the distribution of the ERP signal over the 27 scalp electrodes sites (collapsed across the current – trial teardrop orientation).

Data were decoded independently at 200 time points from -500 – 1480 ms. The decoding procedure used threefold crossvalidation, iterated 100 times. Therefore the data were divided into three separate training (2) and test (1) sets, resulting in three groups of 13 trials for each of the 16 orientations (3 groups x 16 orientations x 27 electrodes) at each time point. To improve signal-to-noise ratio, the 13 trials were averaged per group for each orientation, resulting in 3 ERP waveforms, and decoding analyses were implemented on these averages (not single trials; see Bae & Luck, 2018).

A separate classifier (16 SVMs) was trained to discriminate between each orientation and the other 15 orientations (one vs. all). When classifying orientation in the test set, decoders were combined to create a single decision about the most likely (predicted) orientation for a given (test) case. Decoding accuracy was calculated for the average of the set of trials for each orientation, and was only considered correct when the classifier could pick the correct orientation

out of the sixteen possibilities (chance is .0625 or 1/16). This procedure was repeated 3 times at each time point with each group of data serving as the test data. This operation was then iterated 100 times, shuffling the trials before each iteration. Decoding accuracy was then collapsed across orientations (16), crossvalidations (3), and iterations (100) to produce a percentage decoding accuracy at each time point. We separately decoded the orientation of the stimuli at each contrast level.

Statistical Analysis of Decoding Accuracy

To test whether the decoding accuracy was greater than chance (.0625), we calculated one-sample t -tests at each individual time point during the 0–1500 ms epoch. Below-chance decoding accuracy is meaningless in this context so a one-tailed test was used at each time point. To control for multiple comparisons, we used a false discovery rate (FDR) correction (Benjamini & Yekutieli, 2001). FDR correction allows us to identify with reasonable confidence as many significant points as possible while preserving a low false positive rate. Unlike familywise error rate correction procedures like Bonferroni, which controls the probability of at least one Type 1 error, FDR has less strict control over Type I errors (i.e., of points that are still significant after correction only 5% of them we would expect to be spurious).

RSA & Convolutional Neural Network Model Procedure

In an effort to investigate the correspondence between the layers of a deep convolutional neural network and the human visual processing stream, we linked EEG data to a convolutional neural network (CNN) model using representational similarity analysis. Since there is no strict isomorphism between brain data and computational models, representational similarity analysis (RSA) allows us to abstract away from activity patterns in a particular modality (e.g., fMRI, scalp electrophysiology) and to compare the geometrical structure of the representations (e.g.,

not the pattern of activation itself; Kriegeskorte et al., 2008). In the current study, we used RSA to compare how brain activity during high and low contrast condition related to different layers of a convolutional neural network.

First, we computed the representational similarity matrices (RSM) for the EEG data, which quantified the similarity between the scalp topography for the 16 different orientations. RSMS were computed for every subject for every timepoint by correlating the scalp topography for all 16 different orientations. This left us with a 16x16 (orientation) RSM at each of the 500 time points for each of the 22 subjects.

Second, we used Alexnet (Krizhevsky et al., 2012) as our neural network model of the visual system. This model was inspired by the human ventral visual pathway, wherein as information flows through the ventral stream, receptive fields integrate visual information over larger and larger receptive fields as response properties become more complex layer by layer. The model consists of 12 purely feedforward layers: 5 convolutional layers and 3 pooling layers, followed by 3 fully connected layers. In this study, we trained Alexnet on Imagenet to recognize objects (Deng et al., 2009) and then fed in all 16 different orientations into the CNN with both low and high contrast stimuli. For each orientation we then obtained the activation values from the rectified linear activation (ReLU) function for each unit of the CNN at each layer, so for each of the 12 layers we had different activations for every orientation for both low and high contrast, producing 12 16x16 RSMS for both low and high contrast.

Once the EEG and model data were in the same units (i.e., RSMS), we could determine the relationship between the two different representational spaces. Because the relationship between the two different measurement spaces may not be linear, we used a Spearman Rank correlation between each subject's EEG RSMS (at each timepoint) and each of the convolutional

layer RSMs, giving us 22 (subject) correlation values at each of the 12 layers for each of the 500 time points. In this way, we could assess the strength of the relationship between each representational space, wherein greater similarity in the representational geometry between the brain data and the model yields greater correlational strength. This enabled us to see which layers of the visual processing hierarchy in the model corresponded to the information represented in the brain data. For a more detailed version of the methods see Kiat and Luck, 2020 (Kiat et al., 2020).

Results

Experiment 2 Behavioral Data

Participants' mean time to report the orientation of the teardrop (via the mouse button) was 2200 ms (single trial range was 600-7500 ms after subtracting the fastest and slowest 2.5% of trials). The behavioral data for Experiment 2 were just like those in Experiment except 1) two contrast values were use (80.5%, 1%), and 2) participants were probed to respond on 10% of trials instead of reporting on every trial. Probability distributions of response errors collapsed across all participants as a function of contrast are summarized in Figure 5a. Figure 5b provides the parameter estimates from the mixture model.

A one-way ANOVA on each parameter revealed no significant differences between contrast level for either the guess parameter (p_{mem}), $F(1,18) = 0.49$, $p = 0.49$, or the precision parameter (κ), $F(1,18) = 0.43$, $p = .052$. Note that the low-contrast condition of this experiment was the same as the second-lowest contrast used in Experiment 1, so the behavioral pattern observed here replicates the findings of Experiment 1.

To reduce motor artifacts and to collect a sufficient number of trials for the purposes of EEG decoding analyses, we probed participants to report the orientation on only 10% of trials, leaving us with fewer trials for Experiment 2 than Experiment 1. Therefore these results are less robust than those in Experiment 1 (but still yielded the same pattern, with no difference in either precision or guess rate between the highest contrast and the second-lowest contrast used in Experiment 1).

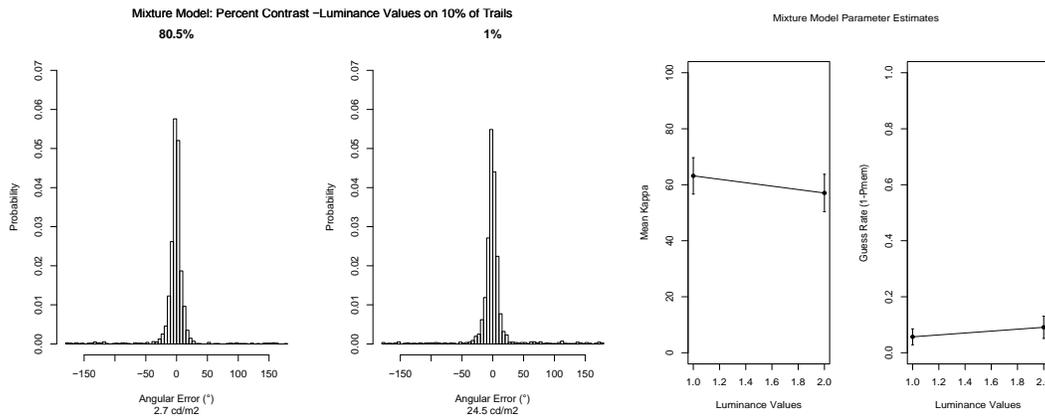


Figure 5a. Probability distribution of response errors on 10% of trials, collapsed across all participants at two different luminance contrast values. Teardrop-shaped stimuli were presented on a dark stimuli on a grey background. One stimulus was black, 80.5% contrast, while the other stimulus was 1% contrast. **Figure 5b.** Parameter estimates of a two component mixture model. Mean kappa (left) represents precision of the von-Mises distribution, while the guess rate, or 1- Pmem parameter, represents the all-or-none access to awareness (right).

Univariate EEG Analyses

Figure 6a shows grand average ERPs at the midline electrodes Fz, Cz, Pz, Oz for low- and high-contrast stimuli, averaged across orientations. We found that low contrast stimuli are more salient in frontal electrodes at early time points but this pattern reverses in more posterior electrodes where high contrast stimuli become more salient. ERPs for high and low contrast stimuli were significantly different between 200 – 450 ms at Fz, Cz, and Oz, $t(21) = 2.08$, $p < .05$. Although there was no significant difference for high and low contrast stimuli at Pz, $t(21) = 2.08$, $p = .42$.

Figure 6b plots the mean global field power (MGFP), a measure of total activity across all scalp electrode sites. Consistent with previous

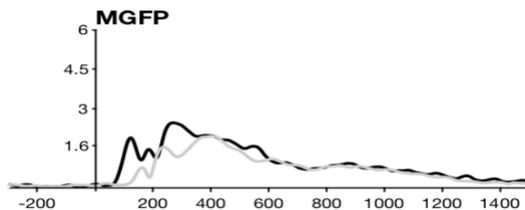


Figure 6b. Mean Global Field Power (MGFP); a measure of total activity across electrode sites.

studies, the early sensory activity was significantly different for high-contrast than for the low-contrast stimuli between 50 and 200 ms, $t(21) = 2.08$, $p < .05$. However, at later points (400 – 1000 ms), the MGFP was quite similar for low- and high-contrast stimuli, $t(21) = 2.08$, $p = .12$.

Experiment 2: ERP Decoding

Figure 7a and 7b show the decoding accuracy at each time point for the low- and high-contrast stimuli, respectively, with gray shading for time points at which the decoding was

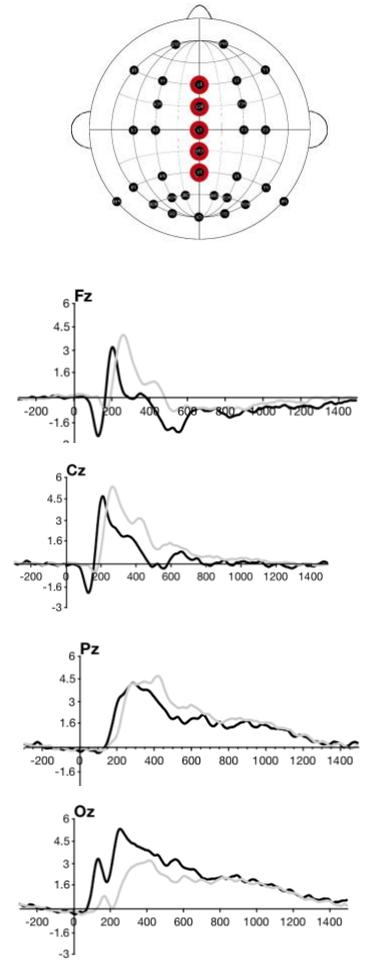
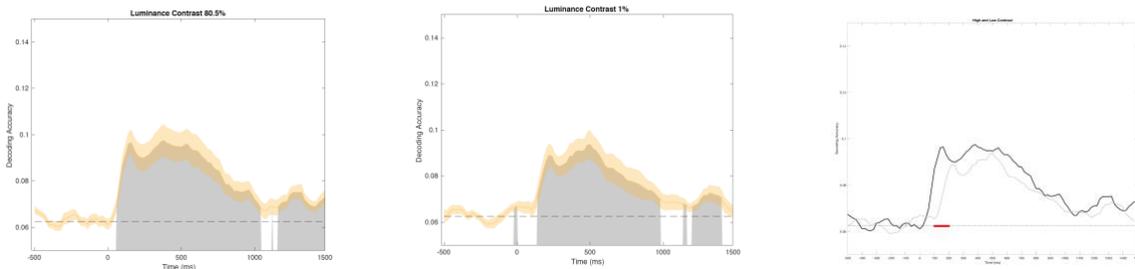


Figure 6a. Grand average ERPs at the midline scalp sites for low- and high-contrast stimuli, averaged across orientations.

significantly greater than chance (with an FDR correction for multiple comparisons). Figure 7c shows decoding accuracy overlaid for the low- and high-contrast conditions, and the red horizontal bar indicates time points in which decoding accuracy differed significantly between the two contrast levels (using 2-tailed t tests with an FDR correction).

Decoding accuracy was significantly above chance beginning at 50 ms for high-contrast stimuli and beginning at 150 ms for low-contrast stimuli. In addition, decoding accuracy was significantly greater for high-contrast stimuli than for low-contrast stimuli from 100 – 200 ms. After approximately 200 ms poststimulus, decoding accuracy was almost as high for low-contrast stimuli as for high-contrast stimuli and no significant differences between the two contrast levels was observed after this time point.



Figures 7a&b. Statistically significant (grey shaded) decoding accuracy for high and low contrast can be seen in Fig. 7a and 7b, respectively. **Figure 7c.** Statistically significant difference in decoding accuracy between high and low contrast (red line; 100-200 ms).

RSA/CNN Model

We did our RSA-CNN analyses on the first 5 convolutional layers of the network. Figure 8 shows the representational similarity values at each time point for the first and last convolutional layers. We found a significant difference between the high contrast and low contrast conditions in the early layers at relatively short latencies (145-250 ms), $t(21) = 2.08$, $p < .05$, paralleling our decoding effects. At higher layers, the peak in representational similarity declined in the early latency range; the difference between high and low contrast conditions was trending but not significant (Figure 8b). At later time points, in early and late layers, there was no

significant difference between high and low contrast conditions; this finding is consistent with the lack of significant difference in decoding accuracy at later time points.

We repeated the analysis using a different neural network, Cornet S, which includes not only feed-forward but recurrent connections and found a similar pattern when comparing the V1 layer and the IT layer.

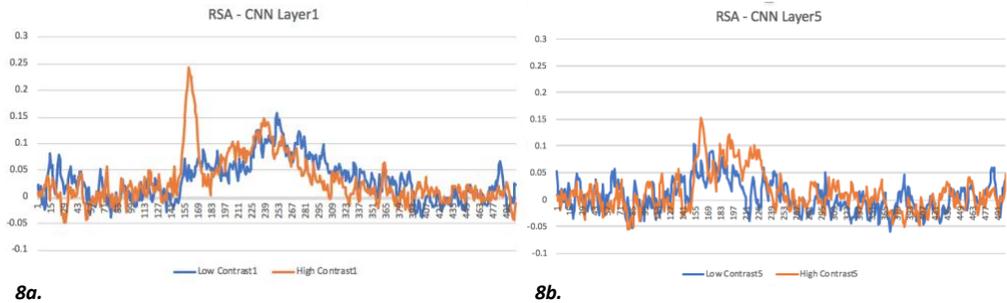


Figure 8a & 8b. RSA/CNN analyses revealed a significant difference between the high contrast and low contrast conditions in the early layers (**8a.**) at relatively short latencies (145-250 ms), while the peak in representational similarity declined progressively in higher layers (**8b.**).

Experiment 3: CNN Neurophysiology Analyses

The ability to recognize a shape over a broad range of contrasts is important in natural vision. As a result, one might expect that CNNs that are trained to classify objects in natural scenes would also exhibit contrast invariance in the deeper layers. To test this hypothesis, we asked whether units in the different layers of AlexNet exhibit the general pattern of contrast invariance that we observed in the behavioral results from Experiments 1 and 2. We took an approach that is analogous to the approach used in classic primate electrophysiology, in which we “presented” teardrops of different orientations and different contrast values to the network and

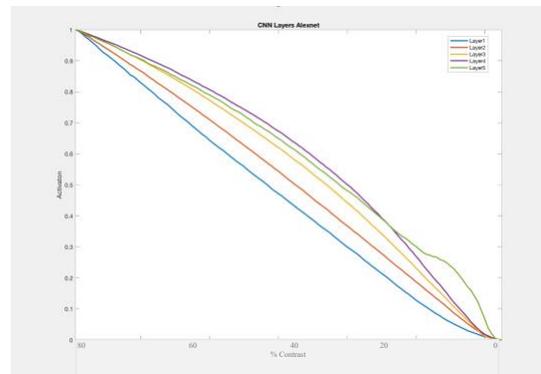


Figure 9. Taking the ten most responsive units from each layer (legend) and plotting the average activation level after the ReLU operation for each unit across the ten units for their preferred orientation as a function of contrast.

then plotted the magnitude of a given unit's response as a function of the orientation to get tuning curves.

In these analyses, we used a version of Alexnet that had been trained on Imagenet to recognize photographs of real-world objects. We fed in teardrop stimuli at 125 different contrast levels ranging from nearly zero percent contrast to 80% contrast, when compared to the grey background (25 cd/m²). Just as classic electrophysiologists often searched for responsive neurons and then obtained their tuning curves, we found the ten most responsive units from each layer and plotted the average activation level (after the ReLU operation) for each unit across the ten units for their preferred orientation as a function of contrast level (Figure 9). The data were normalized so that the smallest response for a given unit was coded as 0 and the largest response was coded as 1 ($[\text{value} - \text{minimum}] \div [\text{maximum} - \text{minimum}]$).

As shown in Figure 9, layer 1 exhibited a nearly linear decrease in activation as the contrast decreased. However, the contrast function became more bowed at higher layers, meaning that the response did not drop as steeply as the contrast dropped (until the very lowest contrast levels, where the activation dropped very rapidly). The function was highly nonlinear for layer 5, where the activation remained reasonably strong until the contrast neared zero. Therefore, we found reduced sensitivity to contrast in deeper layers of the network, confirming what we saw in the neural data and in the behavioral data.

Discussion

As we saw in Figure 2 when viewing the word CONTRAST at low and high contrasts, the perception of the shape of an object was largely independent of the contrast. As long as the contrast was high enough that the observers could see that a teardrop was present, the observers

were able to have a very precise representation even when measured in working memory. In other words, the stimulus was subjectively contrast invariant. Experiment 1 illustrated this invariance behaviorally, and the mixture model indicated that precision did not become impaired at low contrast, but instead participants sometimes exhibited random guesses for low-contrast stimuli as if they were completely unaware that a teardrop had been presented. In other words, subjects could report the orientation of a teardrop shape equally precisely whether it was presented at .03% or 80% contrast—the fact that there was absolutely no detectable impact on precision in behavior is a striking, nonobvious finding.

Experiment 2 examined this contrast invariance neurally, using decoding to test the hypothesis that the visual system begins with a contrast-dependent representation of orientation (lower orientation decoding accuracy for low-contrast relative to high-contrast stimuli) and then shifts to a contrast-independent representation of orientation (equal orientation decoding accuracy for low- and high-contrast stimuli). We found that during the early sensory period (~50-150 ms), decoding accuracy was greater for high contrast than low contrast stimuli, but beginning around 200 ms, decoding accuracy was nearly as great for low contrast stimuli as for high contrast stimuli. Thus, there is a period where the orientation information is contrast dependent and the orientation cannot be decoded from the EEG for low-contrast stimuli, and then the neural representation becomes largely contrast independent. This pattern is consistent with other studies indicating that contrast information is discarded at higher-level stages (Ohayon et al., 2012; Rolls & Baylis, 1986; Strasburger & Rentschler, 1996). We knew that from the single unit recordings but we have now demonstrated it in humans.

With the CNN analysis, we tried to flesh this out a bit more mechanistically. By linking EEG data to a CNN with RSA, we tested the hypothesis that activation in early layers of the

model would show representational similarity with the EEG data in the high contrast condition compared with low contrast condition, an effect that would reverse in deeper layers. Results of the model indicated a significant difference between the high and low contrast condition in the early layers, but the peak in representational similarity declined as you got to higher and higher layers; a finding that was trending but not significant.

One of the most interesting aspects about the behavioral data was the increase in guess rate at the lowest contrast with little impact on the precision. Follow up studies examining this lack of change in the precision by comparing the precision of the behavioral data to the decoding data to assess if behavioral precision is correlated with the decoding precision subject by subject, would be informative. By treating the decoder output like a behavioral response (i.e., what the decoder reported the orientation was on each iteration), one could plot the decoding error distributions for high and low contrast. A beautiful finding from the brain data would be that the precision is the same across contrasts, but there is a higher guess rate for the low contrast, as in the behavioral data.

An ongoing debate in neuroscientific consciousness research involves the functional localization of conscious perception. Higher Order (H. C. Lau, 2007; H. Lau & Rosenthal, 2011) and Global Workspace theories (Baars, B.J., 1997; Dehaene et al., 2011; Dehaene & Naccache, 2001) argue that conscious perception occurs in the prefrontal cortex, whereas Local Recurrency (Lamme, 2010) and Neural Synchrony theories (Crick & Koch, 1990) argue that conscious perception resides in posterior cortices. Therefore, an ongoing debate is whether V1 is necessary but not sufficient for visual consciousness, or does V1 contribute to visual consciousness directly or indirectly? In the context of examining subjective contrast invariance, one might ask, when subjects are reporting about orientation are they pulling the representation from V1 (analogous to

layer 1 of Alexnet) or inferotemporal cortex (analogous to layer 5)? If subjects must be pulling from one or another pool of representations, our neural and modeling data are consistent with the idea that inferotemporal cortex is the pool from which they are drawing. Together these findings suggest that early visual areas are not encoding the conscious perception of orientation, but awareness instead requires extended processing over time to extract a largely contrast independent perception of orientation.

References

- Alitto, H. J., & Usrey, W. M. (2004). Influence of Contrast on Orientation and Temporal Frequency Tuning in Ferret Primary Visual Cortex. *Journal of Neurophysiology*, *91*(6), 2797–2808. <https://doi.org/10.1152/jn.00943.2003>
- Anderson, J. S. (2000). The Contribution of Noise to Contrast Invariance of Orientation Tuning in Cat Visual Cortex. *Science*, *290*(5498), 1968–1972. <https://doi.org/10.1126/science.290.5498.1968>
- Aru, J., Bachmann, T., Singer, W., & Melloni, L. (2012). Distilling the neural correlates of consciousness. *Neuroscience & Biobehavioral Reviews*, *36*(2), 737–746. <https://doi.org/10.1016/j.neubiorev.2011.12.003>
- Avidan, G., Harel, M., Hendler, T., Ben-Bashat, D., Zohary, E., & Malach, R. (2002). Contrast Sensitivity in Human Visual Areas and Its Relationship to Object Recognition. *Journal of Neurophysiology*, *87*(6), 3102–3116. <https://doi.org/10.1152/jn.2002.87.6.3102>
- Baars, B.J. (1997). *Baars, B. J. (1997). Contrastive phenomenology: A thoroughly empirical approach to consciousness. The nature of consciousness: Philosophical controversies, 187-202.*
- Babiloni, C., Vecchio, F., Miriello, M., Romani, G. L., & Rossini, P. M. (2006). Visuo-spatial Consciousness and Parieto-occipital Areas: A High-resolution EEG Study. *Cerebral Cortex*, *16*(1), 37–46. <https://doi.org/10.1093/cercor/bhi082>
- Bae, G.-Y., & Luck, S. J. (2018). Dissociable Decoding of Spatial Attention and Working Memory from EEG Oscillations and Sustained Potentials. *The Journal of Neuroscience*, *38*(2), 409–422. <https://doi.org/10.1523/JNEUROSCI.2860-17.2017>

- Benjamini, Y., & Yekutieli, D. (2001). The Control of the False Discovery Rate in Multiple Testing under Dependency. *The Annals of Statistics*, 29(4), 1165–1188.
- Blake, R. (2022). The Perceptual Magic of Binocular Rivalry. *Current Directions in Psychological Science*, 31(2), 139–146. <https://doi.org/10.1177/09637214211057564>
- Block, N. (2011). Perceptual consciousness overflows cognitive access. *Trends in Cognitive Sciences*, 15(12), 567–575. <https://doi.org/10.1016/j.tics.2011.11.001>
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10(4), 433–436. <https://doi.org/10.1163/156856897X00357>
- Brouwer, G. J., & Heeger, D. J. (2009). Decoding and Reconstructing Color from Responses in Human Visual Cortex. *Journal of Neuroscience*, 29(44), 13992–14003. <https://doi.org/10.1523/JNEUROSCI.3577-09.2009>
- Buzsáki, G., Anastassiou, C. A., & Koch, C. (2012). The origin of extracellular fields and currents—EEG, ECoG, LFP and spikes. *Nature Reviews. Neuroscience*, 13(6), 407–420. <https://doi.org/10.1038/nrn3241>
- Cheng, K., Hasegawa, T., Saleem, K. S., & Tanaka, K. (1994). Comparison of neuronal selectivity for stimulus speed, length, and contrast in the prestriate visual cortical areas V4 and MT of the macaque monkey. *Journal of Neurophysiology*, 71(6), 2269–2280. <https://doi.org/10.1152/jn.1994.71.6.2269>
- Cichy, R. M., Khosla, A., Pantazis, D., Torralba, A., & Oliva, A. (2016). Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence. *Scientific Reports*, 6(1). <https://doi.org/10.1038/srep27755>

- Crick, F., & Koch, C. (1990). Towards a neurobiological theory of consciousness. *Seminars in The Neurosciences*, 13.
- Dehaene, S., Changeux, J.-P., & Naccache, L. (2011). The Global Neuronal Workspace Model of Conscious Access: From Neuronal Architectures to Clinical Applications. In S. Dehaene & Y. Christen (Eds.), *Characterizing Consciousness: From Cognition to the Clinic?* (pp. 55–84). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-18015-6_4
- Dehaene, S., & Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. *Cognition*, 79(1), 1–37.
[https://doi.org/10.1016/S0010-0277\(00\)00123-2](https://doi.org/10.1016/S0010-0277(00)00123-2)
- Del Cul, A., Baillet, S., & Dehaene, S. (2007). Brain Dynamics Underlying the Nonlinear Threshold for Access to Consciousness. *PLoS Biology*, 5(10).
<https://doi.org/10.1371/journal.pbio.0050260>
- Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1), 9–21. <https://doi.org/10.1016/j.jneumeth.2003.10.009>
- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 248–255. <https://doi.org/10.1109/CVPR.2009.5206848>
- Dietterich, T. G., Hild, H., & Bakiri, G. (1995). A comparison of ID3 and backpropagation for English text-to-speech mapping. *Machine Learning*, 18(1), 51–80.
<https://doi.org/10.1007/BF00993821>

- Gross, C., & Mishkin, M. (1977). *The Neural Basis of Stimulus Equivalence Across Retinal Translation*.
- Haxby, J. V. (2001). Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex. *Science*, 293(5539), 2425–2430.
<https://doi.org/10.1126/science.1063736>
- Haynes, J.-D., & Rees, G. (2005). Predicting the Stream of Consciousness from Activity in Human Visual Cortex. *Current Biology*, 15(14), 1301–1307.
<https://doi.org/10.1016/j.cub.2005.06.026>
- Kamitani, Y., & Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nature Neuroscience*, 8(5), 679–685. <https://doi.org/10.1038/nn1444>
- Kapoor, V., Dwarakanath, A., Safavi, S., Werner, J., Besserve, M., Panagiotaropoulos, T. I., & Logothetis, N. K. (2020). Decoding the contents of consciousness from prefrontal ensembles. *BioRxiv*, 2020.01.28.921841. <https://doi.org/10.1101/2020.01.28.921841>
- Kiat, J., Hayes, T., Henderson, J., & Luck, S. (2020). Assessing the time course of saliency and meaning: Representational similarity analysis of ERP responses to natural scenes. *Journal of Vision*, 20(11), 1629. <https://doi.org/10.1167/jov.20.11.1629>
- King, J.-R., Pescetelli, N., & Dehaene, S. (2016). Brain Mechanisms Underlying the Brief Maintenance of Seen and Unseen Sensory Information. *Neuron*, 92(5), 1122–1134.
<https://doi.org/10.1016/j.neuron.2016.10.051>
- Koivisto, M., Mäntylä, T., & Silvanto, J. (2010). The role of early visual cortex (V1/V2) in conscious and unconscious visual perception. *NeuroImage*, 51(2), 828–834.
<https://doi.org/10.1016/j.neuroimage.2010.02.042>

- Koivisto, M., & Revonsuo, A. (2003). An ERP study of change detection, change blindness, and visual awareness. *Psychophysiology*, *40*(3), 423–429. <https://doi.org/10.1111/1469-8986.00044>
- Koivisto, M., & Revonsuo, A. (2010). Event-related brain potential correlates of visual awareness. *Neuroscience & Biobehavioral Reviews*, *34*(6), 922–934. <https://doi.org/10.1016/j.neubiorev.2009.12.002>
- Kriegeskorte, N., Formisano, E., Sorger, B., & Goebel, R. (2007). Individual faces elicit distinct response patterns in human anterior temporal cortex. *Proceedings of the National Academy of Sciences*, *104*(51), 20600–20605. <https://doi.org/10.1073/pnas.0705654104>
- Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis—Connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*. <https://doi.org/10.3389/neuro.06.004.2008>
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. In F. Pereira, C. J. C. Burges, L. Bottou, & K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems 25* (pp. 1097–1105). Curran Associates, Inc. <http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf>
- Lamme, V. A. F. (2006). Towards a true neural stance on consciousness. *Trends in Cognitive Sciences*, *10*(11), 494–501. <https://doi.org/10.1016/j.tics.2006.09.001>
- Lamme, V. A. F. (2010). How neuroscience will change our view on consciousness. *Cognitive Neuroscience*, *1*(3), 204–220. <https://doi.org/10.1080/17588921003731586>

- Lamy, D., Salti, M., & Bar-Haim, Y. (2008). Neural Correlates of Subjective Awareness and Unconscious Processing: An ERP Study. *Journal of Cognitive Neuroscience*, *21*(7), 1435–1446. <https://doi.org/10.1162/jocn.2009.21064>
- Lau, H. C. (2007). A higher order Bayesian decision theory of consciousness. In R. Banerjee & B. K. Chakrabarti (Eds.), *Progress in Brain Research* (Vol. 168, pp. 35–48). Elsevier. [https://doi.org/10.1016/S0079-6123\(07\)68004-2](https://doi.org/10.1016/S0079-6123(07)68004-2)
- Lau, H., & Morales, J. (2022). The Neural Correlates of Consciousness. In *The Oxford Handbook of the Philosophy of Consciousness*. Oxford University Press.
- Lau, H., & Rosenthal, D. (2011). Empirical support for higher-order theories of conscious awareness. *Trends in Cognitive Sciences*, *15*(8), 365–373. <https://doi.org/10.1016/j.tics.2011.05.009>
- Lee, B. B., Martin, P. R., & Valberg, A. (1988). The physiological basis of heterochromatic flicker photometry demonstrated in the ganglion cells of the macaque retina. *The Journal of Physiology*, *404*, 323–347. <https://doi.org/10.1113/jphysiol.1988.sp017292>
- Leopold, D. A., & Logothetis, N. K. (1999). Multistable phenomena: Changing views in perception. *Trends in Cognitive Sciences*, *3*(7), 254–264. [https://doi.org/10.1016/S1364-6613\(99\)01332-7](https://doi.org/10.1016/S1364-6613(99)01332-7)
- Logothetis, N. K., & Schall, J. D. (1989). Neuronal Correlates of Subjective Visual Perception. *Science*, *245*(4919), 761–763. JSTOR.
- Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: An open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience*, *8*. <https://www.frontiersin.org/articles/10.3389/fnhum.2014.00213>

- Malach, R., Levy, I., & Hasson, U. (2002). The topography of high-order human object areas. *Trends in Cognitive Sciences*, 6(4), 176–184. [https://doi.org/10.1016/S1364-6613\(02\)01870-3](https://doi.org/10.1016/S1364-6613(02)01870-3)
- Mohsenzadeh, Y., Mullin, C., Lahner, B., & Oliva, A. (2020). Emergence of Visual Center-Periphery Spatial Organization in Deep Convolutional Neural Networks. *Scientific Reports*, 10(1). <https://doi.org/10.1038/s41598-020-61409-0>
- Monasse, P. (1999). Contrast invariant registration of images. *1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. ICASSP99 (Cat. No.99CH36258)*, 6, 3221–3224 vol.6. <https://doi.org/10.1109/ICASSP.1999.757527>
- Monasse, P., & Guichard, F. (2000). Fast computation of a contrast-invariant image representation. *IEEE Transactions on Image Processing*, 9(5), 860–872. <https://doi.org/10.1109/83.841532>
- Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: Multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences*, 10(9), 424–430. <https://doi.org/10.1016/j.tics.2006.07.005>
- Ohayon, S., Freiwald, W. A., & Tsao, D. Y. (2012). What Makes a Cell Face Selective? The Importance of Contrast. *Neuron*, 74(3), 567–581. <https://doi.org/10.1016/j.neuron.2012.03.024>
- Pelli, D. G. (1997). *The VideoToolbox software for visual psychophysics: Transforming numbers into movies*. Spatial Vision.
- Pitts, M. A., & Britz, J. (2011). Insights from Intermittent Binocular Rivalry and EEG. *Frontiers in Human Neuroscience*, 5. <https://doi.org/10.3389/fnhum.2011.00107>

- Pitts, M. A., Martinez, A., & Hillyard, S. A. (2010). When and where is binocular rivalry resolved in the visual cortex? *Journal of Vision*, *10*(14), 25–25.
<https://doi.org/10.1167/10.14.25>
- Pitts, M. A., Padwal, J., Fennelly, D., Martínez, A., & Hillyard, S. A. (2014). Gamma band activity and the P3 reflect post-perceptual processes, not visual awareness. *NeuroImage*, *101*, 337–350. <https://doi.org/10.1016/j.neuroimage.2014.07.024>
- Reynolds, J. H., & Chelazzi, L. (2004). Attentional Modulation of Visual Processing. *Annual Review of Neuroscience*, *27*(1), 611–647.
<https://doi.org/10.1146/annurev.neuro.26.041002.131039>
- Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, *2*(11), 1019–1025. <https://doi.org/10.1038/14819>
- Rolls, E. T., & Baylis, G. C. (1986). Size and contrast have only small effects on the responses to faces of neurons in the cortex of the superior temporal sulcus of the monkey. *Experimental Brain Research*, *65*(1), 38–48. <https://doi.org/10.1007/BF00243828>
- Salti, M., Monto, S., Charles, L., King, J.-R., Parkkonen, L., & Dehaene, S. (2015). Distinct cortical codes and temporal dynamics for conscious and unconscious percepts. *ELife*, *4*.
<https://doi.org/10.7554/eLife.05652>
- Sandberg, K., Andersen, L. M., & Overgaard, M. (2014). Using multivariate decoding to go beyond contrastive analyses in consciousness research. *Frontiers in Psychology*, *5*.
<https://doi.org/10.3389/fpsyg.2014.01250>
- Sergent, C., Baillet, S., & Dehaene, S. (2005). Timing of the brain events underlying access to consciousness during the attentional blink. *Nature Neuroscience*, *8*(10), 1391–1400.
<https://doi.org/10.1038/nn1549>

- Seymour, K., Clifford, C. W. G., Logothetis, N. K., & Bartels, A. (2010). Coding and Binding of Color and Form in Visual Cortex. *Cerebral Cortex*, 20(8), 1946–1954.
<https://doi.org/10.1093/cercor/bhp265>
- Skottun, B. C., Bradley, A., Sclar, G., Ohzawa, I., & Freeman, R. D. (1987). The effects of contrast on visual orientation and spatial frequency discrimination: A comparison of single cells and behavior. *Journal of Neurophysiology*, 57(3), 773–786.
<https://doi.org/10.1152/jn.1987.57.3.773>
- Strasburger, H., & Rentschler, I. (1996). Contrast-dependent Dissociation of Visual Recognition and Detection Fields. *European Journal of Neuroscience*, 8(8), 1787–1791.
<https://doi.org/10.1111/j.1460-9568.1996.tb01322.x>
- Supèr, H., Spekreijse, H., & Lamme, V. A. F. (2001). A Neural Correlate of Working Memory in the Monkey Primary Visual Cortex. *Science*, 293(5527), 120–124.
<https://doi.org/10.1126/science.1060496>
- Tononi, G., & Koch, C. (2008). The Neural Correlates of Consciousness. *Annals of the New York Academy of Sciences*, 1124(1), 239–261. <https://doi.org/10.1196/annals.1440.004>
- Weil, R. S., & Rees, G. (2010). Decoding the neural correlates of consciousness: *Current Opinion in Neurology*, 23(6), 649–655. <https://doi.org/10.1097/WCO.0b013e32834028c7>
- Xia, G.-S., Delon, J., & Gousseau, Y. (2010). Shape-based Invariant Texture Indexing. *International Journal of Computer Vision*, 88(3), 382–403.
<https://doi.org/10.1007/s11263-009-0312-3>
- Zou, J., He, S., & Zhang, P. (2016). Binocular rivalry from invisible patterns. *Proceedings of the National Academy of Sciences of the United States of America*, 113(30), 8408–8413.
<https://doi.org/10.1073/pnas.1604816113>

III.

Chapter II: ERP Decoding of Visual Consciousness During Binocular Rivalry

Abstract

EEG decoding techniques provide a powerful tool for examining the temporal dynamics of representational content in the brain, which can help adjudicate between theoretical models of consciousness that delineate when awareness occurs in the visual processing hierarchy (Dehaene & Changeux, 2011; Lamme, 2006). In an effort to isolate the neural correlates of consciousness (NCC) and further understand the time course of conscious perception, the current study used EEG decoding methods to decode conscious and unconscious orientation information during binocular rivalry. We compared decoding accuracy from a) conscious content reported in the perceptually dominant eye with b) unconscious content in the non-dominant eye, suppressed during rivalry. Because the retinal stimulus is the same for different conscious percepts, binocular rivalry is a canonical paradigm for investigating the NCCs. We tested two competing predictions: (1) the Global Neuronal Workspace Theory (GNWT) predicts that the difference between conscious and unconscious processing would occur at least 300 ms after stimulus onset (i.e., the time range of the P3b component); (2) the view that an earlier electrophysiological correlate of visual awareness (the Visual Awareness Negativity) reflects conscious perception predicts that the difference between conscious and unconscious processing occurs earlier than the P3b, at ~200 ms. We found that decoding accuracy for the conscious and unconscious percepts on the same trial both exceeded chance at ~140 ms poststimulus and remained approximately equivalent until ~560 ms; the unconscious percept then returned to chance while the decoding accuracy for the conscious percept

persisted until 1000 ms. These findings are consistent with theories such as the GNWT that propose a late electrophysiological correlate of visual awareness.

Introduction

One major debate in the science of consciousness involves the role of the prefrontal cortex in facilitating conscious perception. Higher Order (H. C. Lau, 2007; H. Lau & Rosenthal, 2011) and Global Workspace theories (Baars, B.J., 1997; Dehaene et al., 2011; Dehaene & Naccache, 2001) postulate that conscious perception is circumscribed to the prefrontal cortex (PFC), whereas Local Recurrency (Lamme, 2010) and Neural Synchrony theories (Crick & Koch, 1990) argue that the PFC is not essential for perceptual awareness (H. Lau & Morales, 2022).

According to Global Neuronal Workspace Theory (GNWT), conscious access constitutes the “selection, amplification, and global broadcasting” of information (Dehaene et al., 2011). On this account, early bottom-up processes from ~100-300 ms are necessary but not sufficient for awareness. To become conscious, the representations created by these early processes must be amplified through top-down processes and maintained in a global workspace (Dehaene et al., 2011). This amplification and broadcasting of neuronal information, which begins sometime after ~300 ms, occurs in conjunction with increased gamma oscillation and phase synchrony across the brain and is referred to as conscious “ignition”. On this model, the P3b, a late electrophysiological signature occurring at ~300-500 ms, consistently correlates with conscious access (Babiloni et al., 2006; Del Cul et al., 2007; Koivisto & Revonsuo, 2010; Lamy et al., 2008; Sergent et al., 2005).

Unlike Global Workspace Theories, which require global activity in frontoparietal networks, Higher Order Theories argue that specific computations in PFC are required for

conscious perception. Consciousness on this account is not for producing superior performance, but instead is based on Bayesian decision theory (H. C. Lau, 2007). On this account, since perceptual decision making capacity is limited by the amount of noise in the system, the best performance is set by decision criteria derived from previous behavior, implying that the system makes many statistical inferences or learns from internal signals over time (e.g., “higher order” learning). Despite these differences from GNWT, Higher Order Theories agree that consciousness does not occur until a relatively late stage of processing.

By contrast, several opposing theories and findings argue that posterior cortices house the critical loci for conscious perception. *Recurrent Processing Theory* claims that a feedforward sweep of processing is not sufficient, and that consciousness requires *recurrent* processing (Lamme, 2006). Evidence for recurrent activity can be found in monkey studies where the initial V1 responses were identical whether monkeys reported seeing a figure or not (identifying when a salient figure on a background), but a late V1 response component (thought to represent recurrent activity) only occurred when the monkey reported seeing the figure (Supèr et al., 2001; Tononi & Koch, 2008). These findings are consistent with the claim that recurrent loops are necessary for conscious access.

Crick and Koch put forth the Neural Synchrony Theory in a 1990 paper, wherein the authors coined the popular working definition of the neural correlates of consciousness as “the minimal set of neuronal events and mechanisms jointly sufficient for a specific conscious percept”. On this account, attention causes neural firing rates to become synchronized between 40-70 Hz. They argued that this oscillatory mechanism underlies all conscious perception (i.e., neural oscillatory activity at 40-70 Hz constitutes a NCC), and the NCC resides in the cortical areas where this neural synchrony takes place. Therefore, various visual features like orientation

and motion are processed in early visual areas and are bound together by synchronous neural activity in these areas. Even though this activity is coordinated with feedback from higher level cortices, the PFC does not play a constitutive role in maintaining conscious perception.

Naturally, the cortical localization of perceptual consciousness bears upon the timing of neural activation. Evidence related to this latter set of models, which claim that awareness arises in early visual cortex, can be found by Koivisto and Revonsuo (2010). They found an early electrophysiological correlate of visual consciousness referred to as the *visual awareness negativity* (VAN). Incidentally, the authors proposed three NCC candidates, P1 enhancement at ~100 ms, an early posterior temporal and occipital negativity at ~200 ms, (i.e., the *visual awareness negativity*), and the P3b component at ~300-500 ms. Results showed a larger VAN when comparing stimuli that reached awareness to stimuli that did not, across studies and conditions (Koivisto et al., 2010; Koivisto & Revonsuo, 2003, 2010). The VAN was found to be independent of attention and often, but not always, followed by a the P3b.

This chapter focuses on the temporal commitments of these two general theoretical camps, which make different predictions about the functional localization and therefore the time course of conscious perception. Specifically, we concentrate on the GNWT and the VAN because these theories and findings make testable predictions regarding the timing of visual awareness. We employ EEG decoding methods during binocular rivalry to test two competing predictions: 1) the Global Neuronal Workspace Theory (GNWT) predicts that the difference between conscious and unconscious processing will occur at least 300 ms after stimulus onset (i.e., P3b time range), 2) the view that the VAN reflects conscious predicts that the difference between conscious and unconscious processing occurs at ~200 ms.

Binocular rivalry is an elegant paradigm for assessing the NCC. In this paradigm, two distinct images are presented separately at spatially corresponding locations, one to each eye. During rivalry, the observer is typically aware of only one of the two images, and conscious perception appears to “switch”, or “alternate”, or to spontaneous “reverse” between the images over time. Binocular rivalry is uniquely suited to isolate the NCCs since the retinal stimulus remains the same while only the conscious percept changes (Leopold & Logothetis, 1999). In other words, because a given eye will sometimes contain the image corresponding to awareness and will sometimes contain the unconscious image, this paradigm makes it possible to compare conscious and unconscious mental representations of exactly the same stimulus, thus isolating awareness. In the words of Randolph Blake, “what you are seeing...is not what you are viewing” (Blake, 2022). Because binocular rivalry occurs frequently in natural perception (e.g., when part of an image is occluded by an object in one eye but not in the other eye), it is also a more natural way to track conscious visibility, and information about the perceived and unperceived items are kept separate at least until area V1. Moreover, because people typically either fully perceive the image in one eye or fully perceive the image in the other eye, rivalry does not require the complex metacognitive processes and thresholds involved in making a visible/invisible decision for a masked stimulus.

Previous research using binocular rivalry has yielded inconsistent results regarding the time course of conscious perception. Single unit recordings of neural activity in monkey visual cortex provided evidence of a relatively late effect. Specifically, only a small percentage of neurons in early cortical areas (V1 and V2) responded in concert with rivalrous alternations (Leopold & Logothetis, 1989). Neuronal firing was higher for rivalrous alternations in extrastriate areas V4, MT and MST than in V1 and V2, while nearly all visual activity in IT and

STS corresponded to the rivalrous percept (Leopold & Logothetis, 1999; Logothetis & Schall, 1989). These findings suggest a relatively late onset of the perceptually dominant percept in the visual processing hierarchy. By contrast, a study using univariate EEG analyses found evidence of a relatively early effect. Specifically, differences in ERP activity between trials on which participants reported they were aware of the stimulus emerged at ~130-160 ms (Pitts, et al, 2010).

It is important to note that univariate analyses are limited for multiple reasons. Since conscious experience can vary trial by trial, univariate tests are not as sensitive to trial by trial differences. For example, there may be some trials where a prerequisite of conscious experience occurs (e.g., arousal, preparatory motor activity) but a conscious experience of the stimuli does not. Univariate tests therefore are less predictive of conscious experience because they can only show an average difference between conscious and unconscious conditions (Sandberg et al., 2014).

More recently a mathematical technique using powerful pattern classification methods called *multivariate decoding* has been used to decode information processing in the brain (Norman et al., 2006). Importantly, this technique can reveal detailed, content-specific aspects of information processing like how the brain processes specific orientation information (Bae & Luck, 2018; Kamitani & Tong, 2005). A typical decoding protocol involves splitting the data into *training data* and *test data*. Researchers then train a classifier (e.g., a support vector machine) on the training data to determine the patterns of brain activity that best distinguish between the to-be-decoded stimulus classes. The training yields an optimal plane in the multidimensional data space for separating the data associated with the two different stimulus classes. After the classifier is trained on the training data, it is tested on the test data to see if the classifier can

predict which class a given test case belongs to. If the classification accuracy is better than chance, this indicates that the brain data contains enough information to satisfactorily differentiate the stimulus classes (Weil & Rees, 2010).

Until recently, decoding has primarily been applied to fMRI data. Unlike conventional univariate fMRI approaches, which track the magnitude of the BOLD signal at a specific voxel or averaged across a region of interest, multivariate pattern analysis (MVPA; of which decoding is a subset) assesses patterns across many voxels. Previous studies using fMRI have decoded color and form, object categories (face/house), and the identity of individual faces (Brouwer & Heeger, 2009; Haxby, 2001; Kriegeskorte et al., 2007; Seymour et al., 2010). These fMRI decoding methods have also been used to decode the neural correlates of consciousness, including a study by Haynes and Rees (2005) which attempted to decode changes in conscious experience during binocular rivalry (Haynes & Rees, 2005).

However, because the BOLD signal integrates neural activity over a period of several seconds, it is difficult to know the precise time when a given neural response has occurred. This makes it difficult to distinguish between models of consciousness that vary largely in whether awareness occurs very early in processing (as in the visual awareness negativity) or relative late in processing (as in the GNWT). More recently, researchers have applied decoding methods to EEG data because of its superior temporal resolution, on the order of milliseconds. In addition, EEG has the advantage of directly reflecting the summed postsynaptic potentials of the active neurons in the brain, whereas the BOLD signal is a downstream, secondary consequence of neural activity. EEG decoding therefore provides a powerful and fairly novel tool for investigating the temporal dynamics of representational content in the brain, especially as it pertains to the content of conscious experience.

Only recently have EEG and MEG experiments used decoding techniques to examine potential NCCs. In one experiment, for example, participants were instructed to report the spatial location (one of eight) and visibility (manipulated by a backward mask) of a target stimulus during a “blindsight” task (Salti et al., 2015). MVPA showed that *some* information about spatial location was found on the blindsight trials, but only at ~270 ms post stimulus onset did information unique to consciously perceived stimuli occur. Another study examining unconscious working memory found that orientation could be decoded independently of visibility reports, and that this information was maintained over a retention period even when reported as unseen (King et al., 2016).

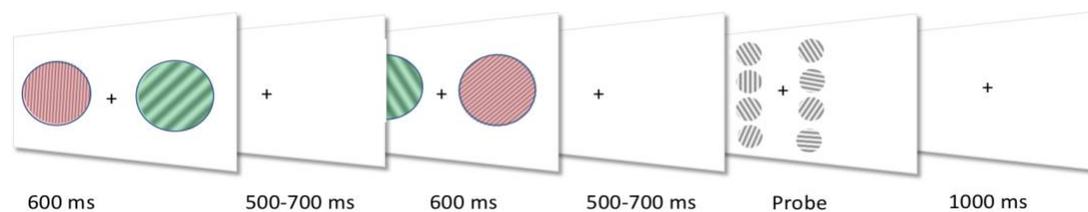
It is typical to manipulate visibility via backward masking to compare conscious and unconscious content. However, backward masking has its limitations. Since backward masking disrupts the representation of information beginning in the retina, reducing the quality or strength of the stimulus, this makes informational content difficult to decode. In this way, binocular rivalry is a superior paradigm to test and compare the timing of conscious and unconscious content without disrupting early representational content (i.e., in V1). Additionally, previous studies relied on participants to report whether a given stimulus was visible or not, which can be unreliable in masking paradigms because research shows that participants can perform better than chance even when they report being unaware of the stimulus.

Another problem in identifying the NCCs results from limitations of the *contrastive* approach that was used in many of the previous studies. The contrastive approach involves comparing conditions in which the stimuli are identical but participants either have or do not have a conscious experience of the stimuli, and then examining the differences in brain activity (Baars, B.J., 1997). One difficulty with the contrastive approach lies in conflating the true NCC

with the prerequisites and consequences of conscious perception; that is, a necessary precondition for awareness will differ between aware and unaware trials, as will a downstream consequence of awareness.

These processes are sometimes distinguished as three distinct NCCs: 1) *NCC-pr*, prerequisite processes carrying no conscious content; 2) *NCC proper*, the true NCC; and 3) *NCC-co*, the consequences of conscious experience (Aru et al., 2012; Pitts et al., 2014). For example, preparatory processes included in the *NCC-pr* may be necessary but not sufficient for the *NCC proper*. Therefore, differentiating these three NCC processes is particularly important when adjudicating between theories of consciousness which make claims about the timing of conscious perception. This problem is notoriously difficult with self-report, because it necessarily reflects both perceptual and post perceptual processing. However, decoding methods like we employ in this study can now be applied to the contrastive approach. That is, comparing content-specific aspects of conscious vs. unconscious contents may help resolve the problem of conflating pre and post NCCs with the *NCC proper*. For example, some of the prerequisites and consequences of consciousness that would vary between aware and unaware trials (e.g., arousal) would not be expected to contain decodable information about the *contents* of awareness. Consequently, if the contents of visual awareness can be decoded from a presumptive NCC, this increases the likelihood that it is a *NCC proper* rather than being a prerequisite or consequence of awareness.

Binocular Rivalry Procedure (with probe)



Physical Alternation Procedure

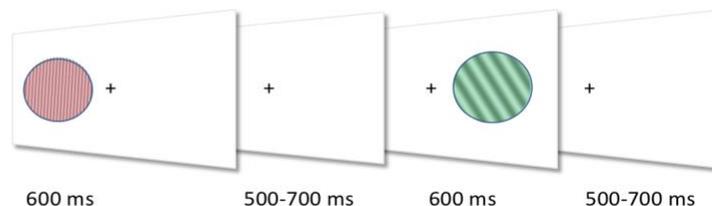


Figure 1: The Binocular Rivalry Procedure (top left), Physical Alternation Procedure (bottom right). Participants viewed the stimuli through a mirror stereoscope. Binocular pairs (left) were presented for 600 ms followed by a variable delay between 500-700 ms of a blank screen. Monocular pairs (right) were presented for 600 ms followed by a variable delay between 500-700 ms. High and low spatial frequencies were always associated with red and green colors, respectively. 8 possible orientations were presented. Participants reported the color they perceived (red/green) by pressing one of two keys. In the probe condition, on a random 10 % of trials participants were asked to report the orientation of the stimulus on the previous trial.

In the Binocular Rivalry condition of the present study (top left Figure 1), a red oriented grating was presented to one eye and a green oriented grating was simultaneously presented to the other eye. Participants viewed the stimuli through a mirror stereoscope that aimed the two gratings so that they appeared at the same foveal location in both eyes. That is, the sensory input contained a red grating of one orientation and a green grating of another orientation that appeared as if they were at the same location in space. In the absence of active mechanisms for dealing with difference between the two eyes, these two gratings would merge together into a single complex grating. However, the visual system contains mechanisms that prevent this kind of merger and instead involve a winner-take-all outcome in which either the red grating or the green grating is consciously perceived on a given trial.

The present study also included a Physical Alternation condition, in which a grating appeared in only one of the two eyes on each trial (bottom right Figure 1). Because there was no significant competition from the other eye, this grating was always consciously perceived. For

both the Binocular Rivalry and Physical Alternation condition, the stimulus was presented for 600 ms, followed by a variable delay between 500-700 ms. Participants pressed one of two keys on each trial to indicate whether they perceived a red grating or a green grating on that trial. They were instructed to press neither key if they perceived a mixture of the red and green gratings (which occurred rarely). On a random 10 % of trials participants, the grating was followed by a cue that indicated they should report the orientation of the grating that they had just seen.

In the Binocular Rivalry condition, each trial contained a red grating presented to one eye and a green grating presented simultaneously to the other eye, but the observer was consciously aware of only one of these two gratings. Because of their keypress responses, we knew which of the two gratings was consciously perceived on each trial. This made it possible for us to take the EEG data—which contained a mixture of brain activity corresponding to the consciously perceived eye and the orientation presented to the unperceived eye—and try to decode the orientation of the perceived grating and the orientation of the unperceived grating. If Global Workspace Theory is correct, one would predict that rivalry is not resolved until ~300-500 ms. That is, this is the time range in which the stimulus in one eye becomes consciously available. Thus, GWT would predict that decoding accuracy for the consciously reported orientation would increase from ~300-500 ms relative to nonperceived orientation presented to the other eye, and decoding accuracy from the nonperceived eye would then quickly return to chance. By contrast, if the Visual Awareness Negativity view is correct, the stimulus in one eye would become available to consciousness at ~200 ms. That is, this model predicts that the difference in decoding accuracy between the perceived and nonperceived orientations would occur at ~200 ms.

The physical alternation (non-rivalrous condition) serves as a baseline, and the two theories do not make differing predictions. We predict that the decoding accuracy for the raw stimulus would initially exceed chance at ~50 ms after stimulus onset (because this is approximately when visual signals reach visual cortex) and should be maintained for at least the 500 ms stimulus duration, reflecting a combination of unconscious and conscious processing.

Traditionally, stimuli in binocular rivalry paradigms are presented for a long duration (e.g., 30 seconds), and observers report spontaneous perceptual ‘reversals’ over time. However, spontaneous reversals are difficult to time-lock, making EEG decoding impractical. That is, if we don’t know exactly when the reversal occurred, it is difficult to determine the exact time at which a given neural signal occurred relative to the reversal. This is why we used the *intermittent stimulus presentation* approach shown in Figure 1 rather than the traditional approach. The intermittent stimulus presentation approach makes it possible to simply use stimulus onset as the time-locking point. Previous studies have shown that with stimulus parameters such as those used in the present study, only one percept is experienced per trial, and transition rates across trials are similar to reversal rates in continuous presentation paradigms (Pitts et al., 2010; Pitts & Britz, 2011). In many designs, participants make differential responses. For example, during an oddball paradigm participants are often instructed to report only the rare stimuli. However, in this design the response related activity will be identical for the aware and unaware orientations because participants are not reporting the orientation on every trial, but instead whether they perceived the color red or a green. In this way, the response related activity is nicely separated from the orientation information and decodability. This unique experimental design—which takes advantage of the intermittent stimulus approach together with the temporal precision of

EEG decoding— will, for the first time, be able to determine the exact time at which the content of the neural representation differs for perceived and non-perceived stimuli.

Methods

Participants

EEG recordings were obtained from 29 UC-Davis students between the ages of 19–28 with normal or corrected-to-normal visual acuity, normal color vision, and normal depth perception, as determined by self-report. An a priori power analysis was conducted using G*Power version 3.1.9.7 for sample size estimation. With a significance criterion of $\alpha = .05$ and power = .90 for a medium effect (Bae & Luck, 2018), the minimum sample size needed with this effect size is $n = 21$. Our $n = 19$ was just shy of the suggested a priori $n = 21$. Prior to the EEG recording, subjects were familiarized with the stereoscope to ensure they could achieve binocular fusion. Six subjects were removed due to the inability to perceive fusion or rivalry. Four were removed due to technical issues, leaving 19 subjects included in the analyses. Participants gave informed consent, and the study was approved by the UC-Davis Institutional Review Board.

Stimuli and Procedure

Stimuli were presented at a viewing distance of 100 cm on a grey background (25 cd/m²), and centered horizontally and vertically on a 60 Hz LCD monitor (Dell U2412M). Participants viewed stimuli through a mirror stereoscope (Stereo Aid; Figure 2) which allows images from different parts of the monitor to stimulate corresponding portions of the left and right eyes. At the beginning of each session, the experimenter adjusted the mirror angles to achieve stereo-fusion in each participant. To maintain fusion, participants were instructed to fixate a centrally located fixation point.

As illustrated in Figure 1, there were two conditions: a) the binocular rivalry condition, and b) the physical alternation condition. In the rivalry condition, stimuli consisting of circular square wave gratings 6 degrees in diameter were presented as pairs, one to each eye. Square wave gratings were used because V1 neurons are tuned for a combination of spatial frequency and orientation, and sine wave gratings would lead to responses only in neurons are tuned for that combination of spatial frequency and orientation. Square wave gratings, by contrast, will lead to responses from all neurons that are tuned to a given orientation, irrespective of their tuning for spatial frequency.

The gratings differed along four orthogonal dimensions, orientation (one of 8 different orientations; every 22.5 degrees starting at 0 degrees from upright), color (red: CIE = (2.1652, 1.1164, 0.1015) or green: CIE = (0.8138, 1.6277, 0.2713), spatial frequency (high (5.85 cpd) or low (4.55 cpd)), and eye (right or left). Eight different orientations were randomly presented with equal probability. The presentation of the orientation to each eye was independent of each other (i.e., equal numbers of all possible combinations of the eight orientations). Because we used circular gratings, there were only 180° of distinct orientations. On each trial, one grating was red and the other was green, but which eye was red and which eye was green varied at random across trials. Orientation gratings were either high spatial frequency or low spatial frequency, with red always being high spatial frequency and green always being low spatial frequency. Color and orientation were perfectly correlated because if the randomly-chosen gratings presented to the two eyes happen to have the same orientation on a given trial, the gratings will not perfectly overlap because of the difference in spatial frequency. All stimuli were equated in luminance. There were 64 possible combinations of red and green gratings (randomly intermixed) presented for 16 blocks (total trials).

In the rivalry condition (Figure 1 upper left), binocular image pairs of were presented for 600 ms followed by a variable delay between 500-700 ms of a blank screen. Because the stimuli were rivalrous, participants perceived either the red grating or the green grating on a given trial. Participants were instructed to report the color they perceive (red/green) on each trial, by pressing one of two keys. Participants were instructed to withhold their response if they perceived an integration of red and green or if they perceived a reversal of the colored grating (transition from red to green or vice versa) mid-trial.

In the physical alternation condition (non-rivalrous; Figure 1 lower right), one stimulus was presented to one eye while the other eye was blank. This condition served as a baseline or raw stimulus response. In this condition, participants were presented with one colored grating, so no binocular fusion was possible. The stimulus was presented randomly to either the right eye or the left eye on each trial. Perceptually, participants experienced this condition that same as the rivalry condition, but behind the scenes there was no binocular competition. The presentation rate mimicked the same presentation rate as rivalry.

On a random 10% of the trials, participants were probed to report the orientation of the perceived grating. These probe trials were inserted to ensure participants remained attentive to the orientation feature to maximize the encoding of orientation information. The probe display consisted of eight (black and white) orientation gratings; participants were instructed to use the gamepad to select the orientation that corresponded to the orientation they perceived on the preceding trial. The location of a given orientation within the probe array varied randomly from trial to trial. This made it impossible for the participant to prepare a motor response prior to the onset of the probe array. Consequently, we could be certain that we were decoding the working memory representation of the orientation rather than decoding motor preparation. There was a

1000 ms intertrial delay after the probe responses to avoid contaminating the following trial with motor related activity.

A flicker fusion photometry test was given prior to the onset of the main task to equate the luminance of the red and green stimuli for each individual subject. Flicker fusion photometry is based on the principle that the perceived flicker between the brightness of two colors is minimal when the brightness of the two colors are equal.

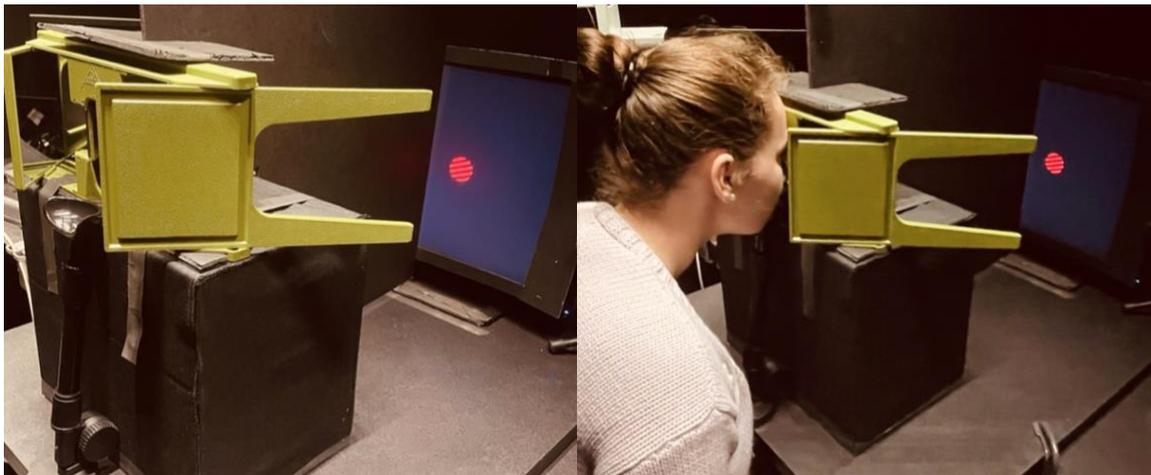


Figure 2: Participants viewed stimuli through a mirror stereoscope (above) which allows images from different parts of the monitor to stimulate overlapping portions of the left and right eyes. At the beginning of each session, the experimenter adjusted the mirror angles to achieve stereo-fusion in each participant. To maintain fusion, participants were instructed to fixate a centrally located fixation point.

EEG collection and analysis

Continuous EEG was recorded from 64 electrodes using a Brain Products ActiCHamp system (Brain Products GmbH). Recordings were obtained from the following 59 scalp sites: AF3, AF4, AF7, AF8, FC1, FC2, FC3, FC4, FC5, FC6, FP1, FP2, F1, F2, F3, F4, F5, F6, F7, F8, C1, C2, C3, C4, C5, C6, CP1, CP2, CP3, CP4, CP5, CP6, P1, P2, P3, P4, P5, P6, P7, P8, P9, P10, PO3, PO4, PO7, PO8, T7, T8, TP7, TP8, O1, O2, Fz, FCz, Cz, CPz, Pz, POz, and Oz, the left and right mastoids, and three electrooculogram (EOG) sites. Two horizontal EOG electrodes were placed adjacent to the outer canthi and were used to record horizontal eye movements, and

a vertical EOG electrode was placed beneath the right eye and used to record eyeblinks and vertical eye movements. These signals were filtered online with a cascaded integrator-comb antialiasing filter (half-power cutoff at 130 Hz) and digitized at 500 Hz. Impedances were <10 k Ω for scalp electrodes, and <20 k Ω for the EOG electrodes.

Offline analysis was performed using EEGLAB (Delorme & Makeig, 2004) and ERPLAB (Lopez-Calderon & Luck, 2014). Data were resampled at 250 Hz. A noncausal Butterworth bandpass filter was applied (half-amplitude cutoffs = 0.1 Hz and 30Hz, slope = 12 dB/octave), and all channels were re-referenced offline to the average of the mastoids. Independent component analysis (ICA) was used to correct for eye blinks and eye movements on the continuous EEG data. Artifact detection was performed on the epoched data to remove eye blinks that occurred at the time of the stimulus presentation using a 200-ms moving window with a step size of 100 ms and a ~ 300 μ V cutoff. Stimulus-locked ERPs were generated using an epoch of -200–1300 ms.

For the decoding analyses, the data were low-pass filtered at 6 Hz and resampled at 50 Hz for the ERP decoding. This resulted in a 4-dimensional matrix for each participant (time point (100) x orientation (8) x trial (~ 40) x electrode site (27)). Following the approach of Bae & Luck (2018), support vector machines with error-correcting output codes were used to perform the orientation decoding.

Decoding was implemented separately at each time point, and the data were divided into separate training and test sets. Data from a given orientation were averaged across multiple trials to improve the signal-to-noise ratio and decoding analyses were applied to these averages (not single trials; see Bae & Luck, 2018). Specifically, the ~ 40 trials for a given orientation were divided into three sets of ~ 13 trials, allowing us to make three different averages for each

orientation. A three-fold cross-validation procedure was used in which two averages for each orientation were used for training the decoder and the third was used for testing the decoder. Trials without a behavioral response were excluded (because this indicated that the participant did not perceive only one of the two gratings), and the number of trials for red and green responses was floored. A different classifier was trained to discriminate between each orientation and all other orientations (one vs. all decoding). When classifying the orientation in the test set, the decoders were combined to create a single decision about the most likely orientation for a given test case. Decoding was considered correct only when the classifier picked out the correct one of the eight orientations (so chance was $1/8$ or 0.125). Decoding accuracy was calculated as the average percent correct for each of the eight averages used for testing. To test whether decoding accuracy was greater than chance (12.5%) we calculated one-sample t-tests at each individual time point during the 0-1200 ms epoch. A false discovery rate (FDR) correction was used to control for multiple comparisons (Benjamini & Yekutieli, 2001).

The EEG data on a given trial contained a mixture of brain activity corresponding to the orientation in the perceived eye and brain activity corresponding to the orientation in the unperceived eye. We separately decoded the orientation of the stimulus in the perceived and unperceived eyes. That is, in one run of decoding, we organized the trials according to the orientation in the perceived eye, creating training and test cases labeled according to the orientation in that eye. The orientation in the unperceived eye varied randomly across these cases, so decoding accuracy reflected the information being coded from the perceived eye. In another run of decoding, we organized the trials according to the orientation in the unperceived eye, and decoding accuracy reflected the information being coded from the perceived eye. Bae

and Luck (2018) have shown that this approach can be applied successfully when two independently varying sources of information are mixed together in the EEG signal.

Results

Behavioral Results

Previous research using the intermittent stimulus presentation design showed that the two colors, and hence the two eyes, were approximately equally likely to be reported as the perceived stimulus (Pitts et al., 2010). We replicated these results in our study, finding that participants reported seeing a red grating on 45% of trials and a green grating on 48% of trials, with no response (indicating a blend of red and green) on 6%

of trials (See Figure 3). In the physical alternation condition, participants reported seeing a red grating on 49% of trials, a green grating on 49% of trials, and they withheld their response on .07% of trials.

Flicker fusion photometry tests revealed very little variability in perceived luminance when comparing red and green stimuli across subjects. Controlling for luminance across subjects ensured we are not decoding a difference in luminance values between red and green (Lee et al., 1988).

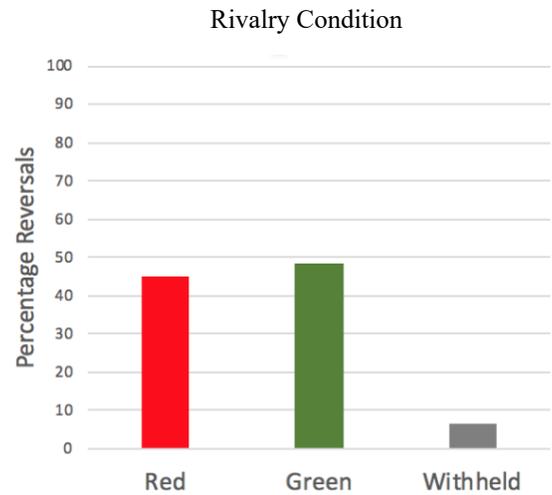


Figure 3: Participants reported seeing a red grating on 45% of trials and a green grating on 48% of trials, and withheld their response on 6% of trials.

Decoding Results

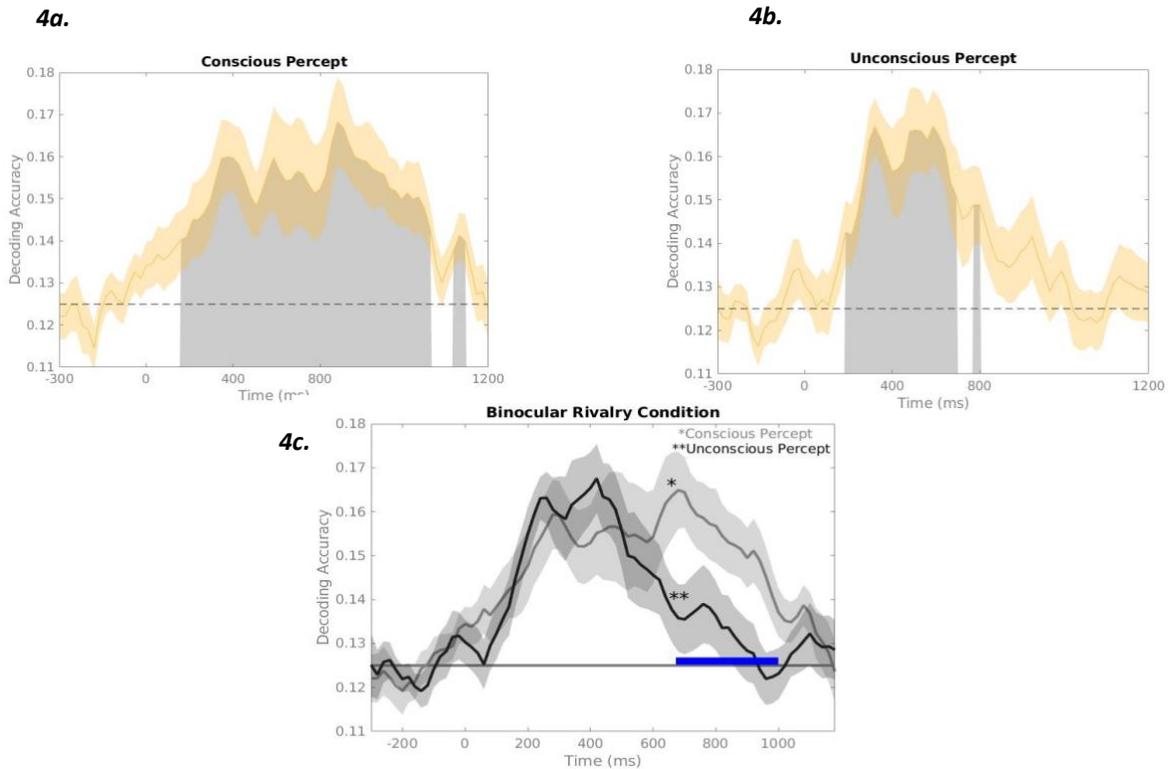


Figure 4a: Decoding accuracy for orientation for consciously reported percepts. **Figure 4b:** Decoding accuracy for orientation for unconscious on the same trial (decoding accuracy significantly greater than chance shaded grey). **Figure 4c.** Decoding accuracy for conscious and unconscious percepts on the same trial. The conscious percept diverges from the unconscious percept at 560 ms and is sustained past 1000 ms, unlike the unconscious percept which returns to chance after the stimulus offset.

Figures 4a & 4b show decoding accuracy for the orientation of the color that was reported (the conscious percept; Figure 4a) and the color that wasn't reported (the unconscious percept; Figure 4b), respectively. We found a period of decoding accuracy that was significantly greater than chance (grey shaded) for both the conscious and unconscious percepts. These figures show that decoding accuracy exceeded chance starting at approximately the same time for conscious (140

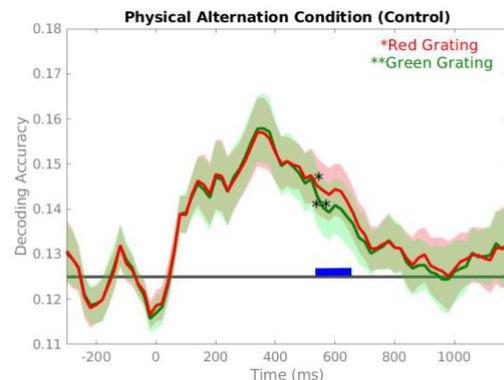


Figure 5: Decoding accuracy for red and green gratings in the non-rivalrous physical alternation control condition. Decoding accuracy was nearly identical for red and green gratings, except for a brief period between 560-660 ms when decoding accuracy was significantly different (blue line).

ms) and unconscious percepts (160 ms). However, Figure 4c shows that decoding accuracy remained well above chance until 1200 ms for the conscious percept whereas it fell quickly to chance by 950 ms for the unconscious percept. Moreover, decoding accuracy was significantly greater for the orientation in the reported color than for the orientation in the unreported color between 660 ms and 1000 ms (Figure 4c; blue line).

Figure 5 compares decoding accuracy for the red and green gratings in the non-rivalrous physical alternation control condition. Decoding accuracy was nearly identical for red and green gratings at most time points, although there was a brief period from 560-660 ms (blue line) in which decoding accuracy was significantly greater for the red gratings than for the green gratings. Note that absolute decoding accuracy for the physical alternation cannot be compared with the absolute decoding accuracy for the binocular rivalry condition, because the decoding was based on different numbers of trials in these two conditions, which directly impacts decoding accuracy.

Discussion

We examined the time course of conscious perception using the temporal precision of EEG decoding methodology during binocular rivalry. We found that neural information about orientation was identical for the perceived and unperceived gratings until at least 500 ms after stimulus onset. After that point, the consciously perceived information diverged from the unreported, unconscious information. The difference between the conscious and unconscious information began at ~560. The decoding of the consciously perceived orientation was sustained until ~1000 ms, whereas the decoding of the unconscious orientation fell to chance rapidly after stimulus offset. This pattern of results is consistent with theories of consciousness (e.g., Global

Neuronal Workspace Theory) which argue that the awareness occurs as a result of an ignition process that is indexed by the P3 wave. The lack of an earlier differentiation between perceived and unperceived information is not consistent with theories proposing that awareness arises rapidly.

Naturally, only correlational claims can be made regarding the decodability of conscious content, even though this experimental design aimed to isolate the neural correlates of awareness by controlling the retinal stimulus while only the conscious percept changed. However, the robust above-chance decoding accuracy of the unperceived stimulus from the time of stimulus onset until the time of stimulus offset is an interesting manipulation check regarding what information decoding techniques are able to capture. That is, the robust decoding of the information presented to the unperceived eye indicates that our method provided substantial sensitivity for assessing the information content present in the neural signal.

The present study was designed to avoid some of the problems with studies using the contrastive approach, which focus on differences in neural activity between trials in which a stimulus is consciously perceived and trials in which the stimulus is not perceived. That design makes it difficult to differentiate between true NCCs and neural signals associated with the prerequisites and consequences of conscious perception. For example, the P3b component is present when someone is aware of a stimulus and absent when they are unaware, but previous studies found that the P3b is a likely signature of report-related processes rather than awareness per se (Pitts et al., 2014). In the present study, we decoded the task irrelevant feature of orientation, not the color participants reported. Therefore, we only examined trials on which a stimulus was reported rather than comparing reported and unreported trials, so this is not an issue in the present study. Instead, we examined the trials in which the color in one eye was perceived

and the color in the other eye was not on the same trial, comparing the ability to decode the orientations presented to these two eyes. We did require participants to report the orientation from the perceived eye on 10% of trials, but the experiment was designed so that we could not be decoding the report per se. That is, we randomized the location of the eight orientation gratings on the probe screen so that we wouldn't be decoding response preparation.

This leads to the question of whether there is something fundamentally unique about binocularly rivalry that would cause the conscious percept to depart from the unconscious percept on the same trial, particularly as late as ~560 ms, which is even later than P3b component. Previous research has shown that reversals in binocular rivalry have occurred as early as 200 ms (Pitts et al., 2010). However, current research in single cell recordings in monkeys show rivalrous perceptual reversals over time occurring at a comparable time course to what we see in our data (Kapoor et al., 2020). Follow-up studies examining the P3b during rivalry could help gauge whether the phenomenon of rivalry is causing conscious perception to resolve later than anticipated since we would ordinarily expect the P3b to occur between ~300-500 ms, not at 560 ms, the time at which we found that decoding accuracy differed between the conscious and unconscious orientations.

Neuroscientists and philosophers alike might argue that the above chance-decoding that was present beginning shortly after stimulus onset also represents conscious contents (see Block phenomenal consciousness; Block, 2011). However, one would still need an explanation for the divergence of decodability at ~560 ms. There is some evidence that rivalrous competition can occur without conscious awareness, and therefore binocular rivalry is not an ideal paradigm for isolating awareness if neural activity correlated with rivalry can occur unconsciously (Zou et al.,

2016). Nonetheless, the divergence between conscious and unconscious percepts at 560 ms still needs explanation.

Characterizing the temporal dynamics of conscious representational content with these methods allows us to home in on the *timing* of specific content, giving us better temporal precision in isolating NCCs, and bringing us closer to the “minimal neural representational system” of a content NCC, helping us to adjudicate between the ongoing debate whether conscious contents are represented in the prefrontal cortex (PFC) or whether activity in the posterior cortical areas represent conscious contents.

References

- Aru, J., Bachmann, T., Singer, W., & Melloni, L. (2012). Distilling the neural correlates of consciousness. *Neuroscience & Biobehavioral Reviews*, *36*(2), 737–746.
<https://doi.org/10.1016/j.neubiorev.2011.12.003>
- Baars, B.J. (1997). *Baars, B. J. (1997). Contrastive phenomenology: A thoroughly empirical approach to consciousness. The nature of consciousness: Philosophical controversies, 187-202.*
- Babiloni, C., Vecchio, F., Miriello, M., Romani, G. L., & Rossini, P. M. (2006). Visuo-spatial Consciousness and Parieto-occipital Areas: A High-resolution EEG Study. *Cerebral Cortex*, *16*(1), 37–46. <https://doi.org/10.1093/cercor/bhi082>
- Bae, G.-Y., & Luck, S. J. (2018). Dissociable Decoding of Spatial Attention and Working Memory from EEG Oscillations and Sustained Potentials. *The Journal of Neuroscience*, *38*(2), 409–422. <https://doi.org/10.1523/JNEUROSCI.2860-17.2017>
- Benjamini, Y., & Yekutieli, D. (2001). The Control of the False Discovery Rate in Multiple Testing under Dependency. *The Annals of Statistics*, *29*(4), 1165–1188.
- Blake, R. (2022). The Perceptual Magic of Binocular Rivalry. *Current Directions in Psychological Science*, *31*(2), 139–146. <https://doi.org/10.1177/09637214211057564>
- Block, N. (2011). Perceptual consciousness overflows cognitive access. *Trends in Cognitive Sciences*, *15*(12), 567–575. <https://doi.org/10.1016/j.tics.2011.11.001>
- Brouwer, G. J., & Heeger, D. J. (2009). Decoding and Reconstructing Color from Responses in Human Visual Cortex. *Journal of Neuroscience*, *29*(44), 13992–14003.
<https://doi.org/10.1523/JNEUROSCI.3577-09.2009>

- Crick, F., & Koch, C. (1990). Towards a neurobiological theory of consciousness. *Seminars in The Neurosciences*, 13.
- Dehaene, S., & Changeux, J.-P. (2011). Experimental and Theoretical Approaches to Conscious Processing. *Neuron*, 70(2), 200–227. <https://doi.org/10.1016/j.neuron.2011.03.018>
- Dehaene, S., Changeux, J.-P., & Naccache, L. (2011). The Global Neuronal Workspace Model of Conscious Access: From Neuronal Architectures to Clinical Applications. In S. Dehaene & Y. Christen (Eds.), *Characterizing Consciousness: From Cognition to the Clinic?* (pp. 55–84). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-18015-6_4
- Dehaene, S., & Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. *Cognition*, 79(1), 1–37. [https://doi.org/10.1016/S0010-0277\(00\)00123-2](https://doi.org/10.1016/S0010-0277(00)00123-2)
- Del Cul, A., Baillet, S., & Dehaene, S. (2007). Brain Dynamics Underlying the Nonlinear Threshold for Access to Consciousness. *PLoS Biology*, 5(10). <https://doi.org/10.1371/journal.pbio.0050260>
- Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1), 9–21. <https://doi.org/10.1016/j.jneumeth.2003.10.009>
- Haxby, J. V. (2001). Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex. *Science*, 293(5539), 2425–2430. <https://doi.org/10.1126/science.1063736>

- Haynes, J.-D., & Rees, G. (2005). Predicting the Stream of Consciousness from Activity in Human Visual Cortex. *Current Biology*, *15*(14), 1301–1307.
<https://doi.org/10.1016/j.cub.2005.06.026>
- Kamitani, Y., & Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nature Neuroscience*, *8*(5), 679–685. <https://doi.org/10.1038/nn1444>
- Kapoor, V., Dwarakanath, A., Safavi, S., Werner, J., Besserve, M., Panagiotaropoulos, T. I., & Logothetis, N. K. (2020). Decoding the contents of consciousness from prefrontal ensembles. *BioRxiv*, 2020.01.28.921841. <https://doi.org/10.1101/2020.01.28.921841>
- King, J.-R., Pescetelli, N., & Dehaene, S. (2016). Brain Mechanisms Underlying the Brief Maintenance of Seen and Unseen Sensory Information. *Neuron*, *92*(5), 1122–1134.
<https://doi.org/10.1016/j.neuron.2016.10.051>
- Koivisto, M., Mäntylä, T., & Silvanto, J. (2010). The role of early visual cortex (V1/V2) in conscious and unconscious visual perception. *NeuroImage*, *51*(2), 828–834.
<https://doi.org/10.1016/j.neuroimage.2010.02.042>
- Koivisto, M., & Revonsuo, A. (2003). An ERP study of change detection, change blindness, and visual awareness. *Psychophysiology*, *40*(3), 423–429. <https://doi.org/10.1111/1469-8986.00044>
- Koivisto, M., & Revonsuo, A. (2010). Event-related brain potential correlates of visual awareness. *Neuroscience & Biobehavioral Reviews*, *34*(6), 922–934.
<https://doi.org/10.1016/j.neubiorev.2009.12.002>
- Kriegeskorte, N., Formisano, E., Sorger, B., & Goebel, R. (2007). Individual faces elicit distinct response patterns in human anterior temporal cortex. *Proceedings of the National Academy of Sciences*, *104*(51), 20600–20605. <https://doi.org/10.1073/pnas.0705654104>

- Lamme, V. A. F. (2006). Towards a true neural stance on consciousness. *Trends in Cognitive Sciences*, 10(11), 494–501. <https://doi.org/10.1016/j.tics.2006.09.001>
- Lamme, V. A. F. (2010). How neuroscience will change our view on consciousness. *Cognitive Neuroscience*, 1(3), 204–220. <https://doi.org/10.1080/17588921003731586>
- Lamy, D., Salti, M., & Bar-Haim, Y. (2008). Neural Correlates of Subjective Awareness and Unconscious Processing: An ERP Study. *Journal of Cognitive Neuroscience*, 21(7), 1435–1446. <https://doi.org/10.1162/jocn.2009.21064>
- Lau, H. C. (2007). A higher order Bayesian decision theory of consciousness. In R. Banerjee & B. K. Chakrabarti (Eds.), *Progress in Brain Research* (Vol. 168, pp. 35–48). Elsevier. [https://doi.org/10.1016/S0079-6123\(07\)68004-2](https://doi.org/10.1016/S0079-6123(07)68004-2)
- Lau, H., & Morales, J. (2022). The Neural Correlates of Consciousness. In *The Oxford Handbook of the Philosophy of Consciousness*. Oxford University Press.
- Lau, H., & Rosenthal, D. (2011). Empirical support for higher-order theories of conscious awareness. *Trends in Cognitive Sciences*, 15(8), 365–373. <https://doi.org/10.1016/j.tics.2011.05.009>
- Lee, B. B., Martin, P. R., & Valberg, A. (1988). The physiological basis of heterochromatic flicker photometry demonstrated in the ganglion cells of the macaque retina. *The Journal of Physiology*, 404, 323–347. <https://doi.org/10.1113/jphysiol.1988.sp017292>
- Leopold, D. A., & Logothetis, N. K. (1999). Multistable phenomena: Changing views in perception. *Trends in Cognitive Sciences*, 3(7), 254–264. [https://doi.org/10.1016/S1364-6613\(99\)01332-7](https://doi.org/10.1016/S1364-6613(99)01332-7)
- Logothetis, N. K., & Schall, J. D. (1989). Neuronal Correlates of Subjective Visual Perception. *Science*, 245(4919), 761–763. JSTOR.

- Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: An open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience*, 8.
<https://www.frontiersin.org/articles/10.3389/fnhum.2014.00213>
- Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: Multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences*, 10(9), 424–430.
<https://doi.org/10.1016/j.tics.2006.07.005>
- Pitts, M. A., & Britz, J. (2011). Insights from Intermittent Binocular Rivalry and EEG. *Frontiers in Human Neuroscience*, 5. <https://doi.org/10.3389/fnhum.2011.00107>
- Pitts, M. A., Martinez, A., & Hillyard, S. A. (2010). When and where is binocular rivalry resolved in the visual cortex? *Journal of Vision*, 10(14), 25–25.
<https://doi.org/10.1167/10.14.25>
- Pitts, M. A., Padwal, J., Fennelly, D., Martínez, A., & Hillyard, S. A. (2014). Gamma band activity and the P3 reflect post-perceptual processes, not visual awareness. *NeuroImage*, 101, 337–350. <https://doi.org/10.1016/j.neuroimage.2014.07.024>
- Salti, M., Monto, S., Charles, L., King, J.-R., Parkkonen, L., & Dehaene, S. (2015). Distinct cortical codes and temporal dynamics for conscious and unconscious percepts. *ELife*, 4.
<https://doi.org/10.7554/eLife.05652>
- Sandberg, K., Andersen, L. M., & Overgaard, M. (2014). Using multivariate decoding to go beyond contrastive analyses in consciousness research. *Frontiers in Psychology*, 5.
<https://doi.org/10.3389/fpsyg.2014.01250>
- Sergent, C., Baillet, S., & Dehaene, S. (2005). Timing of the brain events underlying access to consciousness during the attentional blink. *Nature Neuroscience*, 8(10), 1391–1400.
<https://doi.org/10.1038/nn1549>

- Seymour, K., Clifford, C. W. G., Logothetis, N. K., & Bartels, A. (2010). Coding and Binding of Color and Form in Visual Cortex. *Cerebral Cortex*, *20*(8), 1946–1954.
<https://doi.org/10.1093/cercor/bhp265>
- Supèr, H., Spekreijse, H., & Lamme, V. A. F. (2001). A Neural Correlate of Working Memory in the Monkey Primary Visual Cortex. *Science*, *293*(5527), 120–124.
<https://doi.org/10.1126/science.1060496>
- Tononi, G., & Koch, C. (2008). The Neural Correlates of Consciousness. *Annals of the New York Academy of Sciences*, *1124*(1), 239–261. <https://doi.org/10.1196/annals.1440.004>
- Weil, R. S., & Rees, G. (2010). Decoding the neural correlates of consciousness: *Current Opinion in Neurology*, *23*(6), 649–655. <https://doi.org/10.1097/WCO.0b013e32834028c7>
- Zou, J., He, S., & Zhang, P. (2016). Binocular rivalry from invisible patterns. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(30), 8408–8413.
<https://doi.org/10.1073/pnas.1604816113>

IV. Summary

We have shown that computational cognitive neuroscience can be used to test predictions made by some of the most influential theoretical views of the neural correlates of consciousness. Artificial models and cutting edge methodology enable neuroscientists to investigate the temporal dynamics of representational content with new precision. We empirically tested commitments of theoretical models which assert that the prefrontal cortex plays a crucial role in consciousness experience (e.g., Global Workspace Theory) vs. a class of models which asserts that the prefrontal cortex is not paramount for conscious perception. We found that predictions of this latter class of theories is not supported by evidence.

In Chapter V, consistent with GWT that awareness occurs as a result of an ignition process indexed by the P3 wave occurring at ~300 ms after stimulus onset, our findings suggest that early visual areas are not encoding the conscious perception of orientation but awareness instead requires extended processing over time to extract a largely contrast-independent perception of orientation. Analogously, in Chapter VI the lack of an early differentiation in decoding accuracy between perceived and unperceived information during binocular rivalry is inconsistent with theories that propose awareness arises in early visual areas. Together, these two studies both support theories in which visual awareness of a stimulus arises at a relatively late stage.

In summary, the introduction of new methodological, statistical and computational advancements in neuroscientific research on consciousness can help identify putative NCCs and sharpen theoretical views regarding the functional localization, temporal characteristics, and computational architecture of consciousness.