Color: Functional Organization and Behavior

by Rosa Lafer-Sousa

B.A. Neuroscience
Wellesley College, 2009

SUBMITTED TO THE DEPARTMENT OF BRAIN AND COGNITIVE SCIENCES IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY IN COGNITIVE SCIENCE
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2019

Signature of Author: Signature redacted
Department of Brain and Cognitive Sciences
May 1, 2019

Certified by: Signature redacted
Nancy Kanwisher
Walter A. Rosenblith Professor of Cognitive Neuroscience
Thesis Supervisor

Certified by: Signature redacted
Matthew A. Wilson
Sherman Fairchild Professor of Neuroscience and Picower Scholar
Director of Graduate Education for Brain and Cognitive Sciences

Color: Functional Organization and Behavior

By Rosa Lafer-Sousa

Submitted to the Department of Brain and Cognitive Sciences on May 1, 2019 in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Cognitive Science

Abstract

Color is a fundamental aspect of visual experience that confers a myriad of behavioral advantages: finding objects in cluttered scenes, recognizing familiar objects, and gleaning information about the material composition and state of objects (e.g. the edibility of fruit) and agents in the world (e.g. health or emotional status). As famously pointed out by Marr (1980), a full understanding of perception requires an analysis of the computations performed, the algorithms that carry out those computations, and the implementation of those algorithms in the physical hardware of the brain. This thesis employs psychophysical methods and functional imaging to tackle questions about human color vision at all three levels: what it is used for, how we solve the classic problem of color constancy, and how our color processing machinery is functionally organized in the brain. Chapter 1 provides a brief survey of the background to these questions. Chapter 2 describes functional MRI studies in humans that find both segregation and convergence of the processing of color and shape in the brain, as well as evidence for the homology of the color system between humans and macaques. Chapter 3 uses psychophysics and a recently discovered ambiguous color stimulus ('#theDress') to investigate the cues and assumptions used by the human visual system to constrain the classic ill-posed problem of inferring the intrinsic reflectance of an object by discounting the spectral properties of the illuminant. Specifically, these studies find evidence that color constancy is mediated by sensory, perceptual, and cognitive factors (i.e., low-level features, inferences about 3D scene geometry, prior knowledge, and attention), and provide the first evidence that human skin is a sufficient cue to infer the illuminant and bring about color constant percepts. Chapter 4 uses psychophysics to evaluate the impact of memory on the color appearance of familiar objects and faces. The study finds a novel perceptual illusion that reveals the role of memory for face color in perceptual experience and social communication, sheds light on the selective pressures for the evolution of trichromatic vision in primates, and demonstrates the powerful ability of cognition to influence perception. Taken together, these studies provide clues about the perceptual and neural mechanisms underlying our rich experience of a colorful world.

Thesis Supervisor: Nancy Kanwisher
Title: Walter A. Rosenblith Professor of Cognitive Neuroscience

Acknowledgments

Over the course of my graduate career I have had the good fortune and pleasure of working with a community of deeply insightful and supportive individuals. I am tremendously grateful to these people. This page will go a small way toward expressing that gratitude.

First and foremost, I am indebted to my advisor Nancy Kanwisher for her persistent guidance and support in my scientific endeavors and in life. Her energy, compassion, scientific depth and passion are unmatched and a constant source of inspiration. She is a remarkable person and I am forever grateful to have her in my corner.

I am also grateful to the other members of my committee, Jim DiCarlo, Ted Adelson, and Geoffrey Boynton for their valuable feedback and unique perspectives, which pushed me to develop richer frameworks and directions for my work.

I am deeply grateful to my undergraduate advisor Bevil Conway, who was integral to the work presented in this thesis and my development as a scientist. His creativity and passion for color is infectious, and something I have not been able to shake since first stepping into his classroom a decade ago. I would not be who I am today without his unending support and guidance.

I am grateful to my collaborators, Katherine Hermann, Maryam Hassantash, and Arash Afraz, who were integral to the work presented in this thesis. As well, this work would not have been possible without the support of the fabulous staff in the Martinos Imaging center (Atsushi, Steve, Sheeba, and Dimitrios).

Thanks to my colleagues from the Kanwisher lab (Michael, Alex K., Ben, Matt, Jason, Zeynep, Sam, Idan, Dana, Anna, Jenelle, Leyla, Caroline, Jeff, Katharina, Anya, Rattan, Sarah), Conway lab (Cleo, Galina, Kaitlin, Kathernine, Walid, Josh), and broader BCS community (Alex P., Galen, Chris, Julia, Julian, Shaiyan, Hillary, Dorit, Rishi, Pedro, Jorie) for their intellectual stimulation, constant support, lively banter, and good nature. I look forward to sustaining lifelong friendships and scientific collaborations with these promising young scientists.

I have also had the pleasure of forging relationships with numerous talented individuals at conferences and summer schools. Specific shout outs go to Danny Garside, Stacey Aston, Jake Yates, Qasim Zaidi, Anya Hurlbert, and Sergio Nascimento.

Finally, I am forever indebted to my friends and family for their tireless spiritual, emotional and intellectual support. To Danny and Sarah, thank you for being my rocks. To my parents and brother, thank you for your love, encouragement, and for instilling a deep and unwavering appreciation for science and creativity in my bones.

Table of Contents

Chapter 1: Introduction

Chapter 2: Color-biased Regions of the Ventral Visual Pathway Lie between Face- and Place-Selective Regions in Humans, as in Macaques

Chapter 3: Color constancy: resolving ambiguous color signals

Chapter 4: Paradoxical impact of memory on color appearance of faces

Chapter 5: Conclusions

Chapter 1: Introduction

1.1 Motivation and background

Color is a fundamental aspect of visual experience that helps us solve three central challenges of visual perception. First, color helps us find objects and segment them from cluttered backgrounds, such as finding berries in a leafy bush (Figure 1A), or our car in a crowded parking lot. Second, color helps us identify objects and scenes where color is more diagnostic than shape, as in distinguishing between a lemon and a lime, or a lush scene vs. an arid one. Third, color provides behaviorally relevant information about the material composition and state of entities in the world, such as the ripeness of a banana (Figure 1D), or the health (Figure 1C), emotional state (Figure 1E), or dominance status of another individual (Figure 1B).

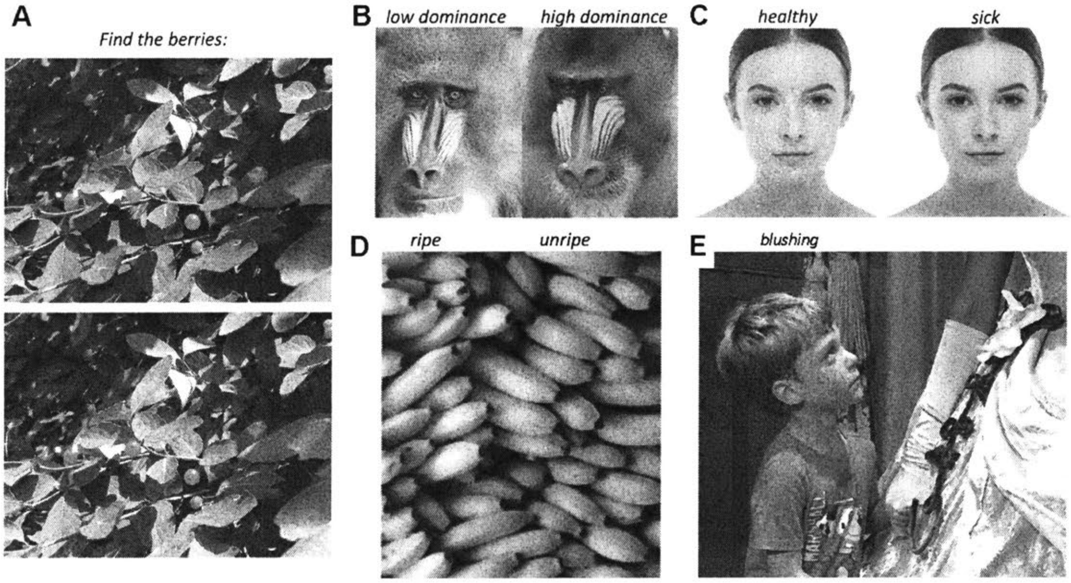


Figure 1. The behavioral utility of color. A) Color is useful in finding the berries amongst the leaves (bottom), which is far more difficult when color is removed (top). B) Color provides information about social status. For example, intensity of facial coloration is associated with dominance rank in the mandrill, with dominant individuals displaying the most vibrant colors. C) Color provides cues to health. For example, reduction in blood oxygenation and flow to certain areas of skin causes a decrease in redness (right panel shows a simulation). D) Color provides information about edibility, for example, the ripeness of a banana. E) Color conveys information about emotional status. For example, blushing upon meeting a Disney princess.

For perceived color to provide meaningful information about entities in the world (objects, agents, scenes, and materials), it must reveal properties that are intrinsic to them, regardless of the specific viewing conditions. But herein lies one of the central, classic challenges in color vision: the spectrum of light reflected from an object's surface (the 'color signal') is not determined only by the intrinsic reflectance properties of the object, but also by the spectral content of the light illuminating it. However, the light illuminating the object is not constant and often varies substantially over space and time. Despite these variations, our experience of the color of objects is remarkably stable, a phenomenon termed "color constancy." As Hering noted long ago, *"The approximate constancy of the colors of seen objects, in spite of large quantitative or qualitative changes of the general illumination of the visual field, is one of the most noteworthy and most important facts in the field of physiological optics" (Hering, 1878).* Despite centuries of inquiry, this fundamental question of how the visual system disentangles color signals to bring about stable color percepts remains unsolved.

As famously pointed out by Marr (1980), a full understanding of perception requires an analysis of the computations performed, the algorithms that carry out those computations, and the implementation of those algorithms in the physical hardware of the brain. This thesis tackles questions about human color vision at all three levels: what it is used for (Chapter 4), how we solve the classic problem of color constancy (Chapter 3), and how our color processing machinery is functionally organized in the brain (Chapter 2). Next I briefly sketch the background and motivation for these questions.

How do we perceive relatively constant colors?

How do humans perceive object color invariant to viewing conditions? Answering this question will require investigating many facets of human color vision. What cues do we use in the scene to constrain this ill-posed problem? How are they integrated, weighed, and updated? What role do priors play? To what extent are low-, mid-, and high-level visual processes involved? Do individuals differ importantly from each other in how they solve these problems? These questions have long been addressed with psychophysical methods, but are far from resolved, as briefly outlined next.

Some of the earliest models of color constancy identified a handful of biologically plausible low-level sensory mechanisms that could account for human constancy well under simple viewing conditions (Von Kries 1878; Stiles, 1959; Helson, 1964; D'Zmura & Lennie, 1986; Foster & Nascimento 1994; Chichilnisky & Wandell, 1995; Webster and Mollon, 1995). For example, retinal adaptation and chromatic contrast are sufficient to explain color constancy for abstract 2-dimensional matte surfaces illuminated by a diffuse light source (i.e., 'Mondrian worlds'). However, these low-level mechanisms fail to explain color constancy for natural surfaces (Brainard & Wandell, 1986; Webster & Mollon, 1997) and three-dimensional scenes (Adelson, 1993; Kraft & Brainard, 1999; Hedrich et al., 2009), which produce more illumination ambiguity in the retinal image than do simple 2-dimensional displays.

On the other hand, the physics of 3D image formation and the reliability of certain intrinsic properties of the environment (e.g. the physical properties of common illuminants and surfaces) produce statistical regularities ('cues') that could constrain the problem and enable the recovery of stable color percepts. In effect, the visual system must exploit these cues to infer the causal structure of world that created the retinal image. This process likely involves mid-level perceptual processes representing surfaces, contours, and grouping, and high-level cognitive processes that incorporate knowledge about objects, materials, and scenes. Over the last several decades, researchers have developed models of biologically plausible, color constant visual systems that exploit these regularities. Such models generally comprise a mathematical description of an idealized world and an algorithm that can be used to compute invariant surface color descriptors within the specified environment. Typically, these algorithms hinge on making an estimate of the illuminant, and differ primarily in the cues they use to do this: e.g. surface specularity (Lee, 1986; D'Zmura & Lennie, 1986; Maloney, 2002), shadows (D'Zmura, 1992), mutual illumination (Funt, Drew, & Ho, 1991; Bloj et al., 1999), reference surfaces (Buchsbaum, 1980), subspace constraints (Maloney & Wandell, 1986), scene averages (Buchsbaum, 1980), memory color for familiar objects (Ling, Y., & Hurlbert, A. 2006, 2008; Crichton et al., 2012), and more.

A substantial body of work has focused on behaviorally testing which of the candidate cues and computational strategies proposed by the modeling literature are actually used by humans to infer properties of the visual world (see Smithson, 2005 for a review). Results from experiments that selectively modulate the availability of potentially informative cues suggest that we use many visual cues to achieve color constancy (Kraft & Brainard, 1999; Yang & Maloney, 2001; Boyaci et al., 2006). Understanding how color constancy is achieved will ultimately require determining the mathematical rules that dictate how different cues are weighted and how dynamic this process is with respect to cue availability, cue reliability, cue covariance, and prior expectations.

One classic empirical strategy in visual psychophysics is to exploit visual illusions to reveal the assumptions the visual system makes about the world (Adelson, 1993; Kingdom 2008). This strategy has proven fruitful across a wide variety of visual domains such as the perception of motion, shape, color, lightness, and depth. One class of illusions that has been particularly informative is ambiguous stimuli, which give rise to more than one stable percept. Ambiguous stimuli abound in the shape and motion literature, with famous examples including the Necker cube, the Schroeder staircase, structure from motion, monocular rivalry, and binocular rivalry. These stimuli have been widely used to gain insight into the sensory, perceptual, and cognitive processes that mediate perceptual constancy (Scocchia, Valsecchi, & Triesch, 2014). Yet there is one domain of visual information processing that has until recently lacked such ambiguous stimuli: color. While there are many examples of color illusions for which physically identical stimuli are made to appear differently colored by varying the context in which they appear (e.g. Figure 2), any given instance gives rise to a single stable color percept within and across individuals.

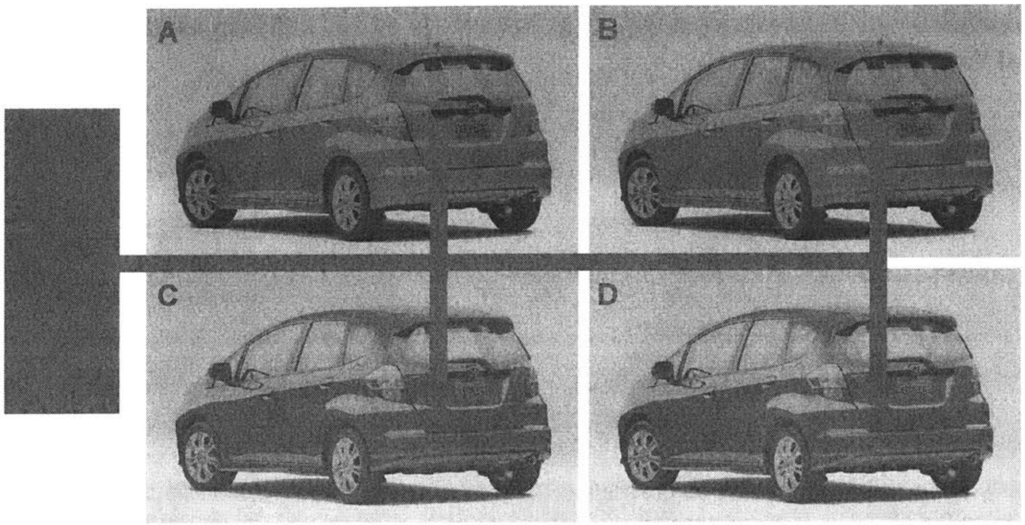


Figure 2. Color ≠ wavelength; color is a sensation created in the brain. The cars in each panel appear to be differently colored: **A**, Orange. **B**, Blue. **C**, Red. **D**, Teal. However, the pixel colors that make up their surfaces are gray when viewed in isolation. That is, the wavelengths that reach the eye are the same, but give rise to different colored percepts. The illusion occurs because the visual system uses the global scene context to infer and discount the cued illuminant. (Stimuli created by RLS, all rights reserved).

The lack of color stimuli prone to multiple interpretations could reflect the robust nature of color constancy. Importantly, however, an image recently emerged that appeared to elicit multiple color percepts, raising that tantalizing possibility that the elusive 'Color Necker cube' had been found: A poor-quality image of a dress ("#theDress") created a stir on social media because of apparent striking individual differences in how it was perceived. While some viewers believed the dress to be blue and black, others insisted that the dress was actually white and gold. Vision scientists became excited about this stimulus and the insights it offered: "The one certainty is that vision scientists have rarely deliberately devised such a powerful stimulus for studying individual differences in colour perception, much less encountered one accidentally. The generation of experiments it spawns will reveal much about how the brain works, both in making colours and in making science." (Hurlbert and Brainard, 2015).

In order to use #theDress as a tool to gain fundamental insights into color perception, we first need to know if it actually perceived as more than one categorical state. That is, do people really see it differently? If there is a simpler, albeit less exciting alternative, that would both seem more likely, and we are obligated to find out. The less exciting alternative is that perception of the dress is unimodal, a normal distribution around blue/brown (the colors of the pixels, when viewed in isolation), and that the differing reports arose as an artifact of how people's color reports were queried. When people were first asked about the dress' colors on social media, the question was posed as a two-alternative forced-choice ("Is this dress blue and black or white and gold"?). If you give people two alternatives, you will get two answers (Figure 3). In Chapter 3 I provide a rigorous assessment of how the dress is perceived using color-matching data collected from 2200 subjects, 53 under controlled conditions in the lab and the rest online. Critically, I tested subjects who had never seen the image before, and so had not been exposed to the 2AFC form of the query.

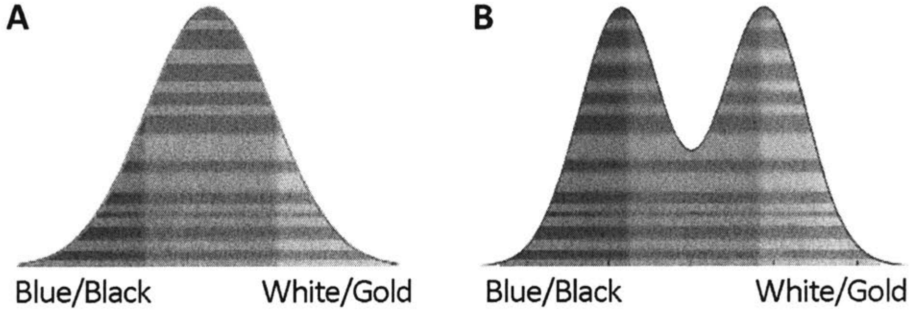


Figure 3. The pixels in #theDress image are light blue and brown. Social media asked: *"Is it blue & black or white and gold?"* Could the purported categorical percepts of #theDress be an artifact of a 2AFC question? A) Unimodal distribution of percepts around blue/brown (the pixel colors). B) Bimodal distribution of percepts, peaking at blue/black and white/gold. Both distributions would give rise to the impression of two categorical percepts when queried with a 2AFC.

The goal of this work is to discover the cues and assumptions the visual systems uses to constrain the ill-posed problem of color perception, and the extent to which individuals vary in this endeavor. To do this, I first rigorously characterize the extent of the individual differences initially reported on social media. I then characterize the properties of the stimulus that give rise to its striking ambiguity, the factors that mediate how it is resolved by different individuals (including inferences about the lighting conditions, prior knowledge and experience, short-term adaptation, image scale and spatial frequency, and attention), and the extent to which the image is subject to reversals (i.e. can an individual switch from one percept to the other?). The work culminates with an experiment leveraging the image as a tool to test a long hypothesized candidate cue for recovering the illuminant: human skin.

Natural scenes usually contain familiar objects and surfaces, and the remembered colors of these familiar surfaces ('memory color') could support color constancy by serving as a reference to ascertain the spectral bias of the light source in the scene, akin to the use of a white balance card in photography (Hering 1878; Bramwell & Hurlbert, 1993; Ling, Y., & Hurlbert, A. 2006, 2008; Mitterer & de Ruiter 2008; Granzier & Gegenfurtner, 2012; Witzel et al. 2011). In particular, skin has been proposed as a potential cue for color constancy, due to its stable color statistics (Hurlbert, 2007; Crichton et al., 2012). These statistics, coupled with the fact that skin is viewable in almost every natural glance, make skin a potentially good cue to estimate the illuminant and discount it from the scene. I test this hypothesis for the first time using manipulations of the dress stimulus. In a separate line of work presented in Chapter 4, I further explore the role of memory color in mediating color perception, beyond supporting color constancy, seeking to shed light on the behavioral significance of color signals for fruit and faces.

What is the brain basis of trichromatic color perception, and is this system homologous in humans and macaques?

To fully understand color perception, we need to not only understand the computations and algorithms associated with color processing, but to also understand how those algorithms are implemented in the brain. The computation of color involves distinct stages of processing. A vast amount of physiological work has been carried out in monkeys to characterize each of these stages, which I detail next.

Color processing begins in the retina, with the differential activation of the three classes of cone photoreceptors sensitive to long (L), middle (M), and short (S) wavelength light. It is the relative activation of the cones that forms the basis for color vision, but substantial cortical processing is required to bring about perceived color. Retinal signals are transmitted through the lateral geniculate nucleus (LGN) of the thalamus to primary visual cortex (V1). The inputs to V1 from the LGN comprise three anatomically separate channels, two that carry color information, and a third for achromatic luminance information (Chatterjee and Callaway, 2003). Because the cones' sensitivity profiles are largely overlapping, the color channels encode color information as difference signals, pitting cone activities against each other in an opponent fashion: L vs M, 'red-green' signals; S vs L+M; 'blue-yellow' signals. Color-responsive neurons in V1 are also cone-opponent, but comprise two types: single-opponent and double-opponent. Double-opponent cells generate a *spatial* chromatic contrast calculation, by pitting single-opponent signals against each over space (Conway, 2001). These local color contrast operations inform the brain about chromatic borders and form the building blocks for color constancy.

Are color and shape segregated in V1? Double-opponent cells are often orientation tuned and sensitive to luminance, which suggests color and shape are not segregated at this stage. On the other hand, their large receptive fields suggest they would not contribute to form processing nearly as precisely as would the strictly luminance-responsive orientation-tuned cells that populate much of V1, whose receptive fields are much finer (Conway, 2009).

Of course, color processing does not end in V1. V1 sends color signals on to the second visual area (V2), whose precise role in color processing is not well resolved (for a review see Conway, 2009). From V2, color signals are passed on to cortical area V4, the best-known and most fiercely debated stage in the cortical hierarchy for color processing. In 1973 Semir Zeki famously recorded from V4 and reported that 100% of V4 cells were hue-tuned, leading him to dub the region "the color center" (Zeki, 1973). Subsequently, Schein and Desimone (1990) recorded from the region and found that only a fraction (16%) of the cells were tuned to hue. And successive work has firmly established that neurons in this region are tuned to multiple other visual dimensions such as orientation, shape, motion, and depth (for a review see Roe et al., 2012). These conflicting findings were reconciled by a landmark study, carried out by Conway et al., (2007) who, using fMRI-guided single unit recording, found that V4 is patchy, containing sub compartments dedicated to hue, which they dubbed "globs." The globs are comprised of narrowly hue-tuned cells spatially arranged into chromotopic maps that reflect perceptual color space (i.e. nearby cells code nearby colors in perceptual color space) (Conway and Tsao, 2009; Bohon et al., 2016). This work not only resolved the conflicting prior reports, but also underscored the benefit of fMRI to overcome the severe sampling limitations of single-unit recording. Taken together, these observations make clear that V4 is involved in color processing, but not exclusively so. They also indicate that color responses in V4 mark a significant transformation from color encoded as cone-difference signals (in V1) to color encoded as hue, and establish the first representation of color in the visual pathway that corresponds with perceptual color space.

Anterior to V4, in IT (where high-level visual processing culminates), sensitivity to color has been found at multiple sites (Takechi et al 1997; Tootell et al., 2004; Koida & Komatsu, 2006; Conway et al., 2007), but until recently the extent of these regions and their relationship to shape/category selective regions was unresolved. In my prior work, using fMRI in macaques to localize neural biases for color, shape, faces, and scenes in IT (Lafer-Sousa and Conway, 2013), I found multiple color-biased regions at highly stereotyped locations spanning the length of IT (see Figure 4). The systematic arrangement of color-biased tissue mirrored the organization of the well-characterized face-patch system (Tsao and Livingstone, 2008). Further, while the color-biased regions were spatially segregated from face and scene selective cortex, they were systematically sandwiched between them, forming a tripartite map. These observations promoted the tantalizing notion that the color-biased regions might constitute a specialized multi-stage subsystem within IT, organized by principles common to the subsystems for faces and scenes. While the precise functions of the color-biased regions in IT are not well understood, recent work offers some clues: The responses of hue-tuned cells deep in IT (TE) can be modulated by task demands (Koida & Komatsu, 2006), implying an integration of color and behavior at this stage. As well, color-tuning biases in IT are predicted by object color statistics (estimated from an analysis of a massive image database of natural and artificial objects) (Rosenthal et al, 2018), implying color signals in IT play a role in real world object representations.

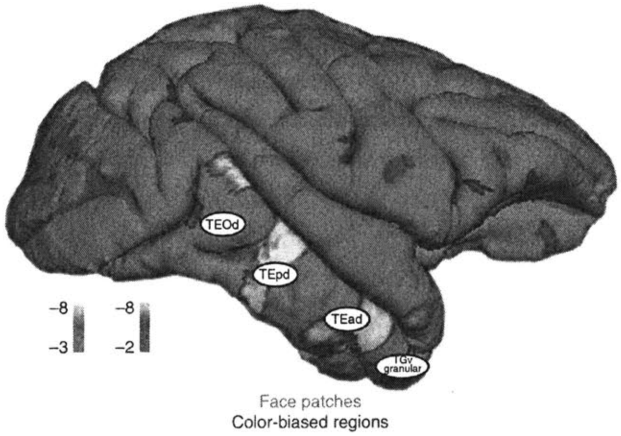


Figure 4. Color-biased patches and face patches on the lateral surface of one monkey along the superior temporal sulcus. In this case, the color-biased patches are shown in cyan/blue while the face patches are shown in yellow/orange.

Up to this point I have described work characterizing color processing in monkeys. Ultimately, we are interested in human visual processing. In humans, considerable evidence from patient and neuroimaging studies implicates specific cortical regions in the analysis of color. For example, stroke patients with damage to a region of the brain in posterior ventral temporal cortex acquire a condition called achromatopsia, which is a total or partial loss of color perception that is not necessarily accompanied by deficits in form or motion processing (though it is often accompanied by prosopagnosia) (e.g. Meadows, 1974; Bouvier and Engel, 2005). This observation prompted neuroimaging studies, which uncovered regions in posterior ventral temporal cortex (hV4 and VO-1) that respond preferentially to passively-viewed abstract stimuli that contain color compared to those that don't (McKeefry and Zeki, 1997; Hadjikhani et al., 1998; Wade et al., 2002, 2008; Cavina-Pratesi et al., 2010). As in monkeys, the representation of color in these regions corresponds well with perceptual color space (Brouwer and Heeger, 2009, 2013). Imaging studies requiring more effortful color processing, such as attention to detailed color information during the sequencing of hues or the retrieval of object-color knowledge, uncover a third, more anterior region (Beauchamp, 1999; Simmons et al., 2007).

However, little is known about the precise computations conducted within these regions, their specific causal role in perception, or their underlying connectivity. The tools we have available to answer these questions in humans are limited. It would be useful if we could extend the wealth of physiology data available in monkeys to understanding humans. This strategy makes the most sense if the macaque visual system is not only loosely similar to that of humans, but also homologous to it (that is, inherited from a common ancestor). Humans and monkeys diverged from our last common ancestor 30 million years ago, so it is possible our visual cortices have diverged considerably. Indeed, while early visual cortex of humans and monkeys is remarkably similar in its gross location and organization, the large swath of cortex anterior to it, responsible for high-level visual processing (IT in monkeys; ventral temporal cortex, VTC in humans) is found in different locations in the two species. In humans, it tracks the ventral surface of the brain, whereas in monkeys it falls on the lateral surface. Could these regions be homologs despite their very different locations in the brain?

Establishing homology is not trivial, and cannot be proven, only inferred. Common criteria for establishing homology include similarities in anatomy (e.g., similar connectivity, cyto- and myeloarchitecture) or function (similar neural responses and/or functional organization). When suspected homologies correspond to regions in different geographic locations, one strategy is to look for evidence of a 'regional homology'—a neighborhood of cortical regions present in a common ancestor whose geographic location was altered by the non-uniform expansion of cortex as new cortical areas emerged nearby (Orban et al, 2004). Functional evidence for a regional homology would be suspiciously similar internal organization of the suspected neighborhood, and an account for how cortical expansion patterns could have relocated the neighborhood to its present location.

In Chapter 2, I describe work investigating the similarity of the internal organization of macaque IT and human VTC, by assessing the relative location of color-biased regions to face and scene-selective regions, as a measure of the homology of these systems. Specifically, I ask whether color-biased regions are sandwiched between (and non-overlapping with) face and scene selective regions. Answering this question by comparing results from prior imaging studies is not tractable. There is significant individual variability in the precise locations of these regions, making comparisons of results obtained in different individuals prohibitive. As well, prior color imaging efforts have tended to employ simple abstract color stimuli, unlikely to drive responses in category selective regions, leaving open the possibility that category-contingent color responses have gone undetected. Critically, the present work localizes these regions in the same individual subjects, uses rich naturalistic stimuli to do so (movies of faces, bodies, scenes, objects, and scrambled objects, rendered in color and greyscale), and in a subset of subjects, compares the color activation patterns acquired with naturalistic stimuli to those acquired with simple abstract stimuli (drifting equiluminant chromatic gratings and achromatic luminance-contrast gratings).

To what extent are color and shape segregated in human VTC?

A fundamental question at the core of our understanding of color processing is whether or to what extent color is processed separately from other visual information. As noted above, this question was heatedly debated in the context of early work on V4: was V4 a "color area," or just another stage in the visual processing hierarchy, where neurons were tuned to many features, spectral and otherwise?

On the one hand, the spatial configuration of colors within a scene has a profound influence on color perception, leading some to argue that color and form are linked inextricably in visual cortical processing, especially in V1 (e.g. Gegenfurtner and Kiper, 2003). On the other hand, color presents computational challenges quite distinct from those associated with shape processing. The information provided by color and shape is often independent, for example in signaling a person's health state independent of face features, or the edibility of a piece of fruit, as when bananas ripen, turning from green to yellow, without changing shape. Further, some color behaviors are carried out independent of shape, like categorizing, discriminating, and detecting colors. For example, when searching for your car in a crowded parking lot it is natural to search for its color rather than its shape, as its shape will be largely occluded. These observations suggest that one might expect some segregation of color from shape. Evidence from the lesion literature in humans has not yielded a clear answer on this question, as deficits in color perception due to cortical lesions (achromatopsia) almost always co-occur with prosopagnosia (Bouvier & Engel, 2005). fMRI offers the possibility of addressing this question with finer spatial resolution than possible with lesion studies. In Chapter 2, I use fMRI in humans to ask to what degree cortical preferences for color (versus greyscale) stimuli overlap in the brain with cortical preferences for stimuli with coherent shapes (versus scrambled stimuli).

1.2. References

Adelson, E. (1993) Perceptual organization and the judgement of brightness. *Science*, 262, 2042-2044.

Beauchamp, M. S. (1999). An fMRI Version of the Farnsworth-Munsell 100-Hue Test Reveals Multiple Color-selective Areas in Human Ventral Occipitotemporal Cortex. *Cerebral Cortex,* (3).

Bohon, K. S., Hermann, K. L., Hansen, T., & Conway, B. R. (2016). Representation of Perceptual Color Space in Macaque Posterior Inferior Temporal Cortex (the V4 Complex). *ENeuro, 3*(4).

Bouvier, S. E., & Engel, S. A. (2005). Behavioral Deficits and Cortical Damage Loci in Cerebral Achromatopsia. *Cerebral Cortex,* (2).

Boyaci, H., Doerschner, K., Snyder, J., & Maloney, L. T. (2006) Surface color perception in three-dimensional scenes. *Visual Neuroscience,* 23, 311-21

Brainard, D. H., & Wandell, B. A. (1986) Analysis of the retinex theory of color vision. *Journal of the Optical Society of America A*, 3, 1651-1661.

Brainard, D.H. & Hurlbert, A.C. (2015) Colour vision: understanding #theDress. *Current Biology*, 25, R551-R554.

Bramwell, D.I. & Hurlbert, A.C. (1993) The role of object recognition in colour constancy. *Perception*, 22, 62.

Brouwer, G. J., & Heeger, D. J. (2009). Decoding and Reconstructing Color from Responses in Human Visual Cortex. *Journal of Neuroscience,* (44).

Brouwer, G. J., & Heeger, D. J. (2013). Categorical Clustering of the Neural Representation of Color. *Journal of Neuroscience,* (39).

Buchsbaum, G. (1980) A spatial processor model for object colour perception. *Journal of the Franklin Institute*, 310, 1-26.

Cavina-Pratesi, C., Kentridge, R., Heywood, C., & Milner, A. (2010). Separate Channels for Processing Form, Texture, and Color: Evidence from fMRI Adaptation and Visual Object Agnosia. *Cerebral Cortex,* (10).

Chatterjee, S. & Callaway, E.M. (2003) Parallel colour-opponent pathways to primary visual cortex. *Nature*, 426, 668-671.

Chichilnisky, E. J., & Wandell, B. A. (1995) Photoreceptor sensitivity changes explain color appearance shifts induced by large uniform backgrounds in dichoptic matching. *Vision Research*, 35, 239–254.

Conway, B. (2001) Spatial structure of cone inputs to color cells in alert macaque primary visual cortex (V1). *Journal of Neuroscience*, 21, 2768-2783.

Conway, B. (2009) Color vision, cones, and color-coding in the cortex. *The Neuroscientist*, 15, 274-290.

Conway, B. R., Moeller, S., & Tsao, D. Y. (2007). Specialized Color Modules in Macaque Extrastriate Cortex. *Neuron,* (3).

Conway, B. R., & Tsao, D. Y. (2009). Color-tuned neurons are spatially clustered according to color preference within alert macaque posterior inferior temporal cortex. *Proceedings of the National Academy of Sciences,* (42).

Crichton, S., Pichat, J., Mackiewicz, M., Tian, G., Hurlbert, A. (2012) Skin chromaticity gamuts for illumination recovery, Conference on Color in Graphics Imaging, and Vision, pp. 266-271.

Desimone, R. & Schein, S.J. (1987) Visual properties of neurons in area V4 of the macaque: sensitivity to stimulus form. *Journal of Neurophysiology*, 57, 835-868.

D'Zmura M, Lennie P, (1986) Mechanism of colour constancy. *Journal of the Optical Society of America A*, 10 1662–1672.

D'Zmura, M. (1992). Color constancy: Surface color from changing illumination. *Journal of the Optical Society of America A*, 9, 490-493.

Foster, D. H., Nascimento, S. M. C. (1994). Relational colour constancy from invariant cone-excitation ratios. *Proceedings of the Royal Society of Londong B Biological Sciences,* 257, 115-121.

Funt B.V., Drew, M.S., & Ho, J. (1991) Color constancy from mutual reflection. *International Journal of computer Vision.* 6, 5-24.

Gegenfurtner, K. R., & Kiper, D. C. (2003). Color vision. *Annual Review of Neuroscience*, 26, 181-206.

Granzier J, Gegenfurtner K (2012) Effects of memory colour on colour constancy for unknown coloured objects. *iPerception* 3.

Hadjikhani, N., Dale, A., Liu, A., Cavanagh, P., & Tootell, R. (1998). FMRI of Retinotopy and Color Sensitivity in Human Visual Cortex. *NeuroImage,* (4).

Hedrich, M., Bloj, M., & Ruppertsberg, A. I. (2009) Color constancy improves for real 3D objects. *Journal of Vision*, 16, 1-16.

Helson, H. (1964) *Adaptation-Level Theory.* New York: Harper & Row.

Hering, E. (1878) *Grundzüge der Lehre vom Lichtsinn.* Berlin: Springer.

Hurlbert, A. (2007) Colour constancy. *Current Biology*, 17, R906-907.

Kingdom F. A. A. (2008) Perceiving light versus material. *Vision Research*, 48, 2090–2105.

Koida, K., & Komatsu, H. (2006). Effects of task demands on the responses of color-selective neurons in the inferior temporal cortex. *Nature Neuroscience,* (1).

Kraft, J. M., & Brainard, D. H. (1999) Mechanisms of colour constancy under nearly natural viewing. *Proceedings of the National Academy of Sciences USA*, 96, 307–312

Lafer-Sousa, R. & Conway, B. (2013) Parallel, multi-stage processing of colors, faces, and shapes in macaque inferior temporal cortex. *Nature Neuroscience*, 16, 1870-1878.

Lee, H. -C. (1986). Method for computing the sceneilluminant chromaticity from specular highlights. *Journal of the Optical Society of America A*, 3, 1694-1699.

Li, M., Liu, F., Juusola, M., & Tang, S. (2014) Perceptual color map in macaque visual area V4. *Journal of Neuroscience*, 345, 202-217.

Ling, Y., & Hurlbert, A. (2006). Colour-memory-dependent colour constancy: 2D vs. 3D real surfaces. Paper presented at the CGIV 2006, Leeds, UK.

Ling, Y., & Hurlbert, A. (2008). Role of color memory in successive color constancy. *Journal of the Optical Society of America A*, 25, 1215–1226.

Maloney, L. T., & Wandell, B. A. (1986). Color constancy: A method for recovering surface spectral reflectance. *Journal of the Optical Society of America A*, 3, 29-33.

Maloney L (2002) Illuminant estimation as cue combination. *Journal of Vision* 2, 493-504.

Marr, D. (1980) *Vision*. Cambridge: MIT Press.

Mckeefry, D. (1997). The position and topography of the human colour centre as revealed by functional magnetic resonance imaging. *Brain,* (12).

Meadows, J. C. (1974). Disturbed Perception Of Colours Associated With Localized Cerebral Lesions. *Brain, 97*(1).

Mitterer, H. & De Ruiter, J.P. (2008) Recalibrating color categories using world knowledge. *Psychological Science*, 19, 629-634.

Orban, G., Essen, D., & Vanduffel, W. (2004) Comparative mapping of higher visual areas in monkeys and humans. *Trends in Cognitive Sciences,* 8, 315-324.

Roe, A.W., Chelazzi, L., Connor, C.E., Conway, B.R., Fujita, I., Gallant, J.L., Lu, H., & Vandeuffel, W. (2012) Toward a univied theory of visual area V4. *Neuron*, 74, 12-29.

Rosenthal, I., Ratnasingam, S., Haile, T., Eastman, S., Fuller-Deets, J., & Conway, B. R. (2018). Color statistics of objects, and color tuning of object cortex in macaque monkey. *Journal of Vision,* (11).

Schein, S., & Desimone, R. (1990). Spectral properties of V4 neurons in the macaque. *The Journal of Neuroscience,* (10).

Scocchia, L., Valsecchi, M., & Triesch, J. (2014) Top-down influences on ambiguous perception: the role of stable and transient states of the observer. *Frontiers in Human Neuroscience*, 8, 979.

Simmons, W. K., Ramjee, V., Beauchamp, M. S., Mcrae, K., Martin, A., & Barsalou, L. W. (2007). A common neural substrate for perceiving and knowing about color. *Neuropsychologia,* (12).

Stiles, W. S. (1959) Color vision: the approach through increment threshold sensitivity. *Proceedings of the National Academy of Sciences USA*, 45, 100-114.

Takechi, H., Onoe, H., Shizuno, H., Yoshikawa, E., Sadato, N., Tsukada, H., & Watanabe, Y. (1997). Mapping of cortical areas involved in color vision in non-human primates. *Neuroscience Letters,* (1).

Tootell, R. B. (2004). Search for Color 'Center(s)' in Macaque Visual Cortex. *Cerebral Cortex,* (4).

Tsao, D.Y. & Livingstone, M.S. (2009) Mechanisms of face perception, *Annual Review of Neuroscience*, 31, 411-437.

von Kries, J. (1878) Beitrag zur Physiologie der Gesichtsempfindugen. In: MacAdam DL, editor. Sources of color science,1st edn. Cambridge, MA: MIT press, 101–108.

Webster, M. A. & Mollon, J. D. (1995). Colour constancy influenced by contrast adaptation. *Nature*, 373, 694–698.

Webster, M.A. & Mollon, J.D. (1997) Adaptation and the color statistics of natural images. *Vision Research*, 37, 3283-3298.

Witzel, C., Valkova, H., Hansen, T., & Gegenfurtner KR. (2011) Object knowledge modulates colour appearance. *iPerceptio*n. 2, 13.

Yang, J. N., & Maloney, L. T. (2001). Illuminant cues in surface color perception: Tests of three candidate cues. *Vision Research*, 41, 2581-2600.

Chapter 2: Color-biased Regions of the Ventral Visual Pathway Lie between Face- and Place-Selective Regions in Humans, as in Macaques

The contents of this chapter have been published:
Lafer-Sousa, R., Conway, B.R., and Kanwisher, N.K. Color-biased cortex lies between Face-Selective and Place-selective cortex in humans, as in Macaque monkeys. *J Neurosci* *36*(5): 1682-1697 (2016)

2.1. Abstract

The existence of color-processing regions in the human ventral visual pathway (VVP) has long been known from patient and imaging studies. But their location in the cortex relative to other regions, their selectivity for color compared with other properties (shape and object category), and their relationship to color-processing regions found in non-human primates remains unclear. We addressed these questions by scanning 13 subjects with fMRI while they viewed two versions of movies (colored, achromatic) of five different object classes (faces, scenes, bodies, objects, scrambled objects). We identified regions in each subject that were selective for color, faces, places, and object shape, and measured responses within these regions to the ten conditions in independently acquired data. We report two key findings. First, the three previously reported color-biased regions (located within a band running posterior-anterior along the VVP, present in most of our subjects) were sandwiched between face-selective cortex and place-selective cortex, forming parallel bands of face, color, and place selectivity that tracked the fusiform gyrus/collateral sulcus. Second, the posterior color-biased regions showed little or no selectivity for object shape or for particular stimulus categories, and showed no interaction of color preference with stimulus category, suggesting that they code color independent of shape or stimulus category; conversely, shape-biased region LO showed no significant color bias. These observations mirror results in macaque inferior temporal cortex (Lafer-Sousa and Conway, 2013), suggesting a homology in which the entire tripartite face-color-place system of primates migrated onto the ventral surface in humans over the course of evolution.

2.2. Introduction

Color is a fundamental dimension of visual experience, informing us about myriad facets of our environment. Befitting its importance for human vision, considerable evidence from patient studies (Meadows, 1974; Newcombe and Ratcliff, 1975; Heywood et al., 1987, 1991; Zeki, 1990; Bouvier and Engel, 2006; Nijboer et al., 2007) and human neuroimaging (McKeefry and Zeki, 1997; Hadjikhani et al., 1998; Beauchamp et al., 1999; Wade et al., 2002, 2008;

Brewer et al., 2005; Simmons et al., 2007; Cavina-Pratesi et al., 2010) implicates specific cortical regions of the ventral visual pathway (VVP) in the analysis of color. But little is known about the causal role of these regions in perception (Murphey et al., 2008; Rangarajan et al., 2014), their connectivity, the precise computations conducted within them, or the underlying neural circuits. The most powerful methods for answering these questions require a primate model. Such a model would be most informative if the processing mechanisms were homologous to those found in humans. How similar then are the cortical mechanisms for color vision in non-human primates and humans?

Macaques and humans have very similar color behavior (Stoughton et al., 2012; Gagin et al., 2014), raising the possibility that the two species have similar cortical machinery for color vision. Here we test this hypothesis by asking whether humans show the same organization for color in the VVP as found in macaques, where color-biased regions are sandwiched between face and place selective regions (Lafer-Sousa and Conway, 2013; Verhoef et al., 2015). No prior study has directly assessed the topographic relationship of these regions in humans. A strong test of the idea requires a systematic and quantitative evaluation of category, shape, and color selectivity, in the same subjects, across human visual cortex. To do this, we scanned 13 human subjects using functional magnetic resonance imaging (fMRI) while they viewed either full-color or achromatic movie clips of faces, objects, scenes, bodies, or scrambled objects. We first determined the spatial layout of color, face, and place selectivity by projecting color, face, and place preferences onto each subject's inflated cortical surface. We then defined color, face, and place preferring regions within each subject and measured the response magnitude of each region to each of the ten conditions in independently acquired data. These analyses enabled us to map the relative locations of shape-selective, category-selective, and color-biased regions in each subject individually, and to quantify the selectivity of each region to each dimension.

Our experimental design enabled us to address a second open question: Is the processing of color and shape inextricably linked throughout the VVP? On the one hand, one might think of color as just another object property, likely processed in the same cortical regions (and neurons) as those implicated in computing object shape. On the other hand, the computations entailed in processing color, and the role of color in behavior, often diverge from those for shape (Conway, 2009, 2014). Some micro-electrode studies in monkeys have been interpreted as supporting the idea that color and shape are processed together (Schein and Desimone, 1990; Gallant et al., 2000; Gegenfurtner and Kiper, 2003; Shapley and Hawken, 2010; Yasuda et al., 2010), but other results in macaques (Zeki, 1980, 1983; Bartels and Zeki, 2000; Conway, 2001; Lafer-Sousa and Conway, 2013; Conway, 2014) and humans (Newcombe and Ratcliff, 1975; Milner and Heywood, 1989; Heywood et al., 1991; Bouvier and Engel, 2006; Cavina-Pratesi et al., 2010) suggest some dissociation of color and shape processing. We addressed this question by determining the extent to which regions (e.g. LO) that responded preferentially to intact shape (compared to scrambled counterparts) were modulated by the presence of color, and the extent to which regions responding preferentially to color were modulated by the presence of intact shape. These experiments enabled us to test the degree to which color, category and shape preferences are segregated in the human brain, and to investigate possible homologies of cortical regions in humans and macaques.

2.3. Methods

Human Participants

Thirteen subjects (age 20-39, 7 female) participated in the study. Participants had no history of neurological or psychiatric impairment, had normal or corrected vision, normal color vision (tested with Ishihara plates), and were native English speakers. All participants provided written, informed consent.

Stimuli: Dynamic Localizer

To localize category, shape, and color-biased activity we used functional magnetic resonance imaging (fMRI) to scan subjects while they viewed natural video clips corresponding to one of five stimulus classes (Faces, Bodies, Scenes, Objects, and Scrambled Objects) presented in either full color (chromatic; the same stimuli used in Pitcher et al., 2011) or gray scale (achromatic) (Fig. 1a). Each run consisted of twenty-five 18-second long blocks (20 stimulus blocks and 5 full-field gray fixation blocks to allow the signal to return to baseline). Each stimulus block contained six three-second long video clips randomly drawn from 60 clips of a specific stimulus condition (e.g, chromatic faces). In a single scanning session, subjects viewed 16-20 blocks of each of the ten stimulus conditions (five classes x chromatic/achromatic), or a total of 32-40 blocks for each class (e.g, 32-40 blocks of faces, comprising 16-20 chromatic and 16-20 achromatic); 80-100 chromatic blocks (comprising all classes of chromatic stimuli) and 80-100 achromatic blocks. The stimuli were presented in the central 20 degrees of visual field. The order of conditions was palindromic (e.g. A-B-C-C-B-A) within a run and was counterbalanced across runs such that each condition happened equally often in each serial position in the run (10 run orders). The Matlab function rgb2gray was used to render the clips in gray-scale while retaining luminance structure (confirmed with photometric measurements).

Drifting Gratings

In control experiments we localized color responses using drifting grating stimuli comparable to those used in prior work (Lafer-Sousa and Conway, 2013); these stimuli allow more stringent control of the physical parameters of the stimulus but contain no recognizable objects. The chromatic gratings were designed to be equiluminant; the achromatic gratings were designed to comprise some luminance contrast. The stimulus paradigm was similar to that used to identify color-biased regions in macaque (Lafer-Sousa and Conway, 2013). Three achromatic gratings conditions (luminance contrasts 50%, 25%, and 7%) and eight color directions defined by the cardinal and intermediate directions of the equiluminant plane in cone-opponent color space (Derrington et al., 1984; Macleod and Boynton, 1979) were used. Block-lengths were shortened compared to those used in the macaque experiments (18 seconds instead of 32 seconds) to match the hemodynamic response function of the BOLD signal in humans (block lengths needed to be longer in the macaque experiments to account for the longer hemodynamic delay that accompanies the use of an intravenous contrast agent). Stimuli were calibrated using spectral readings taken with a PR-655 spectroradiometer (Photo Research). The spectra were multiplied with the Judd-revised Commission Internationale de l'Eclairage (CIE) 1931 color matching functions to derive CIE xyY coordinates of the monitor primaries (Hansen et al., 2008), and cone excitation was calculated using the Smith and Pokorny cone fundamentals (Smith and Pokorny, 1975). Stimuli were presented as vertical trapezoid-wave gratings (Conway and Tsao, 2006). Each chromatic and achromatic grating was drifted back and forth for 18 seconds, switching directions every 2 s (2.9 cycles per degree, drifting 0.75 cycles per s). Within a run, the grating blocks were interleaved with equiluminant neutral full-field gray (100 cd/m2) (12-s) to allow the signal to return to baseline [e.g. Gray, Achromatic-25%, Gray, Color-1, Gray, Achromatic-50%, Gray, Color-8, Gray, etc.] There were 9 run orders, counterbalanced across runs such that each condition happened equally often in each serial position in the run. Each achromatic grating appeared three times per run, and was shown between chromatic blocks (with neutral gray blocks interleaved). Subjects were required to fixate a central fixation cross and carried out a difficult motion-detection task to maintain attention: the gratings drifted left to right, switching direction every 2s except once per block, when the grating would drift 3s in one direction. Subjects had to report when this occurred (via a button press in the scanner).

Retinotopic mapping

In 9 subjects, we identified the borders of visual areas V1 through V4 by mapping the responses to horizontal and vertical meridians following standard procedures (Sereno et al., 1995). The stimuli were flickering (counterphase at 7.5 Hz) checkerboards restricted to wedges along either the horizontal or vertical meridian, spanning11.8 deg radius and subtending 8.2 deg. Blocks of checkerboards were interleaved with neutral full-field gray. Each run comprised 36 ten-second blocks. Subjects were required to fixate throughout. Responses to vertical meridians were contrasted with responses to horizontal meridians and projected on each subject's cortical surface, producing a striped map used to demarcate the boundaries of V1, V2, V3 and V4. Because the anterior border of V4 (between V4/hV4 and VO-1) is difficult to define functionally (McKeefry and Zeki, 1997; Bartels and Zeki, 2000; Tootell and Hadjikhani, 2001; Brewer et al., 2005; Winawer et al., 2010), its location was determined using the anatomical landmark ptCoS (Winawer et al., 2010).

Data acquisition

Data were acquired using a Siemens 3T MAGNETOM Tim Trio scanner (Siemens AG, Healthcare, Erlangen, Germany) with a 32-channel head coil (at the Athinoula A. Martinos

Imaging Center of the McGovern Institute for Brain Research at MIT). Functional data were collected using a T2*-weighted echo planar imaging (EPI) pulse sequence sensitive to blood-oxygen-level-dependent (BOLD) contrast. Given the known susceptibility artifact of the anterior regions of the ventral temporal lobe (caused by the ear canals), we carried out extensive piloting to determine the parameters that produced the highest voxel SNR. Pilot data were collected with and without iPAT (Griswold et al., 2002) and field correction; we varied the voxel-resolution, slice angle, TE, and fraction of k-space sampled. We obtained improved tSNR and BOLD sensitivity by increasing voxel resolution by a factor of 4 (from the standard 3 mm iso to 2 mm iso), forgoing image acceleration (iPAT), and collecting field maps prior to each run. Reducing TE (to 14 or 23 msec) enhanced the tSNR, but compromised BOLD sensitivity. TE was set30 msec.

The functional volumes were acquired for a restricted portion of the brain spanning the ventral surface of the temporal lobe. Each functional volume comprised 25 slices (2 mm isotropic; field of view [FOV] = 192mm, matrix = 96x96 mm) oriented roughly parallel to the temporal lobe, covering the occipital cortex (V1-V4) and the full length of the temporal lobe ventral to and including some of the superior temporal sulcus (2.0 s TR, 30 ms TE, 90 degree flip angle, 6/8 echo fraction). The first 5 volumes of each run were discarded to allow for T1 equilibration. Field maps (2 mm isotropic, 25 slices) were collected prior to each functional run to measure magnetic inhomogeneities and used to estimate spatial distortions in the functional volumes that were then removed during analysis. High-resolution T1-weighted anatomical images were collected for each subject using a multi-echo MPRAGE pulse sequence (1 mm isotropic voxels; FOV = 256mm, matrix = 256x256).

Data preprocessing and modelings

Data were processed using Freesurfer (http://surfer.nmr.mgh.harvard.edu/) and custom MATLAB scripts. Freesurfer was used to segment white and grey-matter structures from the anatomical volumes (Dale et al., 1999; Fischl et al., 1999, 2001). Individual subject's functional data were field-corrected, motion corrected using rigid-body transformations to the middle image of each run, intensity normalized (non-brain tissue was masked), spatially smoothed using an isotropic Gaussian kernel (3mm FWHM) to improve SNR, and aligned to the individual's anatomical volume using a rigid-body transformation determined by Freesurfer's bbregister (Greve and Fischl, 2009). Digitally inflated (Figures 1, 2, 5, 6, 9) and flattened (Figures 2, 6, 7, 9) cortical surface reconstructions were generated using freesurfer.

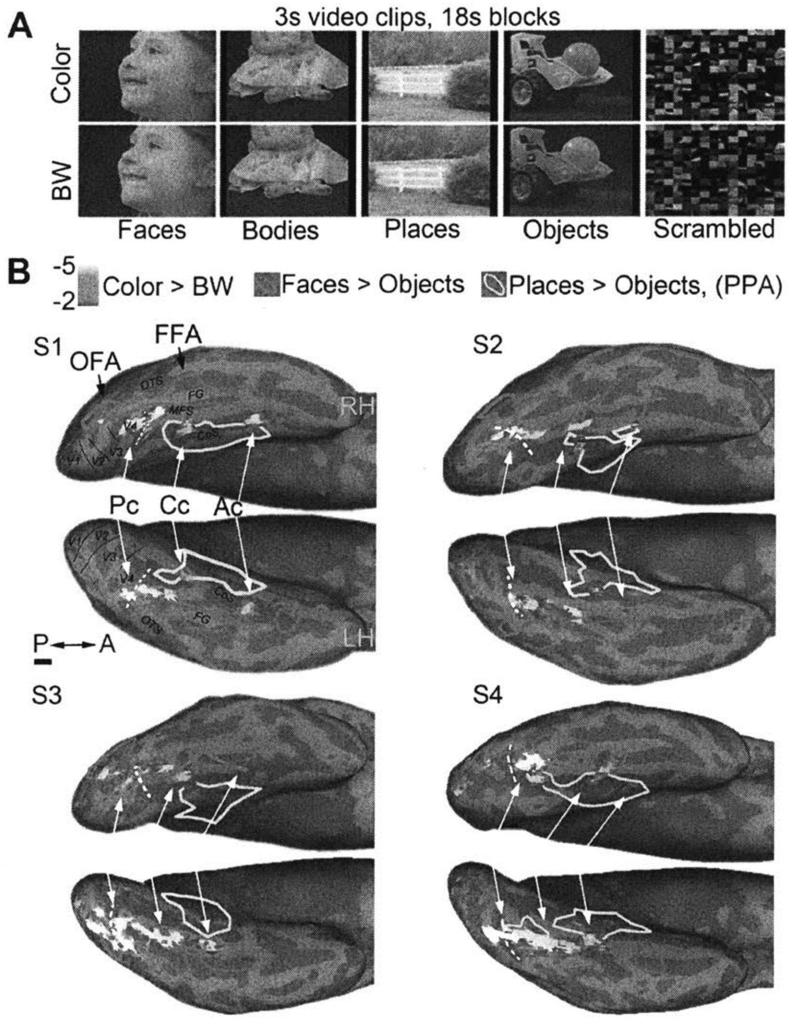


Figure 1. Sandwiching of color-biased regions between face-selective cortex and place-selective regions. A, Frames from the video clips used to define functional regions of interest. Each stimulus block comprised video clips from one stimulus category (Faces, Bodies, Scenes, Objects, and Scrambled Objects) presented in color or luminance-matched grayscale (six 3-second clips/block). **B,** Color-biased activity on the ventral cortical surface in four subjects (S1-S4; color scale shows exponent of the p level for the chromatic>achromatic contrast defined using video clips from all categories). Color-biased activity in the ventral visual pathway was found between the more lateral face-selective regions (faces > objects contrast, $p < 0.0001$, purple outlines) and the more medial/ventral place-selective regions (scenes > objects, $p < 0.0001$, green outlines). Color-biased activity spans much of the posterior-to-anterior extent of the VVP. Three peaks in the color-biased activity were defined in most subjects: Pc (posterior color region, called elsewhere V4 or hV4), Cc (central color, V4α/V8/VO-complex), and Ac (anterior color). The white dashed line denotes the posterior transverse collateral sulcus (ptCoS), which runs transversely to the base of the collateral sulcus (CoS), and defines the border between V4 and the ventral occipital complex VO (Winawer et al., 2010; Witthoft et al., 2013). The boundaries of early visual areas as defined by standard meridian mapping (vertical/horizontal reversals) are shown for S1. Scale bar represents 1 cm.

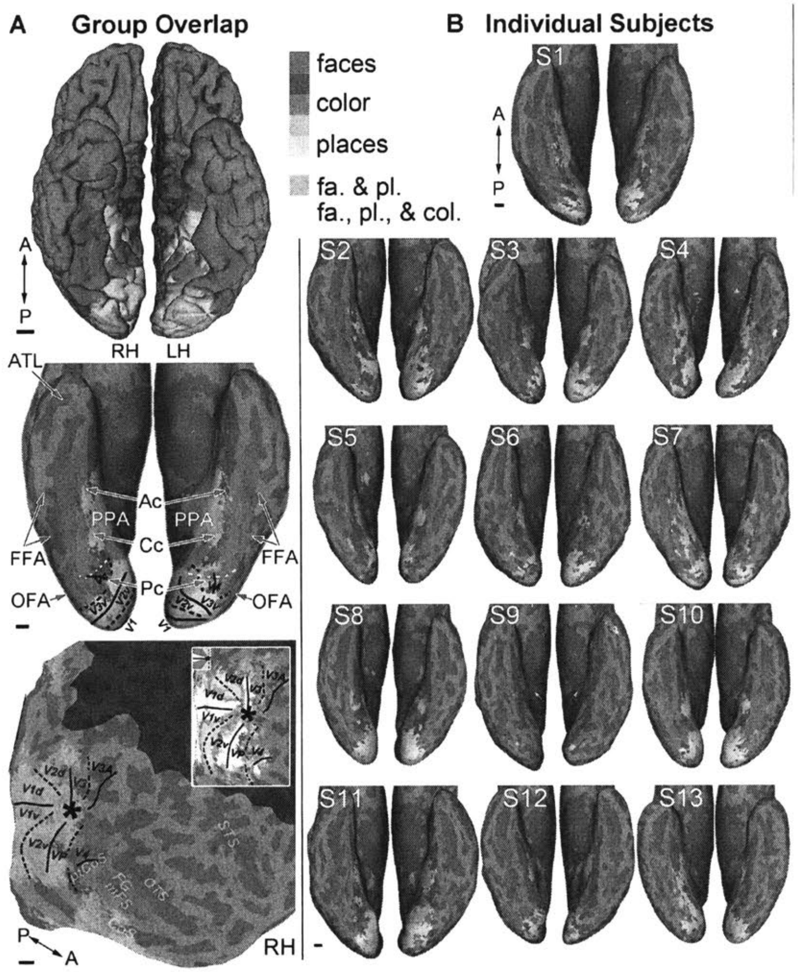


Figure 2. Relative cortical location of face, color, and place activations. A, Group overlap maps. Data from each subject (N=13) was registered to a published group-average template. Individual contrast maps were thresholded ($p<0.01$ uncorrected) and tallied, such that each resultant voxel represented the number of subjects who had above-threshold activity at that location. Voxels for which at least 5 subjects had above threshold activity are colored: face-selective voxels (purple), color-biased (blue), and place-selective (green); dark purple shows overlap of face and color; cyan shows overlap of color and place; orange shows overlap of face and place; yellow shows overlap of face, color, and place. Top panel shows the pial surface (ventral view); middle panel shows the digitally inflated surface (prefrontal cortex excluded); bottom panel shows the digitally flattened surface (RH, right hemisphere; dark gray mask overlays regions outside functional scans). The white dashed line denotes the anterior border of V4 (ptCoS). Horizontal (solid lines) and vertical (dashed lines) meridians were obtained by mapping meridian representations; colors in the flat map inset (bottom) show voxels in which at least 5 subjects had above threshold activity (yellow, horizontal meridian stimulation; cyan, vertical meridian stimulation; $p<0.001$). Asterisk indicates the foveal confluence. **B,** Thresholded contrast maps for each individual subject (same color scheme as in **A**). P-thresholds were set at the levels used to define ROIs in the volume, (color>achromatic, $p < 0.001$; faces > objects, $p < 0.0001$; scenes >objects, $p <0.0001$). In each subject, color-biased activations were sandwiched between face-selective and place-selective regions, with face-selective regions lateral to color-biased cortex, and place-selective cortex medial. Scale bars represent 1 cm.

Whole-volume general linear model (GLM)-based analyses were performed for each subject's even and odd runs separately. Regressors were defined as boxcar functions convolved with a gamma hemodynamic response function (Friston et al., 1994). The boxcar function for each condition included each block from that condition. Nuisance regressors for motion (three translations, three rotations) and a linear trend to capture slow drifts were included.

Analysis of Color, Face, Place, and Shape preferences across cortex (Dynamic Localizer Data)

To show the large-scale spatial organization of the ventral visual pathway within each individual, subjects' statistical contrast maps were projected on the digitally inflated surface of their own cortical anatomy. Following standard practice (Julian et al., 2012), functional regions were defined as follows. Face-selective voxels were defined as those showing higher responses to video clips of faces than objects ($p < 0.0001$; see purple regions in Figure 1B and Figure 2B). Place-selective voxels were defined as those showing a higher response to places than objects ($p < 0.0001$; green regions in Figure 1B and Figure 2B). Color-biased voxels were defined as those showing a higher response to chromatic video clips than achromatic versions of the same clips matched in luminance, averaging responses obtained in all stimulus classes ($p <0.001$; activation maps shown in Figure 1B; blue regions in Figure 2B). The conclusions were not affected if color-biased regions were only defined using movie clips of a single category (data not shown). Shape-selective voxels were defined as those showing a higher response to intact objects than scrambled objects ($p < 0.0001$; see Figure 5).

The major conclusions presented here are derived from an analysis of individual subject data. But to illustrate overall trends in the spatial layout of color-biased regions relative to face, place, and shape-selective regions, we generated group overlap maps. To do this, statistical contrast maps from each of the thirteen individual subjects were registered to a published group-average template (Freesurfer, CVS_avg35, an average of 35 people). The group-registered contrast maps from each individual were then thresholded and each voxel was assigned a binary value (1 or 0—above threshold or not). For a given contrast, all 13 subject maps were tallied and the voxels in the resulting volume were assigned a value according to the number of subjects who had above-threshold activity at that location. Figure 2A shows the group summary map for face, color, and place contrasts, thresholded at 5 subjects (individual contrast maps thresholded at $p < 0.01$), on the reconstructed cortical surface of the CVS_avg35 volume. Figure 5A shows the group map for the shape contrast (intact objects > scrambled objects; $p < 0.0001$) as a surface histogram where voxel color (red to yellow) reflects the agreement of 2 to 5+ subjects.

The surface maps enabled us to examine the spatial segregation of the activation patterns for color, faces, and places. But region of interest (ROI)-based analyses make possible stronger tests of our hypotheses. Specifically, they enabled us to cross-validate the contrast by which each region was defined, and to quantify the effect size and selectivity of that contrast. Effect size matters: a region may have highly significant selectivity but of small magnitude, an important distinction that is not apparent in standard activation maps showing only significance levels. By quantifying activity in ROIs, we were able to determine the profile of response of each region to each of the ten conditions tested, providing a richer profile of the selectivity of each region. From these results, we ask if the selectivity for color is present to a similar degree for each stimulus category, and the extent to which it depends on stimulus category (which would be manifested by an interaction of color/achromatic by category). Significant interactions between two factors (e.g. color and shape) may be evidence that the two factors depend upon common neural populations, rather than being analyzed independently in two distinct neural populations within that region (Sternberg, 1969). The ROI analyses also enabled us to test whether two regions differ significantly from each other with respect to some contrast. Standard activation maps do not support such conclusions because they only show regions that reach significance in a given contrast and regions that do not—that is, they show differences in significance, but not significant differences across regions. For example, in one ROI analysis we explicitly test if color-biased regions are significantly more color-biased than category-selective regions, and shape-selective regions, contrasting the responses in these regions to each other in an ANOVA with ROI as a factor (described below).

Functional ROIs were defined in individual subjects, in the volume, using data from even runs. The odd runs were used to quantify the responses within each ROI to each of the ten conditions. ROIs were defined using a p-threshold criterion. For a given contrast, all contiguous voxels above the p-threshold constituted an ROI; p-threshold was $p<0.0001$ for OFA, FFA, ATL, PPA, LO, PFS, and pSTS; and $p<0.001$ for color-biased regions (Pc, Cc, and Ac). A less stringent p-threshold was used to define color-biased voxels because color selectivity was weaker overall than shape and category selectivity. Note that the evidence for color selectivity comes not from the statistics used to define the region, but from the statistics of the response magnitudes conducted using independently acquired data (the odd runs). In some subjects, face activations were so robust they formed a contiguous swath extending from the OFA to the FFA; similarly, shape-selective regions sometimes formed a contiguous swath from LO to PFS; and place activations sometimes extended contiguously into early visual cortex. In these cases a published parcel atlas derived from 30 human subjects was used to constrain ROI definition for the separate components within each swath of activation (Julian et al., 2012). The main conclusions regarding the segregation of face, color and place activations do not change if ROIs were defined solely using functional criteria (in this case, for example, lumping the OFA and FFA together).

For each subject, the percent signal change within each ROI (PSC; see Figure 3) was extracted by averaging beta values across the ROI and dividing by the mean BOLD signal in the ROI. Each ROI was detected in most subjects and when a region appeared bilaterally (typical for most subjects and regions, see Results), the results across hemispheres were averaged. Participants who lacked a certain ROI were not included in the statistical analyses (ANOVAs) involving that ROI; if the selectivity of an ROI present in only 10 subjects was being compared to an ROI present in 13 subjects, only the 10 subjects that had both ROIs were included in the analysis. Statistical tests (ANOVAS) for main effects and interactions are described in the results.

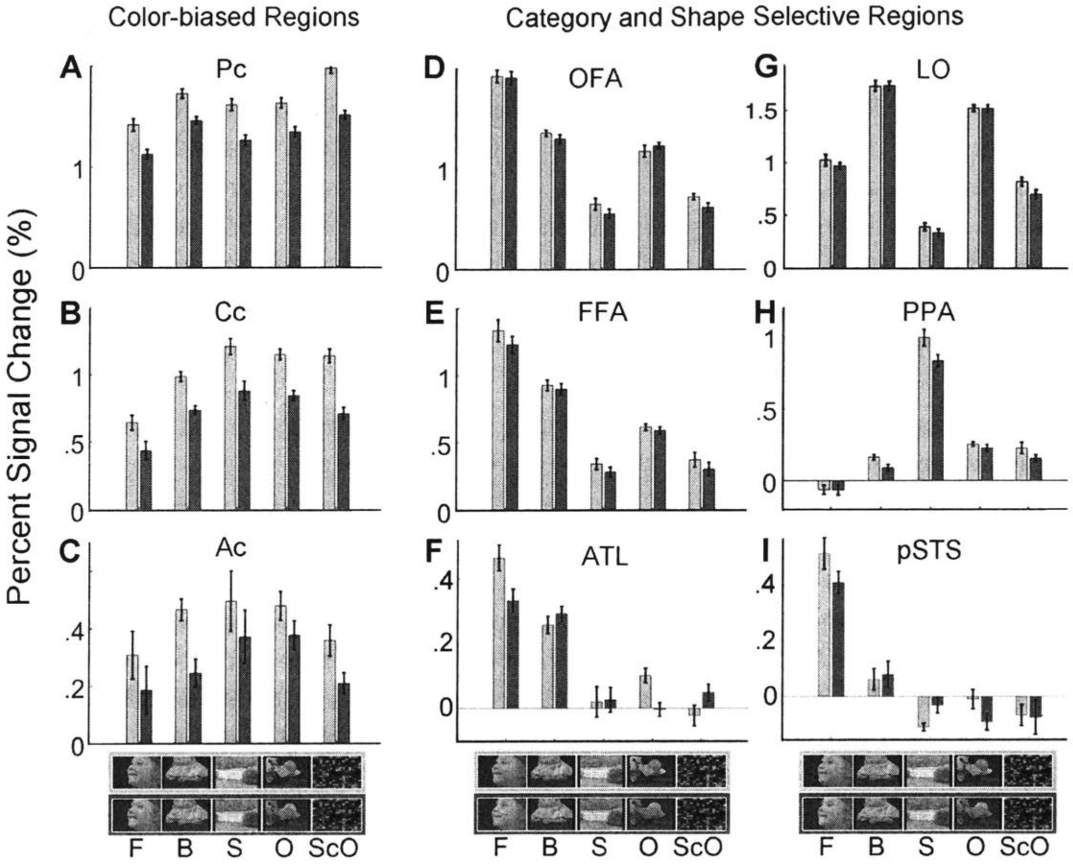


Figure 3. Segregation of color preference and category selectivity in the ventral visual pathway. To quantify the response profiles of specific regions of interest (ROIs), category-selective (OFA, FFA, ATL, PPA) and color-biased (Pc, Cc, and Ac) ROIs were defined in each subject individually using data from even runs, and the response magnitude of each region to each of the ten conditions was quantified using the independent data from odd runs (panels **A-F, H**). For comparison, we also defined a region that responded preferentially to intact objects (defined by shape selectivity not category selectivity) that is adjacent to Pc and OFA (region LO, panel **G**), and a face-selective region (pSTS, panel **I**) on the lateral surface (as opposed to the ventral surface). ROIs were defined using a p-threshold criterion (for the OFA, FFA, ATL, PPA, pSTS, and LO, $p < 0.0001$; for the color-biased regions, $p < 0.001$). For each subject, an ROI's percent signal change (PSC) was extracted by averaging beta values across the voxels in the ROI and dividing by mean BOLD signal in the ROI. When a region appeared bilaterally, the PSC for the left hemisphere ROI was averaged with the PSC of the corresponding right hemisphere ROI for each condition. Bar plots show the average PSC across subjects for each ROI (error bars show within subject standard error). Participants who lacked a particular ROI were not included in the statistical analysis for that ROI. Cyan bars indicate chromatic conditions and gray bars indicate achromatic conditions (F: Faces, B: Bodies, S: Scenes, O: Objects, ScO: Scrambled Objects).

As described in the results, the color activation along the posterior-anterior extent of the VVP was not homogeneous; three main peaks in the activation were defined (a posterior, central and more anterior peak; Pc, Cc, and Ac). To quantify this inhomogeneity, in each subject we defined two ROIs (Figure 4B schematic, black circles) between the peak color-biased activations (white circles): IPCc (intermediate to Pc and Cc) and ICAc (intermediate to Cc and Ac). These

ROIs were defined in each subject individually in the volume using even runs, as the non-significant voxels in the contrast used to define the color-biased regions. Figure 4B bar plots show the PSCs for colored and achromatic stimuli, averaging across faces, bodies, scenes, objects, and scrambled objects.

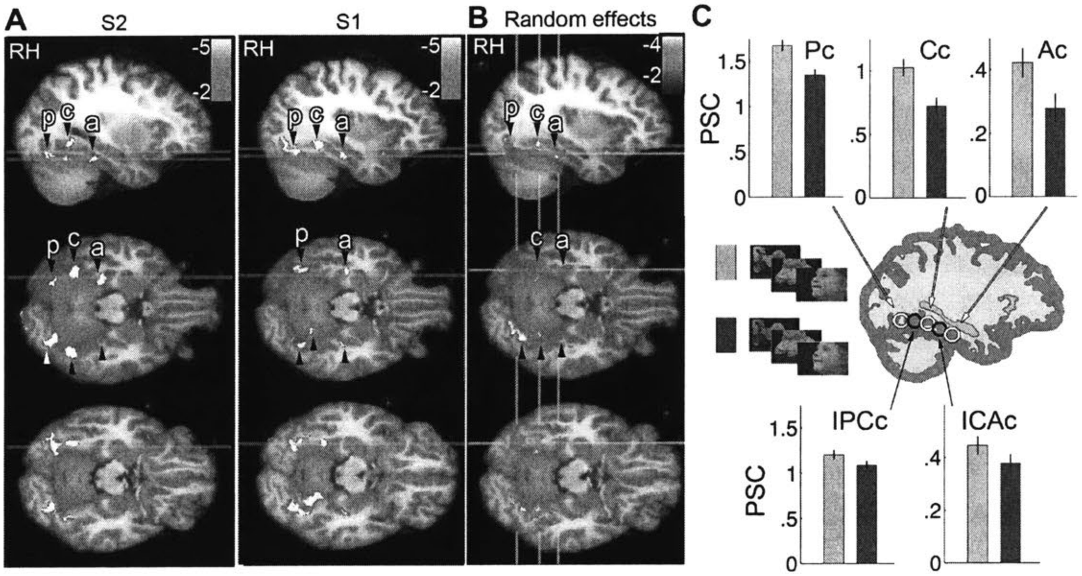


Figure 4. Quantification of the color bias within color-biased ROIs and intervening tissue. A, Individual functional volumes (chromatic > achromatic) for two example subjects (displayed on CVS_35 group anatomical). Top row, sagittal sections through the right hemisphere; middle and bottom rows, horizontal sections at the planes indicated by the yellow lines in the sagittal sections (yellow lines in horizontal sections show the location of the sagittal section). **B,** a random-effects analysis (N = 13); other conventions as for **A**. **C,** Quantification of inhomogeneity along the posterior-anterior band of color activations. Two ROIs (schematic, black circles) spatially intermediate to the color-biased regions (white circles) were defined in each subject: IPCc (intermediate to Pc and Cc) and ICAc (Intermediate to Cc and Ac). Bar plots show each region's mean percent signal change for chromatic stimuli (cyan bar) and achromatic stimuli (gray bar) (pooled across faces, bodies, scenes, objects, and scrambled objects) (error bars show within subject standard error). Color-biased regions (top row) were more color-biased than intermediate regions (bottom row) (p values < 0.003, pairwise 3-way ANOVAs on ROI x color x stimulus class, each contrasting an intermediate region with one of its neighboring color-biased regions).

In order to plot color-bias as a function of shape-bias (see Figure 5C), we used an index of color selectivity (response to color – response to achromatic / response to color + response to achromatic), and a comparable index of shape selectivity (response to intact objects – response to scrambled objects / response to intact objects + response to scrambled objects). The response to color was the average PSC to chromatic intact objects and chromatic scrambled conditions, and the response to achromatic was the average PSC to achromatic-intact and achromatic-scrambled conditions. The response intact objects was the average PSC to chromatic and achromatic intact object conditions; the response to scrambled objects was the average PSC to chromatic and achromatic scrambled conditions. These indices were computed at the group level to avoid complications posed by dividing by noisy negative individual responses. Error bars were calculated using bootstrapping across subjects and represent 95% confidence intervals (10,000 bootstrapped samples) (Efron, 1982).

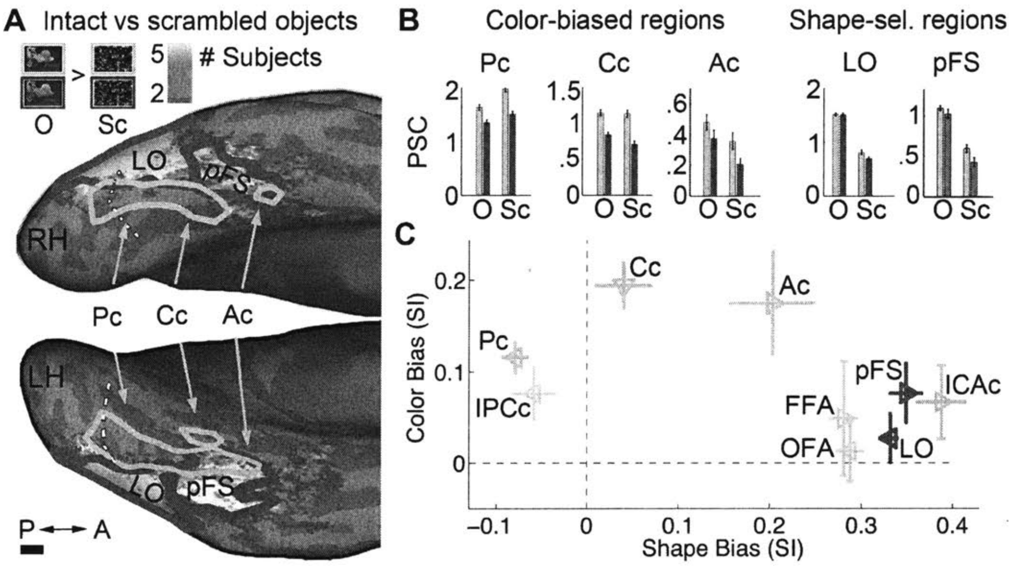


Figure 5. Relationship of shape-biased regions to color-biased regions. A, Group overlap map; color scale reflects number of subjects that showed a significant preference for intact objects ("O") over scrambled objects ("ScO") (p-threshold = 0.0001, N =13 subjects). Light-blue outlines show color-biased regions (from Figure 2a); navy blue outlines show published group parcels for shape-selective regions LO and PFS (Julian et al., 2012). Most of the large swath of shape-biased cortex avoids color-biased cortex, except at the most anterior location. Scale bar,1 cm. **B,** Bar charts quantifying the response of color-biased and shape-biased regions to chromatic (cyan bars) and achromatic (gray bars) intact objects ("O") and scrambled objects ("ScO"). ROIs were defined in individual subjects using data from even runs, and the response magnitude of each region to each condition was quantified using the independent data from odd runs. **C,** Scatter plot showing the extent of color bias and shape bias in color-biased regions (cyan triangles), shape regions (navy triangles), OFA, FFA, IPCc and ICAc (grey triangles). Selectivity indices for shape and color bias were determined for each ROI using the data from panel B. Bias was computed as $[PSC_{preferred} - PSC_{non\text{-}preferred}] / PSC_{preferred} + PSC_{non\text{-}preferred}]$. For color bias, *preferred* is chromatic (intact and scram.) and *non-preferred* is achromatic (intact and scram). For shape bias, *preferred* is intact (chrom. and achrom.) and *non-preferred* is scrambled (chrom. and achrom). Error bars,show 95% C.I. (10,000 bootstrapped samples).

Group Analysis (Random Effects) of Color-biased activations (Dynamic localizer)

In order to test the extent to which color activations were found at consistent locations in the brain across subjects (see Figure 4), we performed a group random-effects analysis on the chromatic versus achromatic contrast, in which data from each subject were registered and normalized to a common template (Freesurfer, CVS_avg35).

Analysis of inhomogeneity in tSNR across the cortical surface

The ability to detect significant differences in the BOLD response in each region of the cortex depends on the signal to noise ratio, which is known to vary across the brain due inhomogeneities in the magnetic field. To visualize this spatial variation in SNR, we computed temporal SNR (tSNR) for each subject as follows: the motion-corrected functional volumes were high-pass filtered with a cut-off of 0.004 Hz to remove slow drifts, and for each voxel we computed the mean of the time course for each voxel was divided by its standard deviation. To generate a mean tSNR map from the group of subjects, the tSNR map for each subject was registered to a common template (Freesurfer, CVS_avg35), and the volumes from all subjects were averaged.

Color localization with colored gratings

In addition to localizing color-biased cortex using a contrast of chromatic>achromatic movie clips, color-biased regions were also localized in a subset of subjects using a more controlled stimulus paradigm containing no recognizable objects (full field drifting gratings), comparable to the one used in an earlier study of macaque monkeys (Lafer-Sousa and Conway, 2013), and similar to those used in most prior work on color-localization in humans (McKeefry and Zeki, 1997; Hadjikhani et al., 1998; Wade et al., 2008). This enabled us to directly compare the human data to the previously published macaque data (obtained using gratings) and establish the extent to which the natural videos recover the same regions localized with standard low-level stimuli.

In the 3 subjects who participated in this control experiment, color-biased activations were defined by contrasting responses to the chromatic grating conditions that elicited the weakest response in MT+ (these colors would suffer the least luminance contamination) with the achromatic grating condition that yielded the same magnitude of response in V1 (50% luminance contrast). We quantified the PSC to gratings in each of the ROIs defined using the dynamic natural stimulus, as well as regions V1, V2, and MT.

Monkey data

If the homology between humans and monkeys hypothesized in the present report is true, color-biased regions in monkeys (Lafer-Sousa and Conway, 2013) and color-biased regions in humans (described presently) should show a similar pattern of color and shape selectivity when scanned using the same dynamic natural videos. In order to test this hypothesis, we scanned the two macaque subjects from Lafer-Sousa and Conway (2013) while the animals were shown the dynamic natural video stimuli, using the same acquisition and pre-processing methods described in the earlier report. MION contrast-agent was used in the monkey experiments. The stimulus block lengths for the dynamic localizer were extended (32 seconds) to account for the slower hemodynamic response function associated with MION (Vanduffel et al., 2001). The task was free-viewing, as in the human experiments; animals were rewarded with juice for looking anywhere on the screen. We extracted PSC responses to the dynamic natural videos using the

ROIs defined in the earlier report (independent data): the middle face patch (ML) and the set of color-biased ROIs (PLc, posterior lateral color, located in PIT; CLc, central lateral color, located in CIT; and ALc, Anterior Lateral color, in AIT). Bar charts (see Figure 8A) show the mean PSC of the last 7 TRs in each stimulus block. Scatter plot (see Figure 8B) shows the relative color and shape bias of each ROI (as computed for Figure 5B in the humans).

2.4. Results

Topography of Face, Color, and Place Responses.

We first asked to what extent the functional organization of the human ventral visual pathway (VVP) resembles that of macaque monkey, where parallel and adjacent processing systems have been reported for faces, places, and color, with color-biased cortex sandwiched between face-selective streams (superiorly) and place-selective regions (inferiorly) (Lafer-Sousa and Conway, 2013; Verhoef et al., 2015). To do so, thirteen subjects were scanned with fMRI while viewing short video clips containing multiple stimulus categories, each presented (in different blocks) in either full color or luminance-matched gray scale (Fig. 1A). Face-selective and place-selective regions were defined in each subject individually (contrasts were faces> objects and places > objects respectively), and are shown in Figure 1B (outlines) on the inflated cortical surface for four example subjects (voxel threshold of $p < 0.0001$). Face-selective regions (purple outlines, Figure 1B) and place-selective regions (green outlines, Figure 1B) were found in most subjects, typically bilaterally. The parahippocampal place area (PPA; Epstein and Kanwisher, 1998) and fusiform face area (FFA; Kanwisher et al., 1997) were detected in all 13 subjects and appeared bilaterally; the occipital face area (OFA; Pitcher et al., 2011) was detected in 12/13 subjects (11/12 bilaterally); and the anterior temporal lobe face area (ATL; Kriegeskorte et al., 2007; Collins and Olson, 2014) was detected in 9/13 subjects (3/9 bilaterally). The FFA frequently consisted of a posterior and anterior component (FFA-1/pFus-faces and FFA-2/mFus-faces; Weiner & Grill-Spector, 2010). In addition to the face-selective regions of the VVP, the dorsal face-selective region of the superior temporal sulcus (pSTS; Puce et al., 1998) was detected in 11/13 subjects (8/11 bilaterally). The scene-selective retrosplenial cortex (RSC; Epstein et al., 2007) and occipital place area (OPA; Julian et al., 2012) were outside of the region scanned in many subjects and were not analyzed.

Activation maps in Figure 1B show those voxels with greater activation to colored versus achromatic video clips. In each subject, the color-biased activation was sandwiched between the face-selective regions and the place-selective regions. The volume-wide contrast (smoothed 3mm gaussian fwhm) of the color-activation pattern showed a mottled band of color-biased activation running posterior-to-anterior along the VVP, within which we typically observed multiple peaks. In order to quantify the activity, we defined three color-biased regions of interest in most subjects (the interdigitating tissue is further examined in Section 3). In all subjects (11/13 subjects bilaterally) we observed a posterior color-biased region that we refer to as Pc (for

"Posterior color"; talairach coordinates: -32, -76, -7), which corresponds to an area originally referred to by Zeki as V4 (Zeki et al., 1991) and hV4 by others (Wade et al., 2002; Brewer et al., 2005). In most subjects this region extended beyond the V4 border (anteriorly) into ventral occipital cortex (VO-1) (Brewer et al., 2005). We observed a second color-biased region in all subjects (12/13 subjects bilaterally), about 1cm anterior to Pc, medial to the mid-fusiform sulcus (MFS), and often extending into the collateral sulcus (CoS). This region, referred to here as the "Central color" region (Cc; tal. coords.: -25, -54, -10) corresponds to V4α (Bartels and Zeki, 2000), V8 (Hadjikhani et al., 1998), and part of the VO-complex (Wade et al., 2002) . In 10/13 subjects we identified a third even more anterior region (called here "Anterior color" or Ac) in at least one hemisphere (bilaterally in 6 of these ten subjects), located anterior to the VO-complex (tal cords.: -32, -37, -8). This region falls in the neighborhood of activations that have previously only been observed in studies involving a demanding color-behavior task (Martin et al., 1995; Simmons et al., 2007), and most closely resembles the region described by Simmons et al (located in the left fusiform gyrus), hypothesized to be a "color-knowledge" region. Color-localization studies that use low-level stimuli and no color-perception task (e.g. standard Mondrians) do not report this region (McKeefry and Zeki, 1997; Hadjikhani et al., 1998; Wade et al., 2002, 2008) (see section 5).

Figure 2 shows thresholded activation maps of face, color, and place preferences for the group summary (Figure 2A) and for each individual subject (Figure 2B). The group summary shows locations where at least 5 subjects had voxelwise activation overlap for each contrast (see Methods). The pial-surface view of the group summary shows three parallel bands in each hemisphere, running along the posterior-anterior axis: color-biased activations (blue) were sandwiched between face-selective (purple) and place-selective activations (green), with face-selective cortex lateral to color-biased cortex, and place-selective cortex medial. This pattern was even more obvious in the inflated and flattened views (Figure 2A, lower panels), and was evident in data from most individual subjects (Figure 2B). Three subjects also showed some color-biased activation lateral to the FFA (S5, S7, S10, Figure 2B). In some subjects, Cc overlapped partially with the PPA (overlap indicated in cyan). While viewing activations on the inflated surface allows us to observe their relative position on the cortical sheet, smaller activations sometimes fail to project, despite being visible in the volume (Fischl et al., 1999; Operto et al., 2007). Activations that fall in regions of high signal inhomogeneity are particularly vulnerable to imperfect registration to the anatomy, which can lead to surface projection failure (Tucholka et al., 2012). Region Ac is acutely susceptible given its small size and proximity to the ear canals, a region of know signal inhomogeneity (discussed in section 5 below). Despite being visible in the native volume (where ROIs were defined), it failed to visibly project to the surface in several subjects (S2, S4, S8, S9, S11, S13). For example, Ac was not clear in the surface for subject 2 (Figure 2B), but was obvious in the native volume for this subject (slices; see Figure 4).

In sum, we find a parallel and adjacent structure of face, color, and place preferences on the ventral surface of the brain in humans. This pattern mirrors the systematic organization found on the lateral surface of macaque IT, suggesting a broad homology between the two regions.

Segregation of color and category in the VVP

To what extent are color and category information processed separately in the VVP? The brain activation maps described in the previous section show some spatial segregation of preferences for color, faces, and places. As described in the Methods, ROI-based analyses make possible stronger tests of our hypotheses about the functional relationship between color and other stimulus dimensions. Thus, category-selective (OFA, FFA, ATL, PPA) and color-biased (Pc, Cc, and Ac) ROIs were defined in each subject individually using data from even runs, and the response magnitude of each region to each of the ten conditions was quantified using the independent data from odd runs (Figure 3A-F, H). For comparison, we also defined a region that responded preferentially to intact objects (defined by shape selectivity not category selectivity) that is adjacent to Pc and OFA (region LO), and a face-selective region (pSTS) on the lateral surface (as opposed to the ventral surface) (Figure 3G, I). We discuss responses in these two regions in Sections 4 and 6 respectively.

An omnibus three-factor (ROI, category, and color/achromatic) ANOVA across the category-selective and color-biased ROIs of the VVP (OFA, FFA, ATL, PPA, Pc, Cc, and Ac) confirmed that these regions differ significantly from each other in their category selectivity (ROI x category interaction, p< 0.0001) and color selectivity (ROI x color interaction, p< 0.0001). These significant differences, as well as our prior hypotheses, license subsequent analyses of each ROI individually. Two-factor ANOVAs on stimulus category (faces, scenes, objects, and bodies) x color (color/achromatic) were run on each of the three color-biased regions (Pc, Cc, and Ac), and each of the four category-selective regions (OFA, FFA, ATL, and PPA) individually. The results of those analyses are shown in Table 1.

Table 1 (2-way ANOVAs Color x Category):

	Color & BW Faces, Bodies, Scenes, & Intact Objects		
Region	**M - Color**	**M - Cat**	**I – Color x Cat**
Pc	**0**	**0.001**	0.6
Cc	**0**	**0**	0.06
Ac	**0.001**	0.3	0.47
OFA	0.123	**0**	0.15
FFA	**0.003**	**0**	0.27
ATL	**0.003**	**0**	0.07
PPA	**0**	**0**	**0.002**
PSTS	0.39	**0**	**0.037**

Table 1: Two-factor ANOVAs on stimulus category (faces, scenes, objects, and bodies) x color (chromatic/achromatic). Significant effects are indicated in bold. P-values less than 0.0005 are indicated as zero. (M = main effect; I = interaction).

As expected, color-biased regions responded more strongly to colored than to achromatic stimuli; the size of this effect in each ROI can be seen in Figure 3 (significance levels are reported in Table 1). Category-selective regions showed the expected strong (Figure 3) and significant (Table 1) selectivity for their preferred category. However, most of the ROIs were selective for both color and category (Table 1). This result reflects in part the great statistical power of the ROI method (Saxe et al., 2006), which is able to detect statistically significant but very small effect sizes, e.g. of color in the FFA (the blue bar is slightly higher than the black bar, in the first pair of bars in Figure 3E). But are the color ROIs (Figure 3A-C) more color-selective than the category-selective ROIs (Figure 3D-F, H)? And are the category-selective ROIs are more category-selective than the color ROIs? Because there is no uncontroversial pairing of a particular color-biased ROI with a particular category-selective ROI, we answered these questions using a stringent exhaustive analysis of all twelve pairwise ANOVAs, contrasting each color ROI with each category ROI. This analysis confirmed that each color-biased ROI is more color-selective than each category-selective ROI (i.e., $p< 0.01$ for the interaction of ROI x color in 12/12 pairwise ANOVAs), and each category-selective ROI is more category-selective than each color ROI (i.e., $p< 0.0001$ for the interaction of ROI x category in 12/12 pairwise ANOVAs, $p<0.0001$). Note that the large number of statistical comparisons used here does not require correction for multiple comparisons, because our hypothesis required that each of the tests (rather than any of them) reach significance.

The statistical tests presented thus far provide evidence of some anatomical segregation of category and color responses: color-biased regions were more color-selective than category-selective regions, and category-selective regions were more category-selective than color-biased regions. If color and category processing were strongly independent, each kind of information would be carried by different populations of neurons, with no interaction. This strong hypothesis predicts that responses to color and category would not *interact within an ROI* (instead, the effect of the two factors on the response of a voxel containing the two distinct neural populations would be additive). This prediction follows from an extension to fMRI (Sternberg, 2011) of the classic additive factors logic from (Sternberg, 1969). All seven ROIs were tested (Pc, Cc, Ac, OFA, FFA, ATL, and PPA). Only the PPA showed a significant interaction of color by category (see Table 1).

These analyses provide evidence for a double dissociation between color and category preference across the VVP, as well as statistical independence of color and category preferences within six of the seven ROIs.

Quantification of the magnitude of the color bias across the VVP

As described above, the color-biased activations comprised a mottled band running posterior-to-anterior along the VVP (Figures 1B, 2, 4A); the three main peaks were defined as three color-biased regions, and these were identifiable in most subjects (10/13). These three color activations were identifiable because in each subject they were typically separated by tissue that did not show a significant color bias. Moreover, they were sufficiently stereotyped in their location within the brain such that they are apparent in a random-effects group analysis (Figure 4B). In Section 2, we showed that these regions were significantly more color-biased than category-selective regions. But how strong is their color selectivity relative to the VVP more broadly? In particular, are the color-biased regions more color selective than the cortical tissue that immediately surrounds them?

To answer this question, we used an ROI-based analysis that enabled us to leverage the power obtained by measuring responses in many subjects (see Methods). This method is powerful because it enables one to uncover activations that may not be significant in a given ROI within a single subject, but may be significant within that ROI when pooling results from many subjects. Using this method we ask, are the color-biased regions more color-biased than the immediately intervening patches of cortex?

In each individual subject we used data obtained in even runs to define two ROIs located between the color-biased activations with respect to the posterior-anterior axis (Figure 4C schematic, black circles). These included region IPCc (between Pc and Cc) and ICAc (between Cc and Ac). We used results obtained during odd runs to quantify responses. Figure 4C shows each region's mean percent signal change to all chromatic stimuli (cyan bar) and all achromatic stimuli (gray bar) ("all" = faces, bodies, scenes, objects, and scrambled objects). Although both intermediate regions were not color-biased in the individual subject contrast maps, they showed main effects of higher responses to chromatic than achromatic stimuli in the ROI analysis [both ps < 0.01; two-factor ANOVAs on stimulus class (faces, bodies, scenes, objects, and scrambled objects) x stimulus color]. Nonetheless, the color-biased regions were more color-biased than IPCc and ICAc: Pc and Cc were each more color-biased than IPCc (ps < 0.001); and Cc and Ac were each more color-biased than ICAc (ps < 0.003; pairwise 3-way ANOVAs on ROI x color x stimulus class, each contrasting an intermediate region with one of its neighboring color regions). These results show that although a weak color preference extends into surrounding cortex, the color-biased regions represent peaks in that spatial pattern of color bias.

Anatomical segregation of color and shape preferences

Prior work has documented regions of cortex that respond more strongly to intact than scrambled objects, including a posterior region called lateral occipital (LO) and an anterior region often referred to as posterior fusiform (pFS) (Grill-Spector et al., 1999). How segregated is shape processing from the analysis of color? As with the case of color and category described above, this question can be tested both in terms of anatomical segregation of preferences for shape and color, and functional independence within a region. Do responses within ROIs show an interaction of sensitivity to color and shape?

Following standard practice, we defined shape selectivity as a higher response to intact objects than scrambled objects ($p < 0.0001$) (Malach et al., 1995). The location of cortical regions showing a shape preference was anatomically segregated from the location of color-biased regions for most of the VVP. Figure 5A shows a surface histogram of the voxelwise group overlap map for shape-selectivity. The large swath of shape-selective cortex (consistent with published group-derived parcels for LO and pFS, dark blue outlines, from Julian et al., 2012), showed little overlap with color-biased cortex (outlined in light blue, from Figure 2A), except for the most anterior region where some overlap of shape and color preferences was evident. This pattern suggests that color and shape processing is segregated in posterior VVP but converges at more anterior stages of processing. To test this hypothesis, we analyzed data from individual subjects. We quantified the response profiles of shape-selective ROIs (LO and pFS) and the color ROIs in each subject. Figure 5B shows the average PSC across subjects to intact (O) and scrambled objects (SO), presented with color and without color. Two-factor ANOVAs on stimulus shape (Intact/Scrambled) x color (Chromatic/Achromatic) were run separately on Pc, Cc, and LO. Because Ac overlapped slightly with pFS in 5 of 10 subjects (i.e. the ROIs comprised common voxels), quantification of the responses of these regions was not independent of each other, and they and will be discussed in the next section. As expected, Pc and Cc showed main effects of higher responses to chromatic than achromatic stimuli ($ps < 0.0001$). Neither region showed a preference for intact over scrambled objects; to the contrary, Pc had a significant effect in the opposite direction. Conversely, LO showed a significant preference for intact shape ($p<0.0001$), but only a small preference for chromatic over achromatic stimuli ($p=0.003$), driven mostly by the scrambled object condition (see Figure 5B bar plot). To test whether the color ROIs were *more* color-biased than LO, and whether LO was more shape-selective than the color ROIs, we ran two 3-wayANOVAs, each contrasting one of the two posterior color ROIs with LO. This analysis confirmed that both the posterior and central color regions were more color-selective than LO (i.e., $p< 0.001$ for the interaction of ROI x color in each pairwise ANOVA), and LO was more shape-selective than Pc and Cc (i.e., $p< 0.001$ for the interaction of ROI x shape in each pairwise ANOVA). These analyses confirm the dissociation of color and shape preferences among the posterior color-biased regions (Pc, Cc) and LO. This dissociation can also be appreciated in Figure 5C, which plots the color bias as a function of the shape bias for all the ROIs. In this plot, the three color-biased regions are located in the upper half of the plot, and could be clearly separated by a horizontal line from the other regions; the shape-biased regions (notably LO) are located on the right of the plot, and could be separated by a vertical line from the color-biased regions.

As a final test of any integration of the processing of color and shape in Pc, Cc and LO, we asked whether color preferences depended on the presence of shape information or vice versa. This hypothesis predicts a superadditive interaction in which the presence of both color and shape information in a stimulus has a greater effect on the BOLD response than the sum of the two main effects. Contrary to this prediction, although all three regions (Pc, Cc, and LO) showed significant interactions of color and shape ($ps<0.03$, for the interaction of color x shape), these interactions went in the opposite direction, reflecting a greater sensitivity to color in scrambled rather than intact shapes (presumably because of the larger number of color edges present in the scrambled images).

Taken together, these results show that the processing of color and shape occupy largely segregated portions of posterior VVP. Next we turn to the more anterior portions of the VVP, including color region Ac, which showed evidence for modulation by both color and shape.

The anterior color-biased region

Here we describe the functional properties of the anterior color-biased region in detail. We defined Ac in each of the 10 subjects in whom it could be found in at least one hemisphere, and quantified the magnitude of response to each of the ten conditions in left-out data from these same subjects (Figure 3). As described above, we found (see Section 2) in a two-factor ANOVA of color by stimulus category that this region showed a significant main effect of color but no main effect of category, and no interaction of the two. Next we asked if this region was sensitive to object shape (Figure 5) by conducting a two-factor ANOVA (intact/scrambled objects x chromatic/achromatic). In addition to the expected main effect of color ($p=0.01$), this analysis found a significant main effect of shape ($p=0.01$), suggesting a functional difference between this color-biased region and the two posterior color-biased regions (Pc and Cc) where we did not see significant sensitivity to intact shape. A three-way ANOVA on the three color ROIs (Pc, Cc, and Ac) by Shape (intact/scrambled objects) by Color (chromatic/achromatic) confirmed that the anterior region was significantly more shape-selective (interaction of region by shape, $p < 0.0001$; driven by the preference for intact objects in Ac but not Pc, pairwise ANOVA, $p < 0.0001$).

Figure 5C shows a plot of color vs shape bias across all ROIs, and provides evidence that Ac is sensitive to both color and shape: the color ROIs (cyan triangles) fall in the plot in a pattern that mirrors their arrangement on the ventral surface: Pc falls to the left (prefers scrambled objects to intact objects), Ac to the right (some preference for shape), and Cc in between (no shape bias). Meanwhile, shape-biased regions (LO and pFS, dark blue symbols), cluster on the right (high shape-bias), and are low on the y-axis (weak color bias). Face regions and the intermediate-to-color regions (gray triangles) all reside on the right side of the plot (high shape-bias) and low on the y-axis (weak color-bias), with the exception of IPCc. We found no interaction between color and shape information in the anterior color region ($p=0.6$), suggesting that although both kinds of information are present, each dimension is processed independently within the region (see Methods).

Despite being reported less often in the literature than the posterior and central color regions, the anterior color region was evident both in the random effects group analysis (Figure 4b) and in individual-subject data (in 10 out of 13 subjects). Why do we see the anterior region in most subjects whereas many previous studies have not? First, we went to lengths to optimize tSNR in this region (see Methods), which lies in a part of the brain notorious for signal dropout (due to the presence of the ear canals) (Figure 6). Although tSNR was still low in this part of the brain, we were also able to reliably detect (in 9 of 13 subjects) the anterior temporal lobe face area (ATL), an area that has only been reliably detected (in more than 50% of subjects) through innovative signal enhancing efforts (Axelrod and Yovel, 2013). Also note that the inhomogeneity observed within the mottled band of color-biased activity cannot be attributed to variations in tSNR, as tSNR is strong between the color-biased regions (IPCc and ICAc). Second, more complex or engaging stimuli may be necessary in order to elicit activity in Ac. The few studies that have localized anterior color responses required demanding color-behavior tasks (Martin et al., 1995; Beauchamp et al., 1999; Simmons et al., 2007). Simpler stimuli, such as those used in many studies of color, may elicit weaker responses than dynamic natural movies, especially from regions implicated in high-level object vision. We tested this hypothesis by measuring responses in three subjects, to both video clips and gratings. Figure 7A shows the results for an example subject. The left panel shows the color activations observed using the movie clips (Pc, Cc and Ac outlined in white). The right panel shows the results of the gratings experiment (for the contrast of chromatic gratings > achromatic gratings). Only the two posterior regions, Pc and Cc, were detected in the gratings experiment (right panel; black arrow indicates absence of Ac activation). Responses to the movies (both colored and achromatic) were much higher across the visual system, in both early visual cortex (V1, V2) and the VVP (Figure 7B; average PSCs across the 3 subjects). This higher activation is perhaps unsurprising: the movies are more engaging than the gratings, and they comprise a rich mix of low-level visual features, including both luminance and colored components.

Is it possible that color activations extend beyond Ac but are obscured by low tSNR? The flattened surface view (Figure 6B, bottom panel) shows the extent of signal coverage obtained presently, for an example subject; regions anterior and medial to Ac are particularly impoverished (white asterisks), making it difficult to discern whether or not color activations span the full length of the temporal lobe (as they appear to in monkeys).

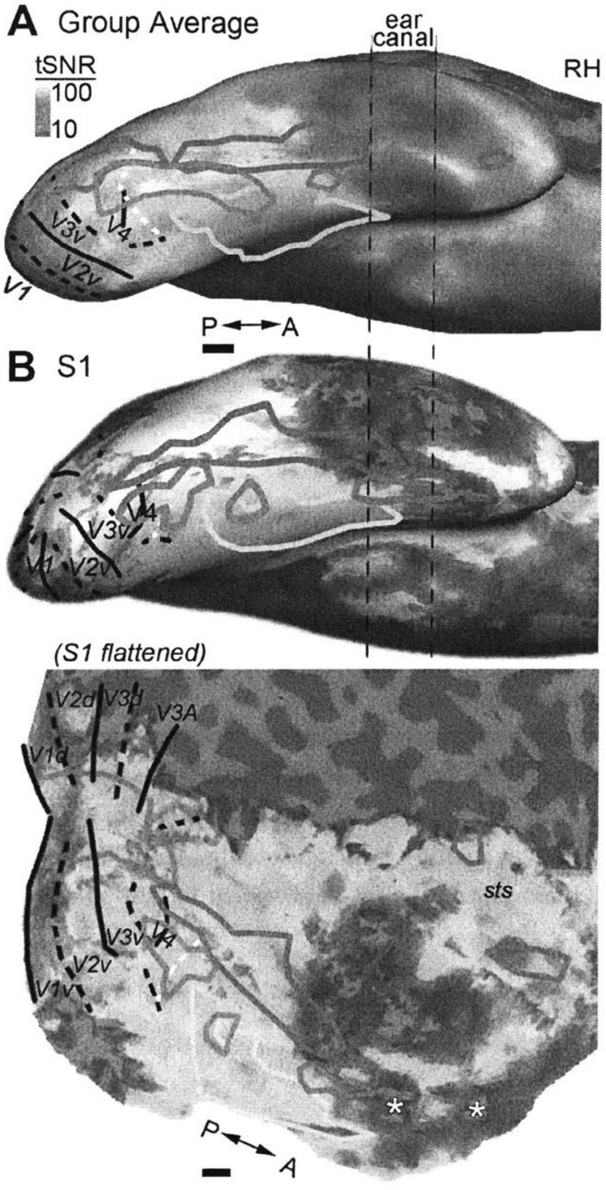


Figure 6. Signal coverage of the ventral surface. To compute the temporal signal-to-noise ratio (tSNR), the motion-corrected and smoothed (3mm) functional volumes were high-pass filtered with a cutoff of 0.004 Hz to remove slow drifts; for each voxel we then computed the mean of the time-course divided by the standard deviation of the time-course. **A,** Mean tSNR map across the group of subjects (N = 13). Regions of interest and other conventions from Figure 2a. The tSNR in cortex near the ear canals is low. **B,** Example subject's tSNR map (S1); inflated view (top) and flattened view (bottom). Gray outline shows the representation of the central 2 degrees of the visual field. Cortex anterior to the anterior color-biased region has low tSNR (white asterisks).

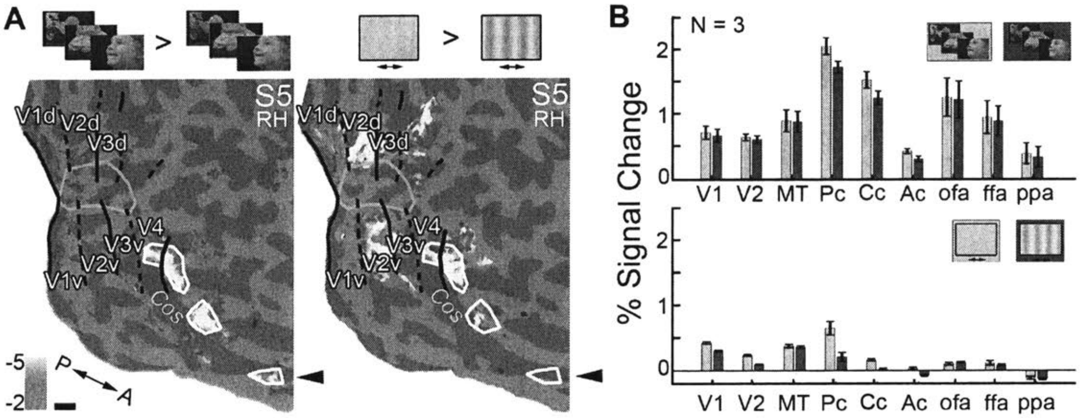


Figure 7. Comparison of responses to movie clips versus gratings. Color activity was measured using drifting equiluminant chromatic versus achromatic gratings. **A,** Comparison of color activations elicited by the movie clips (left) and the gratings (right) on the digitally flattened cortical surface (S5, right hemisphere; white outlines show the location of the color-biased activations localized using the movie clips). The gratings stimulus uncovered regions Pc and Cc, but failed to uncover Ac (black arrow). **B,** Quantification of the responses to movie clips (top) and gratings (bottom) (N=3 subjects); cyan bars (color conditions), gray (achromatic conditions). Functional ROIs (Pc, Cc, Ac, OFA, FFA, PPA) were defined with the movie clips; V1, V2, and MT+ were defined using an automatic parcellation (Freesurfer). Bars in the top graph were computed by averaging responses to all five categories.

Relationship to monkey organization

The topographic relationship of the face, color, and place activation patterns, and the relative functional dissociation of color-biased and category-biased cortex in humans, mirror results we reported in macaque inferior temporal (IT) cortex where color-biased regions are similarly sandwiched between face and place selective regions (Lafer-Sousa and Conway, 2013; Verhoef et al., 2015). In order to further assess the extent to which color and shape information are dissociated in the monkey, we scanned the same animals from the earlier report using the dynamic natural video stimuli used here in humans. We quantified the responses to video stimuli in the set of color-biased ROIs defined in the original report: PLc (posterior lateral color; located in PIT), CLc (central lateral color; located in CIT), and ALc (anterior lateral color; in AIT). The macaque color-biased regions responded similarly to the human color-biased regions: posterior regions (PLc and CLc) showed a color-bias but no shape bias, while the anterior region ALc showed both a color-bias and a shape bias (Figure 8). The middle face patch (ML) lacked a striking color-bias but showed a shape bias, consistent with the results obtained in human FFA.

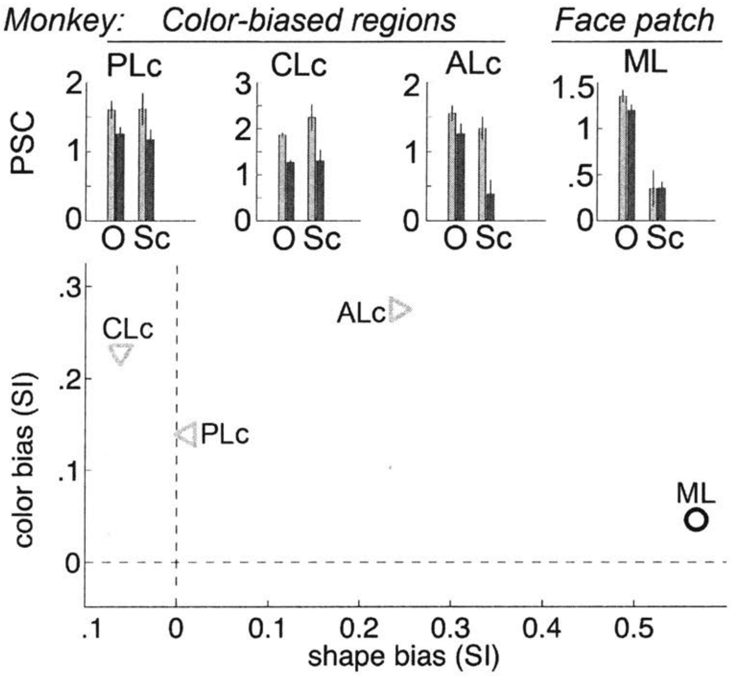


Figure 8. Responses to color and shape in monkeys, measured using movie clips. We measured responses to the movie clips (Figure 1A) in the two animals from an earlier report (Lafer-Sousa and Conway 2013). Bar plots (top) show the percent signal change in the functional regions defined in that study: PLc (posterior lateral color; located in PIT), CLc (central lateral color; located in CIT), and ALc (Anterior Lateral color; in AIT). Cyan bars show responses to colored movie clips, and grey bars show responses to achromatic movie clips, either for intact objects ("O") or scrambled objects ("Sc") (compare with data obtained in humans, Figure 5B). Scatter plot (bottom) shows color as a function of shape preference; conventions as in Figure 5C. The macaque color-biased regions responded similarly to the human color-biased regions; PLc and CLc showed a color bias but no shape preference, while ALc showed both a color bias and a shape bias. The macaque middle face patch (ML) showed a weak color-bias and a strong shape bias.

In addition to the face-selective regions in the human VVP (OFA, FFA, and ATL), a more dorsal face-selective region has been reported in the superior temporal sulcus (pSTS) (Collins and Olson, 2014). We detected this region in 11/13 subjects and quantified its responses in independent data as described for the regions in section 2 (see Figure 3). A two-factor ANOVA on these responses showed a main effect of category, a (weak) interaction of color by category, but no main effect of color (Table 1). The putative homology of human ventral regions and macaque IT raises the question: is there a comparable region in macaque monkeys of the human pSTS region? Freiwald and colleagues have proposed a division of the monkey face system into two streams (Yovel and Freiwald, 2013; Fisher and Freiwald, 2015). On this scheme, the ventral face stream in macaques, which follows the ventral lip of the STS and includes patches PL, ML, and AL, corresponds to regions OFA, FFA, and ATL in humans, whereas the dorsal face stream in macaques, residing along the dorsal lip of STS and comprising MD and AF

(Fisher and Freiwald, 2015), corresponds to human pSTS (and perhaps aSTS). This scheme would account for the functional dissociation found in both species (Haxby et al., 2000; Pitcher et al., 2011, 2014; Fisher and Freiwald, 2015) between face regions identifiable by responses to static images (PL/ML/MF/AL in macaques; OFA and FFA in humans), versus those that require dynamic stimuli (the dorsal face stream, AF/MD in macaques; pSTS in humans). Figure 9 provides a direct comparison of a representative monkey and human subject.

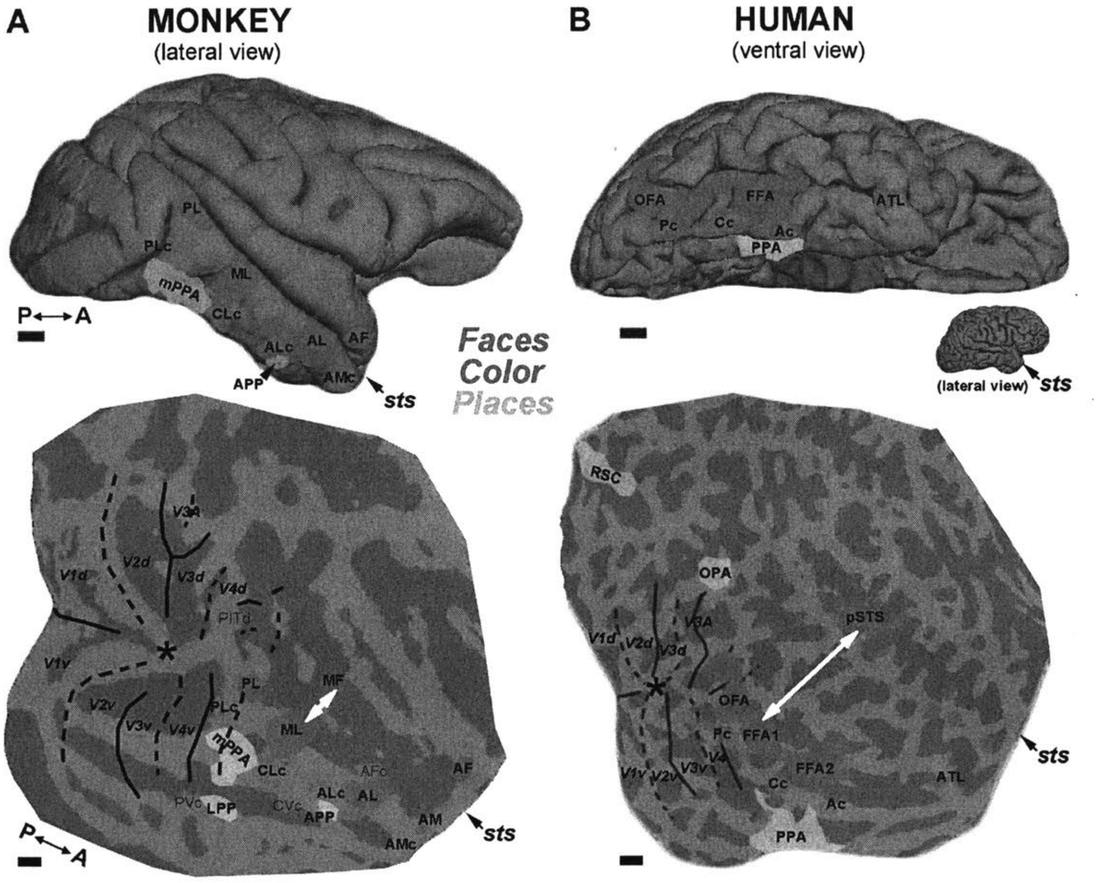


Figure 9. Comparison of the functional topography in monkey IT and human VVP suggests a broad homology. Reconstructed cortical surfaces showing the location of face (purple), color (blue), and place (green) regions of interest for a macaque monkey (panel **A**) and a human (**B**). Top panels show pial surface views (lateral view in the monkey; ventral view in the human); bottom panels show digitally flattened surfaces exposing functional activation buried in the superior temporal sulcus (sts; prefrontal cortex is cut off). In both species, color-biased regions (in macaque, PLc, CLc, ALc; in human, Pc, Cc and Ac) are sandwiched between face-selective regions (in macaque, PL, ML, AL; in human, OFA, FFA1, FFA2) and place-selective regions (in macaque, mPPA, APP; in human PPA). Both species show evidence for an additional face-processing stream in the sts, running parallel to the ventral pathway (in macaque, MF, AF; in human, pSTS, aSTS) (Yovel and Freiwald, 2013; Fisher and Freiwald, 2015). The region of cortex separating the two face-processing streams (white arrow) is relatively larger in humans; this tissue is implicated in communicative and social behaviors and expanded considerably during

evolution (Orban et al., 2004; Hill et al., 2010; Chaplin et al., 2013), accounting for the ventral (and medial) displacement of the ventral visual pathway in humans. Macaque data from Lafer-Sousa and Conway, 2013, and Verhoef et al. 2015.

Taken together these results suggest that the entire face-color-place tripartite cortical system is broadly homologous in macaque monkey and human, and moreover that this system migrated onto the ventral surface in humans as the cortex expanded over the course of evolution (see Discussion). Establishing this homology greatly empowers studies of high-level vision in humans by making directly relevant to humans the results from invasive studies in macaques.

2.5. Discussion

Here we address the functional organization of color processing in the human brain and its relationship to regions implicated in processing faces, objects, and places, using dynamic naturalistic stimuli that robustly drive the ventral visual pathway (VVP). We confirm the existence of a mottled band of color-biased cortex extending along the length of the VVP, within which we identified three main peaks of activation that correspond to previously described color-preferring regions. We found that the series of color-biased regions was sandwiched between bands of face-selective and place-selective regions, mirroring prior results from macaques. In addition, we found substantial anatomical segregation of color-biased regions from those preferring shape or category. Below we discuss two implications of this work: i) the homology it suggests between lateral regions of monkey cortex and ventral regions of human cortex; ii) the segregation it reveals of the cortical processing of color, shape, and category.

Color Regions are sandwiched between Face and Place Selective Regions, as in monkeys

We identified three color-biased regions: two posterior regions, referred to here as Pc and Cc (McKeefry and Zeki, 1997; Hadjikhani et al., 1998; Wade et al., 2002; Brewer et al., 2005; Simmons et al., 2007), and an anterior region, Ac [likely corresponding to an unnamed region reported by Simmons et al. (2007)]. These regions were neatly sandwiched in each subject between face-selective regions on the lateral side, and place-selective regions on the medial side. This systematic arrangement, of three parallel sets of regions running along the posterior-to-anterior axis of ventral occipitotemporal cortex, mirrors the organization of lateral temporal cortex (IT) in macaque monkey (Figure 9) (Lafer-Sousa and Conway, 2013), and suggests that the entire tripartite cortical structure responsible for face/color/place processing constitutes a 'regional homology' (Orban et al., 2004) between the two species.

If this face/color/place system is homologous between humans and macaques, why is it in such a different location within the brain in the two species? We speculate that this system was pushed onto the ventral surface of the brain over the course of human evolution by the expansion of regions engaged in language and social cognition (white arrow in Figure 9). The human cortex expanded 10-fold since humans and macaques diverged from a common ancestor (Orban et al., 2004). But this expansion was not uniform. Regions involved in communicative and social behaviors, including the temporal parietal junction and the ventrolateral prefrontal cortex, expanded as much as 30-fold, while visual areas expanded relatively little (2-6-fold) (Orban et al., 2004; Hill et al., 2010; Chaplin et al., 2013). Therefore the identification of neighborhood relationships, like the parallel organization of face/color/place selectivity, may be a better cue to homology than gross anatomical correspondences (Orban et al., 2004). Our hypothesized homology of the face/color/place tripartite system between humans and macaques is consistent with prior hypotheses about the face system (Tsao et al., 2008; Yovel and Freiwald, 2013), and could be further investigated using tests of connectivity, cytoarchitecture, and gene expression.

Dissociation of Color from Shape and Category

How segregated is the processing of color and shape in the brain? On the one hand, the spatial configuration of colors within a scene has a profound influence on color perception, leading some to argue that color and form are linked inextricably in visual cortical processing, especially in V-1, early in the visual-processing hierarchy (Gegenfurtner and Kiper, 2003; Shapley and Hawken, 2010). On the other hand, color presents computational challenges somewhat distinct from those associated with shape and scene processing. The information provided by color and shape is often independent, for example in signaling a person's emotional state independent of face features, or the edibility of a piece of fruit (e.g., bananas ripen, green to yellow, without changing shape); moreover, color and luminance edges are often independent in natural scenes (Hansen and Gegenfurtner, 2009). These observations suggest that color might be processed separately from shape.

We found substantial segregation of color and shape processing in the VVP, consistent with prior evidence for separate pathways in the VVP for processing surface properties (texture and color) versus geometric/shape properties of objects in humans (Cant et al., 2008, 2009; Cant and Goodale, 2009; Cavina-Pratesi et al., 2010). Specifically, the two posterior color-biased regions did not show a preference for intact objects (responding equally, if not more strongly, to their scrambled counterparts) (Figure 5B). Conversely, shape-biased cortex (LO) showed only a weak preference for color (and this was found only for scrambled, not intact, objects). This dissociation of color and form processing accords with prior fMRI studies, and with double dissociations reported in neurological patients, in which color perception can be impaired while form perception is spared [e.g. patient MS (Bouvier and Engel, 2006); or vice versa, e.g. patient DF (Humphrey et al., 1994)].

Objects can be defined not only by their shape, but also by the more abstract category to which they belong. We found that the VVP showed segregation not only for shape and color, but also for category and color: each color-biased region showed a higher response for colored than achromatic stimuli for each of the five stimulus categories, and this effect did not differ in magnitude across categories. Further, each color-biased region was more color-biased than each category-selective region. This anatomical segregation of regions preferring color versus those preferring object category suggests that category-selective regions rely primarily on shape information; the lack of color selectivity within shape or category selective regions is consistent with the observation that one can recognize most objects from achromatic images (Gregory, 1977). The hypothesized dissociation of color and category processing supported by the present results could be further evaluated using adaptation (Cavina-Pratesi et al., 2010) and multivariate pattern analysis (Brouwer and Heeger, 2013).

Our findings accord with the available data from causal methods. First, the majority (72%) of cortical lesion patients who acquire achromatopsia also acquire prosopagnosia, but rarely object agnosia (8%) (Bouvier and Engel, 2006). This pattern is expected given that cortical lesions are typically larger than the scale of any single functional domain, and color and face regions tend to be adjacent. Second, the few cases of electrical stimulation of the fusiform gyrus in humans produce specific perceptual effects (Rangarajan et al., 2014) consistent with our findings. For example, stimulation of an electrode overlying face-selective tissue (FFA) of one subject elicited the response "You turned into someone else. Your face metamorphosed...your nose got saggy and went to the left". Stimulation of an electrode overlying non-face-selective tissue (posterior to the FFA, likely near Pc) prompted the response, "Light moved across my eyes. It was green, purple, and yellow light together." (Rangarajan et al., 2014).

Overlap of shape and color at more anterior locations of the VVP

The perception of complex scene structure modulates multiple aspects of color perception: color constancy depends on knowledge of 3D scene structure (Bloj et al., 1999; Lafer-Sousa et al, 2015), and pan-field color filling-in fails when scene structure is scrambled (Balas and Sinha, 2007). Moreover, the memory of object colors can modulate object appearance (Hansen et al., 2006; Olkkonen et al., 2010; Witzel and Gegenfurtner, 2013) and facilitate object and scene recognition (Tanaka and Presnell, 1999; Gegenfurtner and Rieger, 2000; Oliva and Schyns, 2000; Nijboer et al., 2007) when color is specifically associated with a given object or scene. These observations suggest that color and shape interact at a high level stage of visual processing (Tanaka et al., 2001). In support of an interaction of color and shape processing, we found anatomical convergence of color and shape bias in the most anterior color-biased region, a region implicated in color knowledge (Simmons et al., 2007). We observed a similar result in macaques, where the anterior color-biased region (ALc) showed a shape bias whereas the more posterior color-biased regions did not (Figure 9). This result is consistent with physiological findings of intermingled cell-types in the neighborhood of ALc, which can be selective for color, shape, or both (Komatsu and Ideura, 1993; Edwards et al., 2003). We speculate that the color-biased regions of the VVP form a multistage system for the analysis of color, analogous to that observed in the adjacent face system (Lafer-Sousa and Conway, 2013; Tsao, 2014). This hypothesized architecture is consistent with functional differences between posterior and anterior color regions reported here and elsewhere (Simmons et al., 2007), and with observed differences in the color processing deficits that can result from lesions in different anatomical locations (Miceli et al., 2001). The specific computations these regions support is largely unknown, but the functional architecture uncovered here provides a roadmap for their investigation.

In sum, our data support two major conclusions. First, humans and macaques show striking similarity in the topographic relationship of color-biased regions, face-selective regions and place-selective regions, suggestive of a broad interspecies regional homology. Second, color, shape, and category selectivity appear to be substantially segregated throughout much of the VVP. These results invite future work into the computations, connectivity, and developmental origins of the tripartite face/color/place system, a research program made more tractable by the system's apparent homology to macaques.

2.6. References

Axelrod V, Yovel G (2013) The challenge of localizing the anterior temporal face area: A possible solution. NeuroImage 81: 371-380.

Balas B, Sinha P (2007) “Filling-in” colour in natural scenes. Visual Cognition 15:765778.

Bartels A, Zeki S (2000) The architecture of the colour centre in the human visual brain: new results and a review. Eur J Neurosci 12:172–193.

Beauchamp M, Haxby J, Jennings J, DeYoe E (1999) An fMRI Version of the Farnsworth–Munsell 100-Hue Test Reveals Multiple Color-selective Areas in Human Ventral Occipitotemporal Cortex. Cereb Cortex 9:257–263.

Bloj MG, Kersten D, Hurlbert AC (1999) Perception of three-dimensional shape influences colour perception through mutual illumination. Nature 402:877–879.

Bouvier SE, Engel SA (2006) Behavioral deficits and cortical damage loci in cerebral achromatopsia. Cereb Cortex 16:183–191.

Brewer A, Liu J, Wade A, Wandell B (2005) Visual field maps and stimulus selectivity in human ventral occipital cortex. Nat Neurosci 8:1102–1109.

Brouwer G, Heeger D (2013) Categorical Clustering of the Neural Representation of Color. Journal of Neuroscience 33:15454–15465.

Cant JS, Arnott SR, Goodale MA (2009) fMR-adaptation reveals separate processing regions for the perception of form and texture in the human ventral stream. Exp Brain Res 192:391–405.

Cant JS, Goodale MA (2009) Asymmetric interference between the perception of shape and the perception of surface properties. J Vis 9:13.1–20.

Cant JS, Large M-EE, McCall L, Goodale MA (2008) Independent processing of form, colour, and texture in object perception. Perception 37:57–78.

Cavina-Pratesi C, Kentridge RW, Heywood CA, Milner AD (2010) Separate channels for processing form, texture, and color: evidence from FMRI adaptation and visual object agnosia. Cereb Cortex 20:2319–2332.

Chaplin T, Yu H-H, Soares J, Gattass R, Rosa M (2013) A Conserved Pattern of Differential Expansion of Cortical Areas in Simian Primates. Journal of Neuroscience 33:15120–15125.

Collins JA, Olson I R (2014). Beyond the FFA: The role of the ventral anterior temporal lobes in face processing. Neuropsychologia, 61: 65-79.

Conway BR (2001) Spatial structure of cone inputs to color cells in alert macaque primary visual cortex (V-1). J Neurosci 21:2768–2783.

Conway BR (2009) Color vision, cones, and color-coding in the cortex. Neuroscientist 15:274–290.

Conway BR (2014) Color signals through dorsal and ventral visual pathways. Vis Neurosci 31:197–209.

Conway BR, Tsao DY (2006) Color architecture in alert macaque cortex revealed by FMRI. Cereb Cortex 16:1604–1613.

Dale AM, Fischl B, Sereno MI (1999) Cortical surface-based analysis. I. Segmentation and surface reconstruction. Neuroimage 9:179–194.

Derrington AM, Krauskopf J, Lennie P (1984) Chromatic mechanisms in lateral geniculate nucleus of macaque. J Physiol (Lond) 357:241–265.

Edwards R, Xiao D, Keysers C, Földiák P, Perrett D (2003) Color Sensitivity of Cells Responsive to Complex Stimuli in the Temporal Cortex. J Neurophysiol 90:1245–1256.

Efron B (1982) The Jackknife, the Bootstrap and Other Resampling Plans. Philadelphia: SIAM.

Epstein R, Kanwisher N (1998) A cortical representation of the local visual environment. Nature, 392: 598-601.

Epstein RA, Parker WE, Feiler AM (2007). Where am I now? Distinct roles for parahippocampal and retrosplenial cortices in place recognition. J Neurosci, 27(23): 6141-6149.

Fischl B, Liu A, Dale AM (2001) Automated manifold surgery: constructing geometrically accurate and topologically correct models of the human cerebral cortex. IEEE Trans Med Imaging 20:70–80.

Fischl B, Sereno MI, Dale AM (1999) Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. Neuroimage 9:195–207.

Fisher C, Freiwald W (2015) Contrasting Specializations for Facial Motion within the Macaque Face-Processing System. Curr Biol 25.

Friston, Holmes, Worsley, Poline J -P., Frith, Frackowiak (1994) Statistical parametric maps in functional imaging: A general linear approach. Human Brain Mapping 2:189–210.

Gagin G, Bohon KS, Butensky A, Gates MA, Hu J-YY, Lafer-Sousa R, Pulumo RL, Qu J, Stoughton CM, Swanbeck SN, Conway BR (2014) Color-detection thresholds in rhesus macaque monkeys and humans. J Vis 14:12.

Gallant J, Shoup R, Mazer J (2000) A Human Extrastriate Area Functionally Homologous to Macaque V4. Neuron 27.

Gegenfurtner K, Kiper D (2003) COLOR VISION. Annu Rev Neurosci 26:181–206.

Gegenfurtner KR, Rieger J (2000) Sensory and cognitive contributions of color to the recognition of natural scenes. Curr Biol 10:805–808.

Gregory RL (1977) Vision with isoluminant colour contrast: 1.A projection technique and observations. Perception 6:113–119.

Greve DN, Fischl B (2009) Accurate and robust brain image alignment using boundary-based registration. Neuroimage 48:63–72.

Grill-Spector K, Kushnir T, Edelman S, Avidan G, Itzchak Y, Malach R (1999) Differential Processing of Objects under Various Viewing Conditions in the Human Lateral Occipital Complex. Neuron 24.

Griswold MA, Jakob PM, Heidemann RM, Nittka M, Jellus V, Wang J, Kiefer B, Haase A (2002) Generalized autocalibrating partially parallel acquisitions (GRAPPA). Magn Reson Med 47:1202–1210.

Hadjikhani N, Liu AK, Dale AM, Cavanagh P, Tootell RB (1998) Retinotopy and color sensitivity in human visual cortical area V8. Nat Neurosci 1:235–241.

Hansen T, Gegenfurtner KR (2009) Independence of color and luminance edges in natural scenes. Vis Neurosci 26:35–49.

Hansen T, Giesel M, Gegenfurtner KR (2008) Chromatic discrimination of natural objects. J Vis 8:2.1–19.

Hansen T, Olkkonen M, Walter S, Gegenfurtner KR (2006) Memory modulates color appearance. Nat Neurosci 9:1367–1368.

Haxby JV, Hoffman EA, Gobbini MI (2000) The distributed human neural system for face perception. Trends Cogn Sci (Regul Ed) 4:223–233.

Heywood CA, Cowey A, Newcombe F (1991) Chromatic Discrimination in a Cortically Colour Blind Observer. Eur J Neurosci 3:802–812.

Heywood CA, Wilson B, Cowey A (1987) A case study of cortical colour "blindness" with relatively intact achromatic discrimination. J Neurol Neurosurg Psychiatr 50:22–29.

Hill J, Inder T, Neil J, Dierker D, Harwell J, Essen D (2010) Similar patterns of cortical expansion during human development and evolution. Proceedings of the National Academy of Sciences 107:13135–13140.

Humphrey GK, Goodale MA, Jakobson LS, Servos P (1994) The role of surface information in object recognition: studies of a visual form agnosic and normal subjects. Perception 23:1457–1481.

Julian J, Fedorenko E, Webster J, Kanwisher N (2012) An algorithmic method for functionally defining regions of interest in the ventral visual pathway. NeuroImage 60:2357-2364.

Kanwisher N, McDermott J, Chun MM (1997) The Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for Face Perception. J Neurosci, 17(11): 4302-4311.

Komatsu H, Ideura Y (1993) Relationships between color, shape, and pattern selectivities of neurons in the inferior temporal cortex of the monkey. J Neurophysiol 70:677–694.

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.

Lafer-Sousa R, Conway BR (2013) Parallel, multi-stage processing of colors, faces and shapes in macaque inferior temporal cortex. Nat Neurosci 16:1870–1878.

Lafer-Sousa R, Hermann K, Conway B (2015) Striking individual differences in color perception uncovered by “the dress” photograph. Current Biology 25.

MacLeod DIA, Boynton RM (1979) Chromaticity diagram showing cone excitation by stimuli of equal luminance. Journal of the Optical Society of America, 69(8), 1183-1186.

Malach R, Reppas J, Benson R, Kwong K, Jiang H, Kennedy W, Ledden P, Brady T, Rosen B, Tootell R (1995) Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. Proc Natl Acad Sci USA 92:8135–8139.

Martin A, Haxby JV, Lalonde FM, Wiggs CL, Ungerleider LG (1995) Discrete cortical regions associated with knowledge of color and knowledge of action. Science 270:102–105.

McKeefry D, Zeki S (1997) The position and topography of the human colour centre as revealed by functional magnetic resonance imaging. Brain 120:2229–2242.

Meadows JC (1974) Disturbed perception of colours associated with localized cerebral lesions. Brain 97:615–632.

Miceli G, Fouch E, Capasso R, Shelton JR, Tomaiuolo F, Caramazza A (2001) The dissociation of color from form and function knowledge. Nat Neurosci 4:662–667.

Milner AD, Heywood CA (1989) A disorder of lightness discrimination in a case of visual form agnosia. Cortex 25:489–494.

Murphey D, Yoshor D, Beauchamp M (2008) Perception Matches Selectivity in the Human Anterior Color Center. Current Biology 18.

Newcombe F, Ratcliff G (1975) Agnosia; a disorder of object recognition. In, pp 317–341. Lyon: Colloque International de Lyon.

Nijboer T, Smagt M, Zandvoort M, Haan E (2007) Colour agnosia impairs the recognition of natural but not of non-natural scenes. Cognitive Neuropsychology 24:152161.

Oliva A, Schyns PG (2000) Diagnostic colors mediate scene recognition. Cogn Psychol 41:176–210.

Olkkonen M, Witzel C, Hansen T, Gegenfurtner KR (2010) Categorical color constancy for real surfaces. J Vis 10.

Operto G, Bulot R, Anton J-L, Coulon O (2007) Projection of fMRI data onto the cortical surface using anatomically-informed convolution kernels. Neuroimage 39:127–135.

Orban G, Essen D, Vanduffel W (2004) Comparative mapping of higher visual areas in monkeys and humans. Trends in Cognitive Sciences 8.

Pitcher D, Vincent W, Bradley D. The role of the occipital face area in the cortical face perception network. Experimental Brain Research 209, no. 4 (2011): 481-493.

Pitcher D, Dilks DD, Saxe RR, Triantafyllou C, Kanwisher N (2011) Differential selectivity for dynamic versus static information in face-selective cortical regions. Neuroimage 56:2356–2363.

Pitcher D, Duchaine B, Walsh V (2014) Combined TMS and FMRI reveal dissociable cortical pathways for dynamic and static face perception. Curr Biol 24:2066–2070.

Puce A, Allison T, Bentin S, Gore JC, McCarthy G Temporal cortex activation in humans viewing eye and mouth movements._J Neurosci, 18(6): 2188-99.

Rangarajan V, Hermes D, Foster B, Weiner K, Jacques C, Grill-Spector K, Parvizi J (2014) Electrical Stimulation of the Left and Right Human Fusiform Gyrus Causes Different Effects in Conscious Face Perception. Journal of Neuroscience 34:12828-12836.

Saxe R, Brett M, Kanwisher N (2006) Divide and conquer: A defense of functional localizers. NeuroImage 30:10881096.

Schein SJ, Desimone R (1990) Spectral properties of V4 neurons in the macaque. J Neurosci 10:3369–3389.

Sereno MI, Dale AM, Reppas JB, Kwong KK, Belliveau JW, Brady TJ, Rosen BR, Tootell RB (1995) Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. Science 268:889–893.

Shapley R, Hawken M (2010) Color in the Cortex: single- and double-opponent cells. Vision Research 51.

Simmons W, Ramjee V, Beauchamp M, McRae K, Martin A, Barsalou L (2007) A common neural substrate for perceiving and knowing about color. Neuropsychologia 45:2802–2810.

Smith VC, Pokorny J (1975) Spectral sensitivity of the foveal cone photopigments between 400 and 500 nm. Vision Res 15:161–171.

Sternberg S (1969) The discovery of processing stages: Extensions of Donders' method. Acta Psychologica 30:276–315.

Sternberg S (2011) Modular processes in mind and brain. Cogn Neuropsychol 28:156–208.

Stoughton CM, Lafer-Sousa R, Gagin G, Conway BR (2012) Psychophysical chromatic mechanisms in macaque monkey. J Neurosci 32:15216–15226.

Tanaka J, Weiskopf D, Williams P (2001) The role of color in high-level vision. Trends Cogn Sci (Regul Ed) 5:211–215.

Tanaka JW, Presnell LM (1999) Color diagnosticity in object recognition. Percept Psychophys 61:1140–1153.

Tootell RB, Hadjikhani N (2001) Where is "dorsal V4" in human visual cortex? Retinotopic, topographic and functional evidence. Cereb Cortex 11:298–311.

Tsao D (2014) The Macaque Face Patch System: A Window into Object Representation. Cold Spring Harb Symp Quant Biol 79:109–114.

Tsao DY, Moeller S, Freiwald W. 2008. Comparing face patch systems in macaques and humans. **PNAS**, 49:19514-9.

Tucholka A, Fritsch V, Poline J-B, Thirion B (2012) An empirical comparison of surface-based and volume-based group studies in neuroimaging. Neuroimage 63:1443–1453.

Vanduffel W, Fize D, Mandeville J, Nelissen K, Hecke P, Rosen B, Tootell R, Orban G (2001) Visual Motion Processing Investigated Using Contrast Agent-Enhanced fMRI in Awake Behaving Monkeys. Neuron 32.

Verhoef B-EE, Bohon KS, Conway BR (2015) Functional architecture for disparity in macaque inferior temporal cortex and its relationship to the architecture for faces, color, scenes, and visual field. J Neurosci 35:6952–6968.

Wade A, Augath M, Logothetis N, Wandell B (2008) fMRI measurements of color in macaque and human. J Vis 8.

Wade A, Brewer A, Rieger J, Wandell B (2002) Functional measurements of human ventral occipital cortex: retinotopy and colour. Philos Trans R Soc Lond, B, Biol Sci 357:963–973.

Weiner S, Grill-Spector K (2010). Sparsely-distributed organization of face and limb activations in human ventral temporal cortex. Neuroimage, 52(4): 1559-1573.

Winawer J, Horiguchi H, Sayres R, Amano K, Wandell B (2010) Mapping hV4 and ventral occipital cortex: The venous eclipse. Journal of Vision 10.

Witthoft N, Nguyen M, Golarai G, LaRocque K, Liberman A, Smith M, Grill-Spector K (2013) Where Is Human V4? Predicting the Location of hV4 and VO1 from Cortical Folding. Cerebral Cortex, 9:2401-2408.

Witzel C, Gegenfurtner KR (2013) Categorical sensitivity to color differences. J Vis 13:1.

Yasuda M, Banno T, Komatsu H (2010) Color Selectivity of Neurons in the Posterior Inferior Temporal Cortex of the Macaque Monkey. Cerebral Cortex, 20(7): 1630-46.

Yovel G, Freiwald WA (2013) Face recognition systems in monkey and human: are they the same thing? F1000Prime Rep 5:10.

Zeki S (1980) The representation of colours in the cerebral cortex. Nature 284:412–418.

Zeki S (1983) The relationship between wavelength and color studied in single cells of monkey striate cortex. Prog Brain Res 58:219–227.

Zeki S (1990) A century of cerebral achromatopsia. Brain 113 (Pt 6):1721–1777.

Zeki S, Watson JD, Lueck CJ, Friston KJ, Kennard C, Frackowiak RS (1991) A direct demonstration of functional specialization in human visual cortex. J Neurosci 11:641–649.

Chapter 3: Color constancy: resolving ambiguous color signals

Chapter 3.1: Striking individual differences in color perception revealed by 'the dress' photograph.

This work has been published:
Lafer-Sousa, R.*, Hermann, K.*, Conway, B.R. Striking individual differences in color perception revealed by 'the dress' photograph. *Cur. Biol.* **25**, R1-R2 (2015)

"The Dress" is a peculiar photograph: by themselves the dress' pixels are brown and blue, colors associated with natural illuminants (Conway, 2015), but popular accounts suggest the dress appears white/gold or blue/black (Rogers, 2015). Could the purported categorical perception arise because the original social-media question was an alternative-forced-choice? In a free-response survey (N=1401), we found that most people, including those naïve to the image, reported white/gold or blue/black, but some said blue/brown. Reports of white/gold over blue/black were higher among older people and women. On re-test, some subjects reported a switch in perception, showing the image can be multistable. In a language-independent measure of perception, we asked subjects to identify the dress' colors from a complete color gamut. The results showed three peaks corresponding to the main descriptive categories, providing additional evidence that the brain resolves the image into one of three stable percepts. We hypothesize these reflect different internal priors: some people favor a cool illuminant (blue sky), discount shorter wavelengths, and perceive white/gold; others favor a warm illuminant (incandescent light), discount longer wavelengths, and see blue/black. The remaining subjects may assume a neutral illuminant, and see blue/brown. By introducing overt cues to the illumination, we can flip the dress color.

Popular accounts suggest The Dress (Figure 1A/B) elicits large individual differences in color perception (Rogers, 2015). We confirmed this in a survey of 1401 subjects (313 naïve; 53 tested in laboratory; 28/53 re-tested). Subjects were asked to complete: "this is a ________and___ ______dress" (Supplemental Experimental Procedures).

Overall, 57% of subjects described the dress as blue/black (B/K); 30%, white/gold (W/G); 11%, blue/brown (B/B); and 2%, other. Redundant descriptions (e.g. "white-golden", "white-goldish") were binned (Supplemental Results). Naïve and non-naïve populations showed similar distributions (Figure 1C), although non-naïve subjects used a smaller number of unique descriptions (Figure S1A). When country (Figure S1B) was removed from the logistic regression (Table S1), experience became a predictor: non-naïve subjects were more likely to choose B/K or W/G, over B/B or other ($p = 0.021$, Wald chi-square; Odds Ratio (OR) = 1.53, 95% C.I. [1.06-2.17]). These results show that experience shaped the language used to describe the dress, and possibly the perception of it. Males were less likely than females to report W/G over B/K ($p = 0.019$, OR=0.75, [0.58-0.95]). Moreover, odds of reporting W/G increased with age (Figure 1D).

Of non-naive subjects, 45% reported a switch since first exposure. Three of 28 subjects retested in laboratory reported a switch between sessions. Subjects whose perception switched were more likely to report B/K (p = 0.0003, OR = 0.60 [0.46-0.79], where W/G=success).

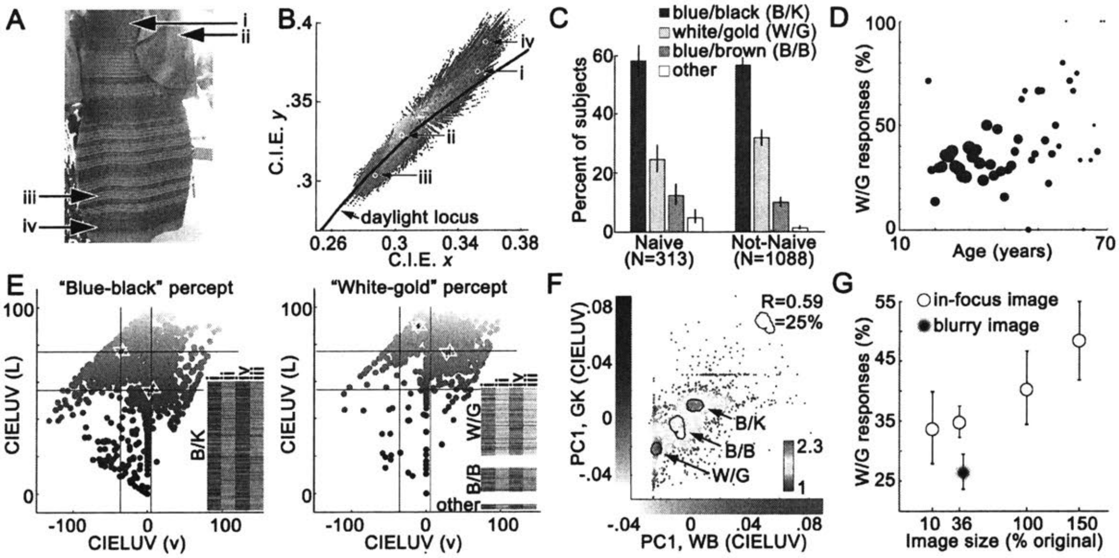


Figure 1. Striking differences in color perception of The Dress. **(A)** Original photograph, reproduced with permission from Cecilia Bleasdale. **(B)** Pixel chromaticities for the dress. **(C)** Histogram of color descriptions, for naïve (N=313) and non-naive (N=1088) subjects. Error bars are 95% C.I. **(D)** Of subjects who reported W/G or B/K (N=1221), the odds of reporting W/G increased by a factor 1.02 per unit age, p=0.0035, 95% C.I. [1.01-1.03] **(Table S1)**. Symbol size denotes number of subjects (largest dot=76; smallest dot=1). **(E)** Color matches for regions i-iv (panel A), sorted by color description (B/K, left; W/G, right). Symbols show averages (upward triangles, regions i and ii; downward triangles, regions iii and iv), and contain 95% C.I.s of the mean. Grid provides a reference across the B/K and W/G panels. Insets depict color matches for individual subjects in each row, sorted by description. **(F)** Color matches for region (i) plotted against matches for region (ii) for all subjects (R=0.59, p<0.0001). Contours contain the highest density (25%) of respondents obtained in separate plots (not shown) generated by sorting the data by description (B/K, W/G, B/B). The first principal component of the population matches to (i, iv) defined the y axis (gold/black, "GK"); the first PC of the population matches to (ii, iii) defined the x axis (white/blue, "WB"). Each subject's (x,y) values are the PC weights for their matches (Supplementary Experimental Procedures). Color scale is number of subjects. **(G)** Among W/G or B/K respondents, percent of W/G responses increased with image size (N=235, 10%; N=1223, 36%; N=245, 100%; N=215, 150%; p < 0.0001, OR=1.004 [1.002-1.007]).

The horizontal dimension of the image was about 2°, 7.2°, 20°, and 30° of visual angle. Blurring the image biased responses towards B/K (N=1048, image was 41% of original size; image was 260x401pixels with a 0.11° pixel radius Gaussian blur; Chi-square, p<0.0001).

Subjects were asked to match the dress' colors. Blue pixels (ii, iii, Figure 1A) were consistently matched bluer by subjects reporting B/K and whiter by subjects reporting W/G, whereas brown pixels (i, iv) were matched blacker by subjects reporting B/K and golden by subjects reporting W/G (Figure 1E; Figure S1C). For a given region, average color matches made by W/G perceivers differed in both lightness and hue from matches made by B/K perceivers (p vals.<0.0001). Intra-subject reliability was significant (Figure S1D,E). Across all subjects, matches for (i) were predictive of matches for (ii); moreover, the density plot showed three peaks (Figure 1F; Figure S1F,G). The highest density of W/G, B/K, and B/B responders (contours Figure 1F) coincide with these peaks, suggesting that the brain resolves the image into one of three stable percepts.

We suspect that priors on both material properties (Zaidi, 2011; Witzel et al., 2011) and illumination (Bloj, 1999) are implicated in resolving the dress' color. In the main experiment, the image was 36% of the original size. In a follow-up experiment (N=853 additional subjects), the fraction of W/G respondents rose with increasing image size (Figure 1G). This suggests that high-spatial frequency information (a cue to dress material), more evident at larger sizes, biases interpretation toward W/G. To further test this, we determined responses to a blurry image: the fraction of W/G respondents dropped. Subjects also rated the illumination for The Dress and two test images showing the dress under cool or warm illumination (Figure S2A). Judgment variance was higher for the original than for either test (cool, $p=10^{-5}$; warm, $p=10^{-7}$, F-test), but similar for the tests (p=0.08), suggesting that illumination in The Dress is ambiguous. When the dress was embedded in a scene with unambiguous illumination cues, the majority of subjects conformed to a description predicted by the illumination (Figure S2B).

A color percept is the visual system's best guess given available sense data and an internal model of the world (Brainard et al., 2006). Visual cortex shows a bias for colors associated with daylight (Lafter-Sousa et al., 2012; Conway, 2014); this bias may represent the brain's internal model. We hypothesize that some brains interpret the surprising chromatic distribution (Figure 1B) as evidence that a portion of the spectral radiance is caused by a color bias of the illuminant (Conway, 2015) (Supplementary Discussion). Some people may expect a cool illuminant, discount short wavelengths, and perceive white/gold; others may favor a warm illuminant, discount longer wavelengths, and see blue/black. The remaining people may assume a neutral illuminant and see blue/brown. But what causes the individual differences? People experience different illuminants and adapt (McDermott et al., 2010). If exposure informs one's prior, we might predict that older subjects and women are more likely to assume sky-blue illumination because they are more likely than younger subjects and men to have a daytime chronotype (Adan et al., 2012). Consistent with this prediction, women and older people were more likely to see white/gold. Conversely, night owls may be more likely to assume a warmer illuminant (Rogers, 2015) common for artificial light, and see blue/black. Alternatively, all people may have the same prior on the illuminant, but different priors on other aspects of the scene that interact to produce different percepts of the dress.

3.1.1. References

Adan, A., Archer, S.N., Hidalgo, M.P., Di Milia, L., Natale, V., and Randler, C. (2012). Circadian typology: a comprehensive review. Chronobiol Int *29*, 1153-1175.

Bloj, M.G., Kersten, D., and Hurlbert, A.C. (1999). Perception of three-dimensional shape influences colour perception through mutual illumination. Nature *402*, 877-879.

Brainard, D.H., Longere, P., Delahunt, P.B., Freeman, W.T., Kraft, J.M., and Xiao, B. (2006). Bayesian model of human color constancy. J Vis *6*, 1267-1281.

Conway, B.R. (2014). Color signals through dorsal and ventral visual pathways. Vis Neurosci *31*, 197-209.

Conway, B.R. (2015). Why do we care about the colour of the dress? The Guardian *http://www.theguardian.com/commentisfree/2015/feb/27/colour-dress-optical-illusion-social-media*.

Lafer-Sousa, R., Liu, Y.O., Lafer-Sousa, L., Wiest, M.C., and Conway, B.R. (2012). Color tuning in alert macaque V1 assessed with fMRI and single-unit recording shows a bias toward daylight colors. J Opt Soc Am A Opt Image Sci Vis *29*, 657-670.

McDermott, K.C., Malkoc, G., Mulligan, J.B., and Webster, M.A. (2010). Adaptation and visual salience. J Vis *10*, 1-32.

Rogers, A. (2015). The Science of Why No One Agrees on the Color of This Dress. Wired *http://www.wired.com/2015/02/science-one-agrees-color-dress/*.

Witzel, C., Valkova, H., Hansen, T., and Gegenfurtner, K.R. (2011). Object knowledge modulates colour appearance. iperception *2*, 13-49.

Zaidi, Q. (2011). Visual inferences of material changes: color as clue and distraction. WIREs Cognitive Science, John Wiley & Sons, Ltd. *Volume 2, November/December 2011*, 686-700.

3.1.2. Supplemental Figures

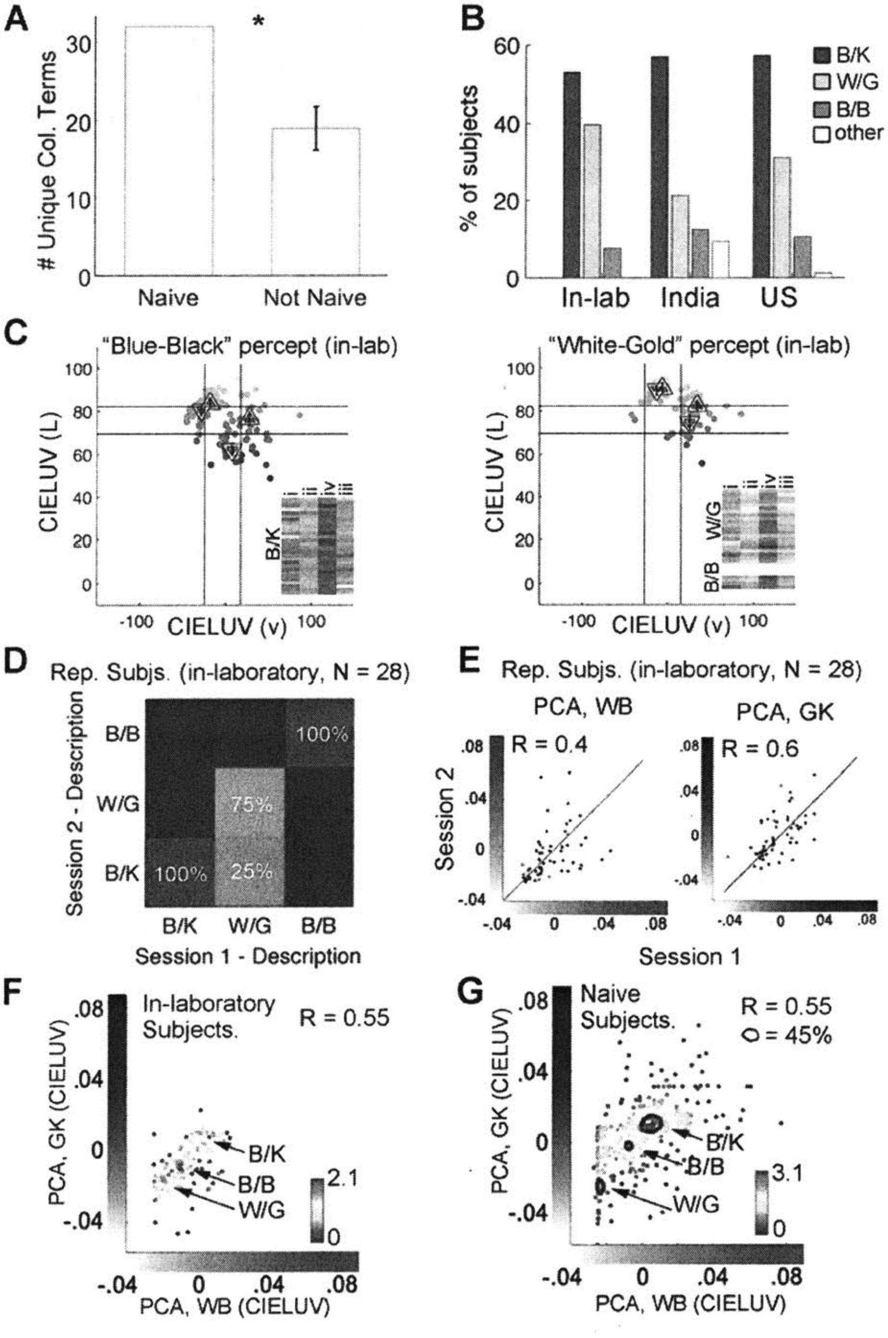


Figure S1. Analysis of the subset of subjects who performed the task in the laboratory under controlled viewing conditions (N=53), and naïve subjects tested on-line who reported having not seen the image before (N=311), also show striking individual differences in color perception of #TheDress that can be recovered both from the words used to describe the image and from the results of color matching. **(A)** Bar plot showing the number of unique color terms used by naïve (N=313) and non-naïve (N=1088 subjects).

To control for the number of subjects, the bar for the non-naïve subjects shows the mean of 10,000 subsampling boot-straps of the non-naïve population. Error bar shows the 95% C.I. of the mean. The naïve subject group used more unique color terms ($p < 0.0001$). **(B)** Histogram showing the color descriptions of the dress for the in- laboratory subjects (N = 53; 51 had seen the image previously), the subset of subjects from India who were tested online (N = 160; 63 had seen the image previously), and the subset of subjects from the United States who were tested online (N = 1147; 945 had seen the image previously). **(C)** Color matches made by control subjects reporting B/K or W/G, for regions in the dress (see i, ii, iii, iv, main Figure 1, panel A). White triangles show population averages (upward triangles for regions i and ii, and downward triangles for regions iii and iv, and contain the 95% C.I. of the mean. Insets depict color matches for individual subjects in each row, sorted by category (B/K, W/G, B/B). Subjects used a chin rest to control viewing angle and distance, and a calibrated display. **(D)** A subset of in-laboratory subjects (N = 28) participated in a test-re-test reliability assessment 3-5 days after participating in the main experiment. Heat map values indicate color descriptions on re-test as a function of how subjects originally reported the dress. Results are predominantly along the x=y diagonal, showing strong test-retest reliability. **(E)** In-laboratory-re- test subjects' color matches for regions (left shows white/blue regions (ii, iii); right shows gold/black regions (i, iv) from experiment 1 ("Session 1") plotted against matches for those same regions on re-test ("Session 2"). Note that the color terms "blue" and "brown" refer to the colors that most people assign to the pixels when the pixels are presented in isolation on a neutral background. See Supplementary Experimental Procedures for the definition of the axes. All data (from the entire study) in which subjects were asked to match a blue region in the image (ii, iii) contributed to the PCA used to define the axes in the left panel; and all data in which subjects were asked to match a brown region (i, iv) were used to define the axes in the right panel. The x=y diagonal is plotted for reference. If subjects show perfect reliability, the results should fall along the diagonal. Subjects whose responses changed in the color-description task are indicated in red. Re-test reliability was significant (regions ii, iii, R = 0.4, p = 0.002; ii, iv, R = 0.6, $p = 3*10^{-6}$). **(F)** In-laboratory subject color matches for regions (i and iv) plotted against matches for regions (ii and iii). The x value for each subject was defined as the weight along the first principal component for their color match; the y-axis values were determined similarly, using matches for the brown pixels of the image (i, iv, Figure 1A). Axes in all density plots shown in the present report (Figure 1F, S1G) are identical (R = 0.55, $p = 7*10^{-10}$). **(G)** Naïve subject color matches for 'brown' region (i) plotted against matches for neighboring 'blue' region (ii). Contours contain 45% of the respondents within each category (B/K, W/G, B/B), defined independently by the color descriptions provided by the subjects. Axes defined as for D. (R = 0.55, $p = 6*10^{-26}$).

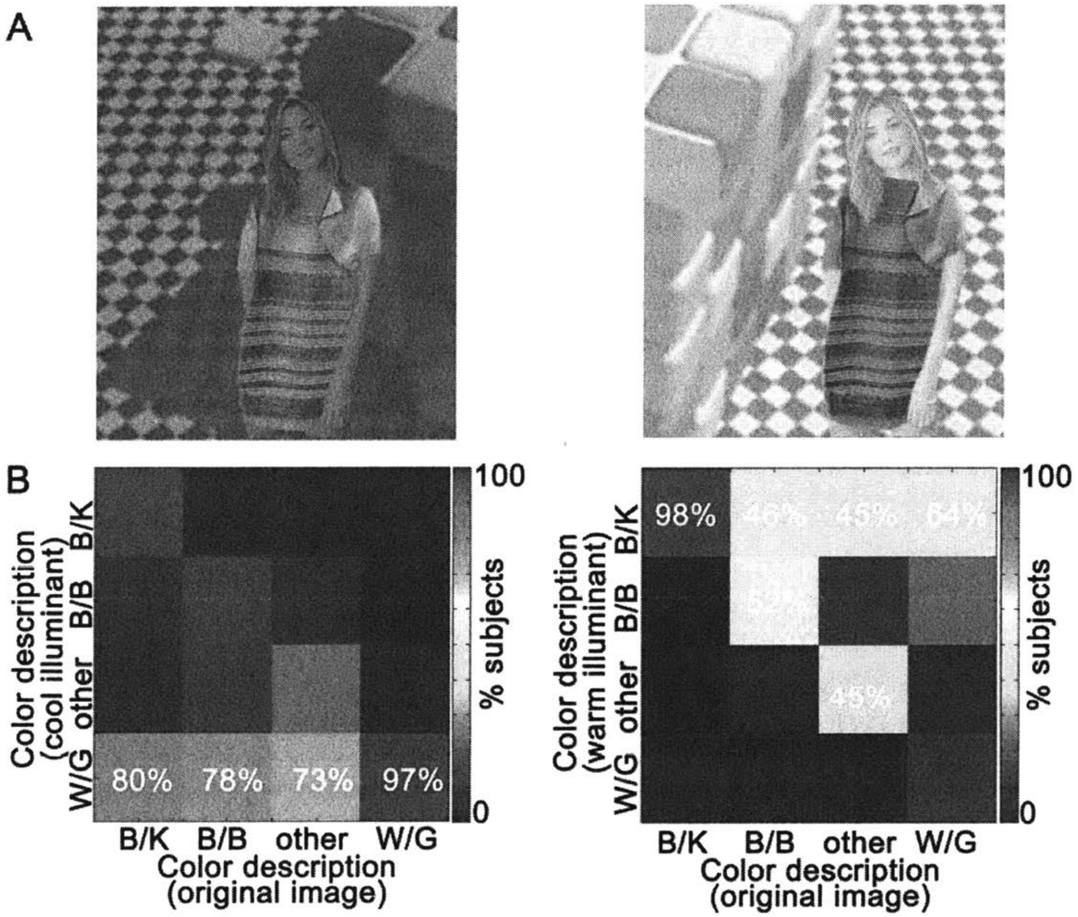


Figure S2. Embedding the image in a scene with overt cues to illumination causes most subjects to perceive the dress in a predictable way: W/G when seen under a cool illuminant; B/K when seen under a warm illuminant (N = 1127). Cues to the illuminant are provided by a global color tint applied to the whole scene, including the skin of the model. **(A)** The dress embedded within a scene simulating a cool illuminant (left), and the dress embedded within a scene simulating a warm illuminant (right) **(B)** Color descriptions under these conditions as a function of how subjects originally described the dress.

3.1.3. Supplemental Results

Results of logistic regression models described in **Supplemental Experimental Procedures**. Odds ratio [95% C.I.]; p calculated by Wald chi-square test.

	Model 1	**Model 2**	**Model 3**
Description	white/gold (success) versus blue/black response	main category (white/gold or blue/black, success) versus non-main category response	same as Model 2, but with "country" removed as a predictor variable
Intercept	0.31 [0.17-0.56]; **p < 0.0001***	20.86 [9.81-45.14]; **p < 0.0001***	15.50 [7.60-32.12]; **p < 0.0001***
Condition: Main experiment, order B (reference: main experiment order A)	1.15 [0.88-1.51]; p = 0.32	1.04 [0.72-0.49]; p = 0.84	1.03 [0.72-1.47]; p = 0.89
Condition: In-laboratory control condition (reference: main experiment order A)	1.46 [0.78-2.70]; p = 0.23	1.09 [0.41-3.75]; p = 0.88	1.33 [0.52-4.56]; p = 0.60
Condition: Scale experiment, 36% (reference: main experiment order A)	0.74 [0.52-1.04]; p = 0.08	1.21 [0.77-1.93]; p = 0.41	1.19 [0.76-1.89]; p = 0.46
F.lux software installed	0.75 [0.50-1.10]; p = 0.14	1.10 [0.67-1.91]; p = 0.71	1.11 [0.68-1.91]; p = 0.68
Male	0.75 [0.58-0.95]; **p = 0.019***	0.92 [0.66-1.28]; p = 0.63	0.89 [0.64-1.24]; p = 0.48
India (reference: USA)	0.84 [0.54-1.28]; p = 0.42	0.48 [0.31-0.76]; **p = 0.0016***	▨
Other country (reference: USA)	0.82 [0.43-1.52]; p = 0.55	6.48 [1.38-115.75]; p = 0.07	▨
Had seen image prior to study	1.21 [0.88-1.68]; p = 0.24	1.27 [0.86-1.86]; p = 0.22	1.53 [1.06-2.17]; **p = 0.021***
Age	1.02 [1.01-1.03]; **p = 0.0035***	0.97 [0.95-0.98]; **p < 0.0001***	0.97 [0.95-0.98]; **p < 0.0001***

Model 4

Description	white/gold (success) versus blue/black response; non-naïve participant group
Intercept	0.52 [0.30-0.91]; **p = 0.02***
Condition: Main experiment, order B (reference: main experiment order A)	1.09 [0.81-1.48]; p = 0.56
Condition: In-laboratory control condition (reference: main experiment order A)	1.54 [0.81-2.92]; p = 0.19
Condition: Scale experiment, 36% (reference: main experiment order A)	0.76 [0.51-1.12]; p = 0.17
F.lux software installed	0.81 [0.53-1.24]; p = 0.34
Male	0.74 [0.56-0.97]; **p = 0.03***
India (reference: USA)	0.83 [0.43-1.52]; p = 0.55
Other country (reference: USA)	0.85 [0.40-1.71]; p = 0.65
Had seen image switch colors prior to study	0.60 [0.46-0.79]; **p = 0.0003***
No response to switching question	0.97 [0.28-3.02]; p = 0.95
Age	1.01 [1.00-1.03]; **p = 0.032***

3.1.4. Supplemental Methods

The majority of participants were recruited and tested through Amazon's Mechanical Turk using a combination of template (Morris Alper's Turk Suite Template Generator 2014, available online at *http://mturk.mit.edu/template.php*) and custom HTML and JavaScript code. A smaller number of subjects (N=53) were recruited from the MIT University and Wellesley College campus through word of mouth and social media, and also tested using the M-Turk platform on a calibrated display in the laboratory. We adhered to the MIT Committee on the Use of Humans as Experimental Subjects policies on using Amazon's Mechanical Turk for research purposes. Informed consent was obtained for those subjects who performed the study in the laboratory. Procedures were approved by the institutional review board of Wellesley College.

We conducted three experiments: 1, a main experiment; 2, a follow-up experiment to assess the role of image size in determining what colors people report; 3, an in-laboratory, controlled experiment. Data were pooled from the various experiments depending on the analysis performed.

Subjects were between 18 and 69 years of age. To control for subject quality among the Mechanical Turk participants, we required that subjects have Mechanical Turk approval ratings of 95% or higher, and have completed at least 1000 human intelligence tasks (HITs) on Mechanical Turk previously.

Experiment 1: Main

Of the M-Turk participants who enrolled in the study, we rejected 70 as repeat participants. Of the remaining 1175, we rejected 20 subjects who indicated they were or might be colorblind, and 11 subjects who provided non-color word responses to our free response questions about dress color, leaving 1074 participants. Sample was gender balanced: $N_{males} = 563$, $N_{females} = 510$, $N_{other} = 1$. We estimate subjects spent approximately twenty minutes completing the survey (subjects who performed the tasks in the laboratory took about the same amount of time).

We assigned subjects to one of two conditions (A and B). In both conditions, we presented three images and asked questions about the dress color and lighting in each. The first image shown was always the same, but the order of the second two images differed between the two conditions. In addition, all subjects were queried on their perception of a blurry version of the first image.

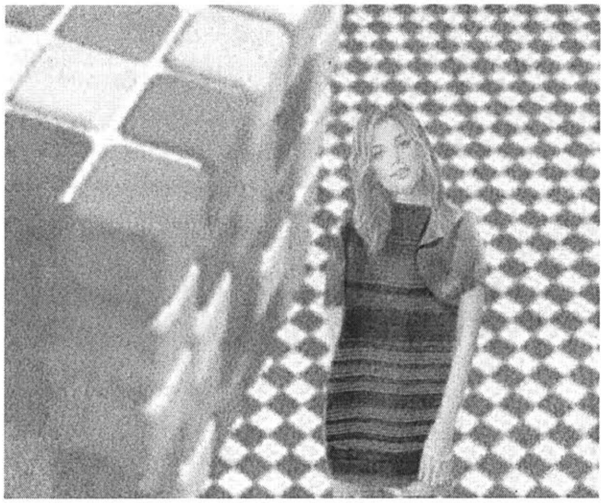
To ensure that stimuli were the same size across displays for on-line subjects, we specified the sizes of stimulus images in absolute pixels in the HTML experiment code. There is some variability from display to display in terms of the actual physical size of a pixel. We measured the images on a typical monitor in the laboratory, to provide a reasonable estimate for how the pixel values correspond to degrees of visual angle. These values are given in the legend of Figure 1, and below. We estimate that among different displays the variance in actual display size was ~+/- 10% of the size measured in laboratory.

The stimulus images were:

1) The original dress image that circulated on the internet ("original") was presented at 36% of its original size so that the entire image would be visible on all displays. We restricted the study to subjects who performed the tests on monitors large enough to display the entire 36% image; this precluded participants who might do the study using a hand-held device. The dress image was presented at an absolute size of 226 x 337 pixels. This corresponded to 7.2° of visual angle

width on the 18.65" wide iMac (21.5" diagonal) used for the in-laboratory experiment, using a 40-cm viewing distance.

2) A cut-out of the original dress superimposed on a picture of a model, embedded in a scene with a simulated *warm illuminant (*model was placed so as to appear in direct illumination); the dress portion of the stimulus was 76% of the size of the dress in the original image. The scene was provided, with permission, by Beau Lotto and Dale Purves. **Figure S2** shows a cropped version of what was shown. The complete picture was 518x429 pixels (~16.5° of visual angle on the horizontal axis) is shown below.



3) A cut-out of the original dress also superimposed on a picture of a model, but embedded in a scene with a simulated *cool illuminant (*model was placed in the shadow cast by a scene element). The scene was provided, with permission, by Beau Lotto and Dale Purves. **Figure S2** shows a cropped version of what was shown. The complete picture was 518x429 pixels (approximately 16.5° of visual angle on the horizontal axis), shown below.

The skin and hair of the model in (2) and (3) were tinted to reinforce the perception that the model is within a particular scene, illuminated by the same light source that is illuminating the rest of the scene. The chromaticity of skin-tint was matched to the overall illuminant (estimated from the "white" patches in the original Lotto cubes), and the lightness of the tint was varied to indicate either shadow or direct illumination. The dress in all three images (1, 2, and 3) was physically identical.

4) A blurry version of the original image was presented at 41% of the source size, with a Gaussian blur radius of 3.3 pixels (0.11°). The image was 8.3° of visual angle along the horizontal axis.

In condition A, subjects saw the cool-illuminated image (3) first and then the warm-illuminated image (2). In condition B, subjects saw image (2) first and then image (3). All subjects saw image (1) before seeing either (2) or (3). And all subjects saw image (4) last.

Each image was shown for fifteen seconds, and then we asked the subjects to make color and lighting judgments. The task was programmed so as to require that the subjects were exposed to the image for at least 15 seconds before they could answer the questions. The image was on the screen continuously while the subjects answered the questions (this was not designed as a test of color memory). Between the presentation of the first and second images, we collected basic demographics information. Between the presentations of the last two images, we asked subjects about the environment in which they were completing the study. We also asked subjects whether they had viewed the original dress image prior to this study, and if so, whether they had experienced multiple percepts of the dress (i.e. "switching"). We inquired about the rate of switching of the dress colors in pre-study viewings of the image. The full survey is given in the Appendix below.

Experiment 2: Scale

To explore the effect of image size on percept, we performed a second experiment in which subjects viewed the original dress image at one of four sizes, defined as a % of the original: 10% (63 x 94 pixels, or 2.0° of visual angle on a 21.5-inch iMac), 36% (the image size we presented in our main experiment; 226 x 337 pixels, or 7.2° of visual angle), 100% (628 x 937 pixels, or 20.0° of visual angle; and 150% (942 x 1405 pixels, or 30.0° of visual angle). The stimuli in the 10% and 36% conditions fit fully in all browser windows. For the 100% and 150% images, only part of the image was visible at once in the height direction, but was completely visible in the horizontal dimension. As a result, subjects had to scroll over the stimulus image, from top to bottom, to view it. Scrolling to see the entire image was required in order for subjects to access the buttons to move through the study, ensuring that all subjects saw the entire image even in the 100% and 150% displays. The rest of the scale experiment consisted of a subset of the questions included in the main experiment; see full surveys given in the Appendix below.

The same subject screening and rejection criteria were used as in the main experiment. In addition, we rejected subjects who had previously participated in the main experiment, leaving N = 1127 (N = 288 in 10% condition, N = 274 in 36% condition, N = 292 in 100% condition, and N = 273 in 150% condition). In the analysis of image size, we pooled results from the main and in- laboratory experiment (described in the next section) to estimate the fraction of people who saw the dress as W/G at the 36% size (see **Figure 1G**).

In-laboratory experiment

One possibility is that the individual differences reported for the dress image are simply the result of differences in the conditions in which different subjects viewed the image, rather than differences in how the brains of different people assign colors to the same stimulus. To test the extent to which individual differences observable in our main experiment were also present in subjects who performed the survey under controlled viewing conditions on the same display, we tested 53 subjects ($N_{males} = 22$, $N_{females} = 31$) in laboratory. Subjects viewed the display at a distance of 40cm. We re-tested 28 of these subjects 3-5 days after their first experimental session in order to estimate intra-subject repeat reliability.

All in-laboratory subjects performed the experiment through the same Mechanical Turk interface used by subjects in the main experiment. The experiment was performed on a calibrated 21.5-inch iMac computer with 1920 x 1080 pixel resolution in a windowless room with LED overhead lighting (C.I.E. xyY = .4814 .4290 4.3587, measured off the MacBeth color checker standard white, held at the same location and viewing angle on the monitor at which we presented the dress image). Normal color vision was confirmed with Ishihara plates.

Analysis Methods

The analysis of free response and color-matching (**Figure 1,C-F**) was conducted using the combined data from the main experiment (N = 1074), the 36% condition of the scale experiment

(N = 274), and the in-laboratory control experiment (N = 53) (pooled N = 1401; sample was gender balanced: $N_{males} = 731$, $N_{females} = 668$, $N_{other} = 2$). Of the pooled data (online and in-laboratory), 1190 subjects reported being from the US (1147 were online subjects), 161 from India (160 were online subjects), and 50 were from another country (41 were online).

Logistic regression models

To identify variables that significantly predicted subjects' free response answers (for the original image, image (1) described above), we created logistic regression models of two binary pairs of interest: reports of white/gold (success) versus blue/black (N = 1221; "Model 1") and reports of white/gold or blue/black (success) versus any other response (N = 1399; "Model 2"). Models were implemented in the R statistical software package using the function "glm" with the logit link function. In both models, we included the following predictor variables: study condition (main experiment order condition A, main experiment condition B, scale experiment 36% condition, in-laboratory experiment), gender (male or female; we removed 2 subjects who indicated "other" for the purposes of the models), country (United States, India, or other), age, whether subject reported having viewed the original dress image prior to this study ("prior view", naïve or non-naïve), and whether the subject reported having the software f.lux installed on their computer. P values for predictors were calculated using the Wald chi-square test. See **Table S1** for model results.

In addition to the full group models, we separately modeled the responses of the subset of subjects who had seen the original image prior to our study and responded with a main category (N = 962; "Model 4" of white/gold [success] versus blue/black responses). We included as predictor variables study condition, gender, country, age, f.lux, and whether subject reported having seen the image change colors at any point prior to our study (**Table S1**).

Logistic regression results documenting the impact of previous view

Overall, 57.0% (95% C.I. = [54.4%-59.5%]) of subjects described the dress as blue/black (B/K); 30.3% [28.0%-32.8%] as white/gold (W/G); 10.6% [9.1%-12.4%] as blue/brown (B/B); and 2.1% [1.4%-3.0%] gave some other description. Non-naïve subjects used a smaller number of unique color terms than naïve subjects (Figure 1C; **Figure S1A**, $p < 0.0001$, one-sample T-test). Residents of India were less likely than residents of the US to use a main category (B/K, W/G) rather than B/B or other ($p = 0.0016$, Wald chi-square; OR=0.48, 95% C.I.=[0.31-0.76]; **Table S1, Figure S1B**). The Indian population contained relatively more naïve subjects than the US population (60% versus 17%). To explore the possibility that country masked an important difference between naïve and non-naïve participants, we created "Model 3," in which we included all original predictors except for country. When country was removed from the logistic regression model ("Model 3"), prior experience became a significant predictor: non-naïve subjects were more likely to choose B/K or W/G, over B/B or other ($p = 0.021$, Wald chi-square; OR= 1.53 [1.06-2.17]) (see **Table S1**). The description B/B (or other) would seem to be categorically different from the descriptions W/G and B/K insofar as B/B comprises two chromatic colors (as opposed to a chromatic color and an achromatic term). That subjects with prior experience of the image are less likely to report B/B is consistent with the conclusion that some subjects actually undergo a change in their perception of the dress following experience.

Together, these results suggest that the media coverage of the image (#theDress) changed the language people used to describe the image and possibly their perception of it.

Size experiment

In **Figure 1D**, we show the percentage of main category respondents who responded white-gold as a function of age. Data was based on subject responses in each condition of the scale experiment, and pooled with the main and in-laboratory experiment data for the 36% image size data point. To explore the effects of image size on percept, we created a logistic regression model of white/gold versus blue/black responses; we included image size as the only predictor.

Confidence intervals

Unless otherwise noted, all reported 95% confidence intervals were calculated analytically in the R statistical package using the modified Wald method [S1].

Analysis of individual differences: variability in lightness (L) and chromaticity (u,v) of color-matching data

Subjects made color matches using the color picker tool from the Raphael JavaScript library (http://raphaeljs.com/license.html). This tool allowed subjects to freely vary the hue, saturation and lightness to obtain the best match. Color matches were recorded as the HEX values generated by the tool, and converted to Luv [S2]. We chose the Luv color space for the analysis because this space is more-or-less perceptually uniform. The color match made by a subject for a given patch of the dress (i, ii, iii or iv, Figure 1A) was defined in the three dimensions of Luv color space. **Figure 1E** and **Figure S1C** show color matches sorted by the description given by the subject, B/K or W/G, for regions in the dress (see i, ii, iii, iv, **Figure 1A**). (Color matches for remaining subjects not shown, except in insets, **Figure 1E, Figure S1C**.) White triangles show population averages (upward triangles for regions i and ii; downward triangles for regions iii and iv; symbols contain the 95% C.I. of the mean). Insets depict color matches for individual subjects in each row, sorted by category (B/K, W/G, B/B, other). The 95% C.I. of the mean was defined as: mean + (T-score x standard error).

For a given region, the average color match made by W/G perceivers differed in both lightness (L)and hue (u, v) from the match made by B/K perceivers (Figure 1E). The magnitude of difference in the lightness dimension (L) was 20, 13.5, 13.3, and 20.7 (for regions i, ii, iii, and iv respectively), and this difference was significant for each region (p values < 0.0001, t-test). The magnitude of difference in the 'u' chromaticity dimension was 9.2, 6.6, 6.5, and 9.9 (for regions i, ii, iii, and iv respectively), and this difference was significant for each region (p values <

0.0001, t-test). The magnitude of difference in the 'v' chromaticity dimension was 23.9, 24.6, 26.6, and 25 (for regions i, ii, iii, and iv respectively), and this difference was significant for each region (p values < 0.0001, t-test).

The same analysis was also conducted separately for subsets of subjects: in-laboratory and online. The differences in magnitude of the lightness dimension and the chromaticity dimensions (u, v) were significant for the online data (and comparable in magnitude to the numbers provided above). The differences in magnitude of the lightness dimension and the chromaticity dimensions of the in-lab data were also significant (although the absolute magnitudes were smaller): The magnitude of difference in the lightness dimension (L) was 5.8, 6.8, 9.1, and 12.4 (for regions i, ii, iii, and iv respectively), and this difference was significant for each region (ps < 0.01, t-test). The magnitude of difference in the 'u' chromaticity dimension was 3, 6, 6.7, and 4.5 (for regions i, ii, iii, and iv respectively), and this difference was significant for all regions except for 'i' (ps < 0.02, t-test). The magnitude of difference in the 'v' chromaticity dimension was 7.2, 12.5, 16.2, and 18.5 (for regions i, ii, iii, and iv respectively), and this difference was significant for each region (ps < 0.01, t-test). Differences in the demographics of the on-line and in-laboratory subjects may partially account for these differences (the in-laboratory group was more homogenous, mostly undergraduates and graduate students).

Principal component analysis (PCA) for color-matching data and test-retest reliability

The color match made by a subject for a given patch of the dress (i, ii, iii or iv, **Figure 1A**) was defined in the three dimensions of Luv color space. In order to obtain a single metric for the color match made by each subject, for a given patch in the dress, we performed PCA on the 3-D cloud of data points for the entire population. We ran two principal component analyses: first, using all data obtained from color matches to the "blue" pixels (ii, iii, **Figure 1A**); second, using all data obtained from color matches to the "brown" pixels (i, iv, **Figure 1A**). Note that the color terms "blue" and "brown" refer to the colors that most people assign to the pixels when the pixels are presented in isolation on a neutral background. The first (white to blue, "WB") component of the PCA produced a vector in the 3-D color space that captured most (86.4%) of the variance in the data. The second vector (gold to black, "GK") also captured a sizable amount of the variance (71.2%). The (x,y) values for each subject in the density plots shown in **Figure 1F**, **Figure S1D, E, F, and G** were defined as the weight of that subject's color match along the first principal components along the WB and GK vectors. Conducting the PCA on the entire set of color matches enabled us to define a common set of axes on which to display color-matches made by the different populations (naïve, non-naïve, in-laboratory). Analyzing the results using PCA conducted on each population separately (not shown) did not alter the main conclusions. When defined only using data collected in laboratory, the WB component explained 91.6% of the variance, while GK component accounted for 77.3% of the variance. Using the PCA weights, we were able to: (1) test the reliability of the color matching in a subpopulation (N=28/53) that was re-tested in the laboratory (testing sessions were 3-5 days apart); (2) assess the extent to which color matches for 'blue' regions predicted matches to 'brown' regions; (3) assess the extent to which the distribution of percepts in the population was continuous or categorical.

(1) Test-retest reliability:

We sought to assess the reliability of the individual differences, to compare with individual differences measured in other circumstances [S3]. The challenge posed by this goal, however, is that the image of the dress is not only seen differently by different people, but also seen differently by the same person at different points of time –i.e. the image is multistable for some subjects. We re-tested about half (N=28) of the in-laboratory subjects in a separate session, 3-5 days after the first testing session. **Figure S1D** shows the intra-subject reliability of color descriptors used in sessions 1 and 2 (11% of subjects switched between sessions). We also compared the color matches that these subjects made for the identical patches seen during the two sessions (**Figure S1E**). Because of multistability, the intra-subject reliability measure provides a low estimate for the true reliability. Nonetheless, the reliability scores were significant (linear correlation: regions ii, iii, $R = 0.4$, $p = 0.002$; regions ii, iv, $R = 0.6$, $p = 3*10^{-6}$).

(2) Color matches for one region predict another:

Figure 1F shows a density plot in which we plot the color matches (PCA weights) for 'brown' region i versus 'blue' region ii for all subjects irrespective of the color description they used for the dress (**Figure S1G** shows the naïve subjects only). Comparing color-match judgments for neighboring regions (i vs ii) controlled for subtle shifts in illumination across the image. Importantly, the results and conclusions did not change when we plotted iii vs iv, or when we plotted both 'brown' regions (i and iv) against both 'blue' regions (ii and iii) (data not shown). In the case of the in-laboratory data, we did not have sufficient data to plot only regions i and ii, so **Figure S1F** shows a density plot in which we plot the color matches for both 'brown' regions (i and iv) against those for both 'blue' regions (iii and iv).

In each density plot, the color scale bar shows the number of subjects at each location in the plot. The use of a density plot is necessary because with over 1400 subjects, data points for many subjects overlap. The maximum color in the scale bar is in **Figure 1F** 2.3 subjects (2.1 in **Figure S1F**, and 3.1 in **Figure S1G**). In addition, the plots were smoothed to reveal any underlying structure. The smoothing was a Gaussian of radius 0.0068 PC units in Figure 1F (0.0027 in **Figure S1F**; 0.0059 in **Figure S1G**). Importantly, the color-matching data are obtained independent of the color descriptions used by the subjects. The density plots clearly show two features: first, color matches to region i are highly predictive of color matches to region ii (**Figure 1F**, $R=0.59$; $p=10^{-264}$; **Figure S1F**, $R = 0.55$, $p = 7*10^{-10}$; **Figure S1G**, $R = 0.55$, $p = 6*10^{-26}$, where the correlation in **Figure S1G** shows how well color matches for both 'blue' regions (ii and iii) predict matches for both 'brown' regions (i and iv)). And second, there is inhomogeneity along the line of maximum correlation: showing evidence for three distinct clusters, discussed next.

(3) Peaks in color match density plots correspond to the three main categories of color descriptions:

We sought to determine the extent to which these clusters in the density plots correspond to the three main groups of subjects, those reporting B/K, W/G or B/B. Using the independent color description data provided by the subjects, we sorted the color matches into four sub- populations defined by main-category description (W/G, B/K, B/B, and other). We then generated a density plot for each sub-population and determined contours that captured the peak density (containing at least 25% of the data). The black squiggle contours in **Figure 1D and Figure S1G** show these contours superimposed on the density plot obtained for the whole population. The contours are coincident with the three clusters in the population data, showing that the color-matching data uncovers the three main descriptive reports. **Figure S1F** does not have squiggles because the N is low; nonetheless, that panel also shows three peaks.

RGB (HEX) to Luv conversion

The white point used to convert RGB (HEX) to CIE values (**Figure 1B**) and Luv values (**Figure 1E,F**) was the standard CIE 1931 F7 illuminant typical of many displays [95.047 100 108.883, XYZ] (Westland, p 65). The white point for the monitor used to collect data in the laboratory was different [34.49 37.67 35.03, XYZ], but we analyzed the in-lab data using the F7 illuminant to facilitate comparison with the MTurk data (**Figure S1C,E,F,G**). The conclusions are no different if we analyze the in-lab data using the lab-specific white point (data not shown). Similarly, when the in-lab white point was defined using the measured mean of the dress image [16.38 16.6 19.41, XYZ], the conclusions remained the same (data not shown).

Histograms for unique color terms

Subjects who had previously seen the image used a smaller number of unique color terms than naïve subjects ($p<0.0001$, one-sample T-test). In Figure S1A we show the number of unique color terms used by the subset of subjects who had never seen the dress image before the study ("Naïve", $N = 313$) and the subset of subjects who had seen the image previously ("Not-naïve", $N= 1088$). Both populations show evidence for three dominant categories: B/K, W/G and B/B. But there are subtle differences in the number of unique color terms used and, as described above, subjects' perception of the dress. In order to quantify the number of unique terms, controlling for the difference in absolute number of subjects in each population, we subsampled the not-naïve population, without replacement 10,000 times. The number of samples drawn on each subsample was 313, and the number of unique color terms was tabulated. A one-sample t-test was performed to determine whether the number of unique color terms derived from the not-naïve population could have come from a distribution whose mean was equal to the number of color terms used by the naïve population ($p < 0.0001$).

Many of the unique color terms appeared to refer to the same perception. For example, "white-golden", "white-goldish", and "white-gold" were deemed redundant. Color descriptions deemed redundant were binned. Four categories emerged: Blue/Black (B/K), White/Gold (W/G), Blue/Brown (B/B), or Other. These categories were distinguished as "Main" categories (B/K or W/G) or not-main categories (B/B or other). This distinction reflects the use of one achromatic and one chromatic color term for the Main categories, as opposed to the use of two chromatic color terms for the not-Main category.

The complete list of unique color terms, and how they were binned is as follows:

Blue/black category: 'blue, black', 'blue, gray', 'light blue, gray', 'light blue, black', 'pale blue, black', 'blue, dark gray', 'dark blue, black', 'deep periwinkle blue, black', 'light blue, light black', 'periwinkle, dark gray', 'black, navy blue'

White/gold category: 'white, gold', 'white, brown', 'white, golden', 'white, tan', 'white, yellow', 'copper, white', 'gold, off-white', 'gold, white (white light blue)', 'gray, brown', 'gray, gold', 'white, golden brown', 'white, khaki', 'white, chocolate', 'white, sandy brown'

Blue/brown category: 'blue, brown', 'blue, gold', 'brown, light blue', 'gold, light blue', 'blue, golden', 'blue, tan', 'blue, dark bronze', 'blue, dark brown', 'dark brown, light blue', 'gold, baby blue', 'light blue, brownish gold', 'periwinkle, gold', 'sky blue, dark brown', 'sky blue, brown', 'dark blue, golden black', 'white-blue, gold', 'blue, bronze', 'blue, gold/black'

Other category: 'purple, brown', 'purple, black', 'ash, black', 'blue, olive', 'golden, lilac', 'lavender, black', 'lavender, gold', 'light violet, black', 'pink, black', 'purple, gold', 'purple, tan', 'violet, black', 'violet, gold', 'white, cream', 'yellow, gold', 'white, pink', 'purple green, light blue', 'gray, black', 'blue, white', 'white, black', 'gold, black'

Figure S1B shows the breakdown of color term usage for in-lab, India (online), and US (online) subjects. The distributions are similar, but the population from India shows more non-main-category terms. Consistent with the hypothesis that experience impacts the likelihood of using a main-category descriptor, the Indian population contained relatively more naïve subjects than the US population (60% versus 17%).

Compelling a particular perception of the dress (W/G or B/K) by embedding it in a scene, on a model

We surveyed color descriptions for the dress embedded two scenes with different simulated lighting conditions (image 2 and 3 described above; N = 1127, 1047 from the main experiment and 53 from the in-laboratory controlled experiment) (**Figure S2**). Of subjects who described the dress originally as white/gold, 64% switched to seeing it as blue/black under the simulated warm light; of those who saw the dress as blue/black originally, 80% switched to seeing it as white/gold under the simulated cool light. Subjects who described the original image as blue/brown (or blue/gold) were just as likely to maintain their description when shown the dress under the warm illuminant, consistent with the idea that they are doing a local pixel-color analysis (or assuming a neutral illuminant) (although these subjects were more likely to switch to white/gold given the cool illuminant).

3.1.5. Supplemental Materials

The full surveys from the main experiment and the scale experiment are provided below.

MAIN EXPERIMENT:

Page 1

In this study, we will show you a number of images. We want you to look at these images carefully, and then answer some questions about them. If you're not sure about an answer, take your best guess. Please answer all questions before proceeding to the next page.
Please try to complete the study in a single sitting, without taking breaks.

Page 2

Stimulus: dress on original background (presented for 15 seconds)

Please look carefully at the dress. A "continue" button will appear just below this text after 15 seconds.

Page 3

Stimulus: dress on original background

This is a picture of a____and____dress. (Fill in the colors that you see

Two textboxes were provided.

Page 4

*Stimulus: dress on original background, with arrows pointing to regions labeled A, B, C, and D***

***regions A, B, C, and D noted here correspond to Figure 1A regions i, ii, iii, and iv, respectively*

Now we would like you tell us about the material properties (fabric) of the dress. For each of the indicated regions (A, B, C, and D) in the diagram above, what kind of material do you think the dress is made of? Below are some examples of each category. Take your best guess if you're not sure.

A panel showing gray-scale images of fabric swatches appeared at the center of the screen.

*1. How would you describe the material at (A)?

Pull-down menu with options as follow:

o semi-transparent, sheer; chiffon, tulle, organza

- flat matte, low-luster, low-sheen; cotton, linen, jersey, matte-satin
- semi-gloss, med-luster, med-sheen; rayon, polyester
- hi-gloss, hi-sheen; silk, glossy-satin

*2. How would you describe the material at (B)?
Pull-down menu with options as follow:
- semi-transparent, sheer; chiffon, tulle, organza
- flat matte, low-luster, low-sheen; cotton, linen, jersey, matte-satin
- semi-gloss, med-luster, med-sheen; rayon, polyester
- hi-gloss, hi-sheen; silk, glossy-satin

*3. How would you describe the material at (C)?
Pull-down menu with options as follow:
- semi-transparent, sheer; chiffon, tulle, organza
- flat matte, low-luster, low-sheen; cotton, linen, jersey, matte-satin
- semi-gloss, med-luster, med-sheen; rayon, polyester
- hi-gloss, hi-sheen; silk, glossy-satin

*4. How would you describe the material at (D)?
Pull-down menu with options as follow:
- semi-transparent, sheer; chiffon, tulle, organza
- flat matte, low-luster, low-sheen; cotton, linen, jersey, matte-satin
- semi-gloss, med-luster, med-sheen; rayon, polyester
- hi-gloss, hi-sheen; silk, glossy-satin

Page 5
Stimulus: dress on original background

*1. Does the lighting on the dress look different than the lighting in the background?
Pull-down menu with options as follow:
- Yes
- No

*2. The lighting in the background is...Check all that apply:

Check boxes appeared next to the following options:
- warm
- bright
- cool
- dim
- dark
- yellowish
- bluish
- reddish
- greenish
- purplish
- glaring
- iridescent
- washed out

- blown out

*3. On a scale from 1 to 5, where 1 is cool and 5 is warm, please rank the lighting conditions in the background.
Pull-down menu with options as follow:
- 1
- 2
- 3
- 4
- 5

*4. The light illuminating the dress is...Check all that apply:
Check boxes appeared next to the following options:
- warm
- bright
- cool
- dim
- dark
- yellowish
- bluish
- reddish
- greenish
- purplish
- glaring
- iridescent
- washed out
- blown out

*5. On a scale from 1 to 5, where 1 is cool and 5 is warm, please rank the light illuminating the dress.
Pull-down menu with options as follow:
- 1
- 2
- 3
- 4
- 5

Page 6 DEMOGRAPHICS

Please help us out by answering the following questions about yourself. You must answer all of the questions marked with an asterisk (*).

*1. What is your gender?
Pull-down menu with options as follow:
- Male
- Female
- Other

*2. How old are you?
Pull-down menu with options 18-100

*3. Is English your mother tongue?
Pull-down menu with options as follow:

- Yes
- No

4. If English is not your mother tongue, what is your mother tongue?
A pull-down menu contained a comprehensive list of languages, including Tamil and other languages spoken in India. List of languages available on request.

*5. What country are you from?
A pull-down menu contained a comprehensive list of countries. List available on request.

5 (A). If you're from the United States, what part of the country are you from?
Pull-down menu with options as follow:

- Northeast - New England
- Northeast - Mid-Atlantic
- Midwest - East North Central
- Midwest - West North Central
- South - South Atlantic
- South - East South Central
- South - West South Central
- West - Mountain

*6. Which levels of education have you completed? Check all that apply.
Check boxes appeared next to the following options:

- Elementary school (primary school)
- High school
- College/university
- Advanced graduate work

*7. Do you have normal or corrected-to-normal vision?
Pull-down menu with options as follow:

- Yes
- N

*8. Are you colorblind?

Pull-down menu with options as follow:

- Yes
- No
- Unsure

*9. On most days, what time do you typically wake up
Two pull-down menus appeared: one containing numbers 1-12, and one containing "A.M. (morning)" and "P.M. (night)"

*10. On most days, what time do you typically go to bed?
Two pull-down menus appeared: one containing numbers 1-12, and one containing "A.M. (morning)" and "P.M. (night)"

11. Do you have a background in the visual arts? Check all that apply:
Check boxes appeared next to the following options:

○ Professionally
○ Hobby
○ Painter
○ Graphic designer
○ Art historian
○ Architecture
○ Illustrator
○ Photographer

Page 7

Stimulus: dress on first colored background (warm or cool illuminant depending on condition; presented for 15 seconds)

Please look carefully at the dress. A "continue" button will appear just below this text after 15 seconds.

Page 8

Stimulus: dress on first colored background

This is a picture of a____and____dress. (Fill in the colors that you see

Two textboxes were provided.

Page 9

Stimulus: dress on first colored background, with arrows pointing to regions labeled A, B, C, and D

Now we would like you tell us about the material properties (fabric) of the dress. For each of the indicated regions (A, B, C, and D) in the diagram above, what kind of material do you think the dress is made of? Below are some examples of each category. Take your best guess if you're not sure.

A panel showing gray-scale images of fabric swatches appeared at the center of the screen.

*1. How would you describe the material at (A)?

Pull-down menu with options as follow:

○ semi-transparent, sheer; chiffon, tulle, organza
○ flat matte, low-luster, low-sheen; cotton, linen, jersey, matte-satin
○ semi-gloss, med-luster, med-sheen; rayon, polyester
○ hi-gloss, hi-sheen; silk, glossy-satin

*2. How would you describe the material at (B)?

Pull-down menu with options as follow:

○ semi-transparent, sheer; chiffon, tulle, organza
○ flat matte, low-luster, low-sheen; cotton, linen, jersey, matte-satin
○ semi-gloss, med-luster, med-sheen; rayon, polyester
○ hi-gloss, hi-sheen; silk, glossy-satin

*3. How would you describe the material at (C)?

Pull-down menu with options as follow:

- semi-transparent, sheer; chiffon, tulle, organza
- flat matte, low-luster, low-sheen; cotton, linen, jersey, matte-satin
- semi-gloss, med-luster, med-sheen; rayon, polyester
- hi-gloss, hi-sheen; silk, glossy-satin

*4. How would you describe the material at (D)?
Pull-down menu with options as follow:
- semi-transparent, sheer; chiffon, tulle, organza
- flat matte, low-luster, low-sheen; cotton, linen, jersey, matte-satin
- semi-gloss, med-luster, med-sheen; rayon, polyester
- hi-gloss, hi-sheen; silk, glossy-satin

Page 10
Stimulus: dress on first colored background

*1. Does the lighting on the dress look different than the lighting in the background?
Pull-down menu with options as follow:
- Yes
- No

*2. The lighting in the background is...Check all that apply:
Check boxes appeared next to the following options:
- warm
- bright
- cool
- dim
- dark
- yellowish
- bluish
- reddish
- greenish
- purplish
- glaring
- iridescent
- washed out
- blown out

*3. On a scale from 1 to 5, where 1 is cool and 5 is warm, please rank the lighting conditions in the background.
Pull-down menu with options as follow:
- 1
- 2
- 3
- 4
- 5

*4. The light illuminating the dress is...Check all that apply:
Check boxes appeared next to the following options:
- warm

- bright
- cool
- dim
- dark
- yellowish
- bluish
- reddish
- greenish
- purplish
- glaring
- iridescent
- washed out
- blown out

*5. On a scale from 1 to 5, where 1 is cool and 5 is warm, please rank the light illuminating the dress.
Pull-down menu with options as follow:

- 1
- 2
- 3
- 4
- 5

Page 11
Tell us a bit about your environment.

*1. What device are you using to view this webpage?
Pull-down menu with options as follow:

- Laptop
- Computer monitor (flat screen
- Computer monitor (rounded screen
- Tablet
- Other

*2. How big is the screen on which you're viewing this webpage? (Take your best guess if you're not sure.)
Pull-down menu with options as follow:

- 10-13 inches wide
- 14-17 inches wide (regular laptop)
- 14-17 inches wide
- More than 17 inches wide

*3. Are you using a Mac or PC?
Pull-down menu with options as follow:

- Mac
- PC
- Unsure

*4. What time is it right now? (select the approximate hour)

Two pull-down menus appeared: one containing numbers 1-12, and one containing "A.M. (morning)" and "P.M. (night)"

*5. Are you indoors or outdoors?
Pull-down menu with options as follow:
- o Indoors
- o Outdoors

*6. If you're indoors, is there natural light entering the room through a window?
Pull-down menu with options as follow:
- o N/A
- o Yes
- o No

*7. What are the lighting conditions where you are right now, on a scale of 1 (darkest) to 10 (brightest)?
Pull-down menu with options as follow:
- o 1 (darkest; as dim as a room without lights at night
- o 2
- o 3
- o 4
- o 5
- o 6
- o 7
- o 8
- o 9
- o 10 (brightest; as bright as the beach on a sunny day

*8. Is the software f.lux installed on the computer you're using to do this HIT?
Pull-down menu with options as follow:
- o Yes
- o No

Page 12

Stimulus: dress on second colored background (warm or cool illuminant depending on condition; presented for 15 seconds)

Please look carefully at the dress. A "continue" button will appear just below this text after 15 seconds.

Page 13

Stimulus: dress on second colored background

This is a picture of a____and____dress. (Fill in the colors that you see
Two textboxes were provided.

Page 14

Stimulus: dress on second colored background, with arrows pointing to regions labeled A, B, C, and D

Now we would like you tell us about the material properties (fabric) of the dress. For each of the indicated regions (A, B, C, and D) in the diagram above, what kind of material do you think the dress is made of? Below are some examples of each category. Take your best guess if you're not sure.

A panel showing gray-scale images of fabric swatches appeared at the center of the screen.

*1. How would you describe the material at (A)?
Pull-down menu with options as follow:
o semi-transparent, sheer; chiffon, tulle, organza
o flat matte, low-luster, low-sheen; cotton, linen, jersey, matte-satin
o semi-gloss, med-luster, med-sheen; rayon, polyester
o hi-gloss, hi-sheen; silk, glossy-satin

*2. How would you describe the material at (B)?
Pull-down menu with options as follow:
o semi-transparent, sheer; chiffon, tulle, organza
o flat matte, low-luster, low-sheen; cotton, linen, jersey, matte-satin
o semi-gloss, med-luster, med-sheen; rayon, polyester
o hi-gloss, hi-sheen; silk, glossy-satin

*3. How would you describe the material at (C)?
Pull-down menu with options as follow:
o semi-transparent, sheer; chiffon, tulle, organza
o flat matte, low-luster, low-sheen; cotton, linen, jersey, matte-satin
o semi-gloss, med-luster, med-sheen; rayon, polyester
o hi-gloss, hi-sheen; silk, glossy-satin

*4. How would you describe the material at (D)?
Pull-down menu with options as follow:
o semi-transparent, sheer; chiffon, tulle, organza
o flat matte, low-luster, low-sheen; cotton, linen, jersey, matte-satin
o semi-gloss, med-luster, med-sheen; rayon, polyester
o hi-gloss, hi-sheen; silk, glossy-satin

Page 15
Stimulus: dress on second colored background

*1. Does the lighting on the dress look different than the lighting in the background?
Pull-down menu with options as follow:
o Yes
o No

*2. The lighting in the background is...Check all that apply:

Check boxes appeared next to the following options:
o warm
o bright
o cool

- dim
- dark
- yellowish
- bluish
- reddish
- greenish
- purplish
- glaring
- iridescent
- washed out
- blown out

*3. On a scale from 1 to 5, where 1 is cool and 5 is warm, please rank the lighting conditions in the background.

Pull-down menu with options as follow:

- 1
- 2
- 3
- 4
- 5

*4. The light illuminating the dress is...Check all that apply:

Check boxes appeared next to the following options:

- warm
- bright
- cool
- dim
- dark
- yellowish
- bluish
- reddish
- greenish
- purplish
- glaring
- iridescent
- washed out
- blown out

*5. On a scale from 1 to 5, where 1 is cool and 5 is warm, please rank the light illuminating the dress.

Pull-down menu with options as follow:

- 1
- 2
- 3
- 4
- 5

Page 16

Stimulus: dress on original background, with arrow pointing to region (A). Color-picker appeared to the right of the image.

Now we would like you to report on the color of the pixels you see in the image. We will provide a color-matching tool. Please adjust the hue (color circle) and brightness (slider bar) to match the pixels you see in the image.

*1. What color is the dress at (A)?

Page 17

Stimulus: dress on original background, with arrow pointing to region (B). Color-picker appeared to the right of the image.

Now we would like you to report on the color of the pixels you see in the image. We will provide a color-matching tool. Please adjust the hue (color circle) and brightness (slider bar) to match the pixels you see in the image.

*2. What color is the dress at (B)?

Page 18

Stimulus: dress on original background, with arrow pointing to region (C). Color-picker appeared to the right of the image.

Now we would like you to report on the color of the pixels you see in the image. We will provide a color-matching tool. Please adjust the hue (color circle) and brightness (slider bar) to match the pixels you see in the image.

*3. What color is the dress at (C)? Page 19

Stimulus: dress on original background, with arrow pointing to region (D). Color-picker appeared to the right of the image.

Now we would like you to report on the color of the pixels you see in the image. We will provide a color-matching tool. Please adjust the hue (color circle) and brightness (slider bar) to match the pixels you see in the image.

*4. What color is the dress at (D)? Page 20

Stimulus: dress on first colored background, with arrow pointing to region (A). Color-picker appeared to the right of the image.

Now we would like you to report on the color of the pixels you see in the image. We will provide a color-matching tool. Please adjust the hue (color circle) and brightness (slider bar) to match the pixels you see in the image.

*1. What color is the dress at (A)?

Page 21

Stimulus: dress on first colored background, with arrow pointing to region (B). Color-picker appeared to the right of the image.

Now we would like you to report on the color of the pixels you see in the image. We will provide a color-matching tool. Please adjust the hue (color circle) and brightness (slider bar) to match the pixels you see in the image.

*2. What color is the dress at (B)?

Page 22

Stimulus: dress on first colored background, with arrow pointing to region (C). Color-picker appeared to the right of the image.

Now we would like you to report on the color of the pixels you see in the image. We will provide a color-matching tool. Please adjust the hue (color circle) and brightness (slider bar) to match the pixels you see in the image.

*3. What color is the dress at (C)?

Page 23

Stimulus: dress on first colored background, with arrow pointing to region (D). Color-picker appeared to the right of the image.

Now we would like you to report on the color of the pixels you see in the image. We will provide a color-matching tool. Please adjust the hue (color circle) and brightness (slider bar) to match the pixels you see in the image.

*4. What color is the dress at (D)? Page 24

Stimulus: dress on second colored background, with arrow pointing to region (A). Color-picker appeared to the right of the image.

Now we would like you to report on the color of the pixels you see in the image. We will provide a color-matching tool. Please adjust the hue (color circle) and brightness (slider bar) to match the pixels you see in the image.

*1. What color is the dress at (A)? Page 25

Stimulus: dress on second colored background, with arrow pointing to region (B). Color-picker appeared to the right of the image.

Now we would like you to report on the color of the pixels you see in the image. We will provide a color-matching tool. Please adjust the hue (color circle) and brightness (slider bar) to match the pixels you see in the image.

*2. What color is the dress at (B)?

Page 26

Stimulus: dress on second colored background, with arrow pointing to region (C). Color-picker appeared to the right of the image.

Now we would like you to report on the color of the pixels you see in the image. We will provide a color-matching tool. Please adjust the hue (color circle) and brightness (slider bar) to match the pixels you see in the image.

*3. What color is the dress at (C)?

Page 27

Stimulus: dress on second colored background, with arrow pointing to region (D). Color-picker *appeared to the right of the image.*

Now we would like you to report on the color of the pixels you see in the image. We will provide a color-matching tool. Please adjust the hue (color circle) and brightness (slider bar) to match the pixels you see in the image.

*4. What color is the dress at (D)?

Page 28

*1. Had you ever seen this image prior to this study?

Pull-down menu with options as follow:

- Yes
- No

If you responded "yes" to questions #1 above, please answer the following questions:

2. When you first saw this image, what color did you see the dress? (Fill in the colors that you saw)

Two textboxes were provided.

3. In your viewings prior to this study, did your perception of the dress colors ever change?

Pull-down menu with options as follow:

- Yes
- No

4. If it did change, please indicate all combinations that you've saw. (Check all that apply)

Check boxes appeared next to the following options:

- blue and black
- blue and olive
- blue and brown
- blue and orange
- white and gold
- white and orange
- lavender and gold
- light blue and gold

5. How often did you see it change?

Pull-down menu with options as follow:

- Never
- Infrequently
- Frequently

Over what time scale did it change?

Pull-down menu with options as follow:

- N/A
- over the course of minutes
- over the course of hours
- over the course of days

Page 29

Stimulus: dress on original background, with alpha-numeric grid superimposed on top

*1. When you first saw the dress (whether in this study or prior), do you recall where you first looked? Give your best estimate by indicating a square in the grid below (e.g. D2).
Two pull-down menus appeared, one with letters, and one with numbers, corresponding to the rows and columns of the grid.

2.When you've look at this image, where have you spent most of your time looking? (again, use the grid to indicate)
Two pull-down menus appeared, one with letters, and one with numbers, corresponding to the rows and columns of the grid.

Page 30
Stimulus: blurred dress on original background (see Methods for stimulus description)

*3. What color is the dress in this blurry image?
Two textboxes were provided.

*4. Do you know what color the dress is in real life? </p>
Two textboxes were provided.

Page 31
Thank you very much for participating in this study.

Please enter any comments you have about this HIT in the box below:
A textbox was provided.

Page 32
Thank you for your help!

SCALE EXPERIMENT:

Page 1
In this study, we will show you a number of images. We want you to look at these images carefully, and then answer some questions about them. If you're not sure about an answer, take your best guess. Please answer all questions before proceeding to the next page.
All questions appear at the bottom of the page. After answering questions, please click the "continue" button at the top of the page to proceed to the next page.

Please try to complete the study in a single sitting, without taking breaks.

Page 2
Do not zoom in on the image. Please give your best estimate at this viewing size.

Stimulus: dress on original background (presented for 15 seconds)

Please look carefully at the dress. A "continue" button will appear just below this text after 15 seconds.

Page 3

Stimulus: dress on original background

This is a picture of a____and____dress. (Fill in the colors that you see

Two textboxes were provided.

Page 4

*Did you have to scroll over the image on the previous page to see all of it?

Pull-down menu with options as follow:

- o Yes, top to bottom
- o Yes, bottom to top
- o No

Page 5

Do not zoom in on the image. Please give your best estimate at this viewing size.

Stimulus: dress on original background, with arrows pointing to regions labeled A, B, C, and D

Now we would like you tell us about the material properties (fabric) of the dress. What kind of material do you think the dress is made of in the region to which the arrow is pointing? Below are some examples of each category. Take your best guess if you're not sure.

A panel showing gray-scale images of fabric swatches appeared at the center of the screen.

*1. How would you describe the material where the arrow is pointing (neckline)?

Pull-down menu with options as follow:

- o semi-transparent, sheer; chiffon, tulle, organza
- o flat matte, low-luster, low-sheen; cotton, linen, jersey, matte-satin
- o semi-gloss, med-luster, med-sheen; rayon, polyester
- o hi-gloss, hi-sheen; silk, glossy-satin

Page 6

Do not zoom in on the image. Please give your best estimate at this viewing size.

Stimulus: dress on original background, with arrows pointing to regions labeled A, B, C, and D

Now we would like you tell us about the material properties (fabric) of the dress. What kind of material do you think the dress is made of in the region to which the arrow is pointing? Below are some examples of each category. Take your best guess if you're not sure.

A panel showing gray-scale images of fabric swatches appeared at the center of the screen.

*2. How would you describe the material where the arrow is pointing (sleeve?

Pull-down menu with options as follow:

- o semi-transparent, sheer; chiffon, tulle, organza
- o flat matte, low-luster, low-sheen; cotton, linen, jersey, matte-satin
- o semi-gloss, med-luster, med-sheen; rayon, polyester
- o hi-gloss, hi-sheen; silk, glossy-satin

Page 7

Do not zoom in on the image. Please give your best estimate at this viewing size.

Stimulus: dress on original background, with arrows pointing to regions labeled A, B, C, and D

Now we would like you tell us about the material properties (fabric) of the dress. What kind of material do you think the dress is made of in the region to which the arrow is pointing? Below are some examples of each category. Take your best guess if you're not sure.

A panel showing gray-scale images of fabric swatches appeared at the center of the screen.

*3. How would you describe the material where the arrow is pointing (body material)?

Pull-down menu with options as follow:

- o semi-transparent, sheer; chiffon, tulle, organza
- o flat matte, low-luster, low-sheen; cotton, linen, jersey, matte-satin
- o semi-gloss, med-luster, med-sheen; rayon, polyester
- o hi-gloss, hi-sheen; silk, glossy-satin

Page 8

Do not zoom in on the image. Please give your best estimate at this viewing size.

Stimulus: dress on original background, with arrows pointing to regions labeled A, B, C, and D

Now we would like you tell us about the material properties (fabric) of the dress. What kind of material do you think the dress is made of in the region to which the arrow is pointing? Below are some examples of each category. Take your best guess if you're not sure.

A panel showing gray-scale images of fabric swatches appeared at the center of the screen.

*4. How would you describe the material where the arrow is pointing (trim)?

Pull-down menu with options as follow:

- o semi-transparent, sheer; chiffon, tulle, organza
- o flat matte, low-luster, low-sheen; cotton, linen, jersey, matte-satin
- o semi-gloss, med-luster, med-sheen; rayon, polyester
- o hi-gloss, hi-sheen; silk, glossy-satin

Page 9 DEMOGRAPHICS

Please help us out by answering the following questions about yourself. You must answer all of the questions marked with an asterisk (*).

*1. What is your gender?

Pull-down menu with options as follow:

- Male
- Female
- Other

*2. How old are you?
Pull-down menu with options 18-100

*3. Is English your mother tongue?
Pull-down menu with options as follow:
- Yes
- No

4. If English is not your mother tongue, what is your mother tongue?
A pull-down menu contained a comprehensive list of languages, including Tamil and other languages spoken in India. List of languages available on request.

*5. What country are you from?
A pull-down menu contained a comprehensive list of countries. List available on request.

5 (A). If you're from the United States, what part of the country are you from?
Pull-down menu with options as follow:
- Northeast - New England
- Northeast - Mid-Atlantic
- Midwest - East North Central
- Midwest - West North Central
- South - South Atlantic
- South - East South Central
- South - West South Central
- West - Mountain

*6. Which levels of education have you completed? Check all that apply.
Check boxes appeared next to the following options:
- Elementary school (primary school)
- High school
- College/university
- Advanced graduate work

*7. Do you have normal or corrected-to-normal vision?
Pull-down menu with options as follow:
- Yes
- No

*8. Are you colorblind?
Pull-down menu with options as follow:
- Yes
- No
- Unsure

*9. On most days, what time do you typically wake up?

Two pull-down menus appeared: one containing numbers 1-12, and one containing "A.M. (morning)" and "P.M. (night)"

*10. On most days, what time do you typically go to bed?
Two pull-down menus appeared: one containing numbers 1-12, and one containing "A.M. (morning)" and "P.M. (night)"
11. Do you have a background in the visual arts? Check all that apply:
Check boxes appeared next to the following options:
- Professionally
- Hobby
- Painter
- Graphic designer
- Art historian
- Architecture
- Illustrator
- Photographer

Page 10
Tell us a bit about your environment.

*1. What device are you using to view this webpage?
Pull-down menu with options as follow:
- Laptop
- Computer monitor (flat screen
- Computer monitor (rounded screen
- Tablet
- Other

*2. How big is the screen on which you're viewing this webpage? (Take your best guess if you're not sure.)
Pull-down menu with options as follow:
- 10-13 inches wide
- 14-17 inches wide (regular laptop)
- 14-17 inches wide
- More than 17 inches wide

*3. Are you using a Mac or PC?
Pull-down menu with options as follow:
- Mac
- PC
- Unsure

*4. What time is it right now? (select the approximate hour)
Two pull-down menus appeared: one containing numbers 1-12, and one containing "A.M. (morning)" and "P.M. (night)"

*5. Are you indoors or outdoors?
Pull-down menu with options as follow:
- Indoors

- o Outdoors

*6. If you're indoors, is there natural light entering the room through a window?
Pull-down menu with options as follow:

- o N/A
- o Yes
- o No

*7. What are the lighting conditions where you are right now, on a scale of 1 (darkest) to 10 (brightest)?
Pull-down menu with options as follow:

- o 1 (darkest; as dim as a room without lights at night
- o 2
- o 3
- o 4
- o 5
- o 6
- o 7
- o 8
- o 9
- o 10 (brightest; as bright as the beach on a sunny day

*8. Is the software f.lux installed on the computer you're using to do this HIT?
Pull-down menu with options as follow:

- o Yes
- o No

Page 11
Stimulus: dress on original background, with arrow pointing to region (A). Color-picker appeared to the right of the image.

Do not zoom in on the image. Please give your best estimate at this viewing size.

Now we would like you to report on the color of the pixels you see in the image. We will provide a color-matching tool. Please adjust the hue (color circle) and brightness (slider bar) to match the pixels you see in the image.
*1. What color is the dress where the arrow is pointing (neckline)?

Page 12
Stimulus: dress on original background, with arrow pointing to region (B). Color-picker appeared to the right of the image.
Do not zoom in on the image. Please give your best estimate at this viewing size.

Please adjust the hue (color circle) and brightness (slider bar) to match the pixels you see in the image.
*2. What color is the dress where the arrow is pointing (sleeve)?

Page 13

Stimulus: dress on original background, with arrow pointing to region (C). Color-picker appeared to the right of the image.

Do not zoom in on the image. Please give your best estimate at this viewing size.

Please adjust the hue (color circle) and brightness (slider bar) to match the pixels you see in the image.
*3. What color is the dress where the arrow is pointing (body material)?

Page 14

Stimulus: dress on original background, with arrow pointing to region (D). Color-picker appeared to the right of the image.

Do not zoom in on the image. Please give your best estimate at this viewing size.

Please adjust the hue (color circle) and brightness (slider bar) to match the pixels you see in the image.
*4. What color is the dress where the arrow is pointing (trim)? P

age 15

Stimulus: dress on original background

*Is the dress in shadow?
Pull-down menu with options as follow:
- ○ Yes
- ○ No

Page 16

Stimulus: dress on original background

*1. Had you ever seen this image prior to this study?
Pull-down menu with options as follow:
- ○ Yes
- ○ No

If you responded "yes" to questions #1 above, please answer the following questions:

2. When you first saw this image, what color did you see the dress? (Fill in the colors that you saw)

Two textboxes were provided.

3. In your viewings prior to this study, did your perception of the dress colors ever change?
Pull-down menu with options as follow:
- ○ Yes
- ○ No

4. If it did change, please indicate all combinations that you've saw. (Check all that apply)
Check boxes appeared next to the following options:

- blue and black
- blue and olive
- blue and brown
- blue and orange
- white and gold
- white and orange
- lavender and gold
- light blue and gold

5. How often did you see it change?
Pull-down menu with options as follow:

- Never
- Infrequently
- Frequently

Over what time scale did it change?
Pull-down menu with options as follow:

- N/A
- over the course of seconds
- over the course of minutes
- over the course of hours
- over the course of days

Page 17
Stimulus: dress on original background

Do not zoom in on the image. Please make your best estimate at this viewing size.

*If you were forced to choose between one of these options, what color is the dress in this photograph?
Pull-down menu with options as follow:

- blue and black
- white and gold

Page 18
*1. Do you know what color the dress is in real life?
Pull-down menu with options as follow:

- Yes
- No

If you answered "yes" to question #1, please answer the following question

2. What is the true color of the dress?
Two textboxes were provided.

Page 19

Thank you very much for participating in this study
Please enter any comments you have about this HIT in the box below:
A textbox was provided.

Page 20
Thank you for your help!

3.1.6. Supplemental Discussion

Many hypotheses have been proposed regarding the source of the individual differences. Our study shows that age and gender are the two most significant predictors. But these correlations do not reveal the causes of the individual differences. A plausible hypothesis is that different people have different priors about the spectral content of the illuminant, with the population of priors varying along the daylight locus [S4]. Consistent with this hypothesis, observers show greater variability along the daylight locus when adjusting a patch on a screen to be "white" (our unpublished results; see also [4 S5], but see [6]). We have not found a strong correlation between the chromaticities subjects report to be "white" and the likelihood that they see the dress as W/G or B/K (data not shown). Nonetheless, it remains possible that people who experience more daylight are more likely to develop a "blue" illuminant prior, since the sky is blue; whereas, people who spend more of their time awake at night are more likely to develop a "warm" illuminant prior, since (historically) most artificial lights are biased for long wavelengths (i.e. "warm"). Other subjects may favor a neutral illuminant (and see the dress as B/B). If the illumination-discounting theory is correct, and different people have different priors on the illumination determined by their exposure, we would predict that "larks" (daytime chrontotype) will show a bias for W/G (discounting the blue in the dress); and "night owls" (night-time chronotype) will show a bias for B/K (discounting the orange in the dress). Does the extent to which one is exposed to a given illuminant (blue-sky daylight versus warm incandescent light) predict the colors one sees the dress? We attempted to answer this question by asking subjects about their sleep-wake patterns. We found a slight, but insignificant trend: subjects reporting more awake time during the day were slightly more likely to see the dress as W/G. Although the results provided by directly surveying sleep-wake patterns are at present inconclusive (we may need to sort data by longitude etc.), our results provide some support for the chronotype hypothesis. Older subjects and women are more likely to have a daytime chronotope [6].

Consistent with this, we found that these two variables significantly predicted a higher likelihood of reporting the dress to be W/G. Lens density, corneal coloration, and other optical properties of the eye vary with age. It is also a possibility that these factors contribute to the individual differences observed presently.

Cultural influences on perception of 'the dress'?

In the group of all male and female subjects ($N = 1399$), age ($p < 0.0001$ Wald chi-square) and a subject's country (India versus USA) ($p = 0.0016$) were significant in predicting a main category response versus a non-main category response (blue/brown or other): older subjects and people from India were less likely to report a main category (see **Table S1** for full model results). These results may provide some evidence for cultural influences in how the dress image is perceived. Alternatively, the Indian population contained many more naïve subjects relative to the USA population. When we removed country from the logistic regression model, prior experience of the image significantly predicted response ($p = 0.021$, Wald chi-square): the odds of reporting a main category increased by a factor 1.53 [1.06-2.17] for subjects who had seen the image previously.

Individual differences in multistability

Non-naïve subjects who saw the dress in the main experiment as W/G rather than B/K were less likely to have reported a switch in the color of the dress prior to our study. Given the results of our size experiment, we can conclude that the size of the image used in the main experiment biased subjects towards reporting B/K. We hypothesize that the subjects who continue to see the dress as W/G at the smaller sizes used in the main experiment were less likely to experience 'the dress' as multistable—these subjects, on average, seemed to be less susceptible to switching. This hypothesis is supported by the data from the experiments using the images embedded in scenes with simulated illuminants. Fewer W/G respondents (64%) switched their perception to B/K when the dress was embedded in a warm-illuminated scene, compared to the fraction of B/K respondents (80%) who switched their perception when the dress was embedded in a cooilluminated scene (Figure S2). These results provide evidence of striking individual differences not only in the perceived color, but also in the multistability of the image.

In the Main experiment, subjects performed the color matching for the various regions of the dress (i, ii, iii, iv, **Figure 1A**), after they had been queried about the terms used for their perception of the dress under the different contexts (images 2, 3). We found that the particular image (2 or 3) that was shown before the color matching was conducted biased the color-matching results (data not shown), providing evidence of hysteresis. We are investigating the extent to which prior exposure, over various time scales, impacts perception of the dress (experiments in progress).

Effects of image size

In the Scale experiment, we found that a reduction in image size biased the population who reported a main category (B/K or W/G) to see the dress as B/K (with a unit increase in image size, the odds of reporting white/gold increased by a factor 1.004 [1.002 – 1.007]; $p < 0.0001$). As described in the main text, we hypothesize that this difference may be attributable to a change in spatial frequency that renders less material information available. Alternatively the impact of size on perceived color may reflect differences in how the image is perceived when more or less of it is seen at a given time. In addition, it is possible that when only a portion of the image is seen at a given time, the particular portion of the image that is seen first has an impact. In the 100% and 150% conditions of the Scale experiment, only part of the image was visible at once: the top of the image was the first part of it seen. Subjects then had to scroll over the image, from top to bottom, to view it in full; in the 10% and 36%, the full image was visible on the screen at once.

The pixel colors of the regions at the top of the dress are lighter than those at the bottom of the dress (**Figure 1A,B**). A subject integrating over the full image at once might arrive at a different percept than a subject who sees the lighter portion first, and then scrolls down over the rest of the image. If true, it is possible that a subject's first fixation on the dress image might influence their report of its color (experiments in progress).

3.1.7. Supplemental References

S1. Agresti, A., & Coull, B. A. (1998). Approximate is better than "exact" for interval estimation of binomial proportions. *The American Statistician*, *52*(2), 119-126.

S2. Westland, S., and Ripamonti, C. (2004). Computational colour science using matlab. John Wiley and Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, England.

S3. Wilmer, J.B., Germine, L., Chabris, C.F., Chatterjee, G., Gerbasi, M., and Nakayama, K. (2012). Capturing specific abilities as a window into human individuality: the example of face recognition. Cogn Neuropsychol *29*, 360-392.

S4. Conway, B.R. (2015). Why do we care about the colour of the dress? The Guardian *http://www.theguardian.com/commentisfree/2015/feb/27/colour-dress-optical-illusion-social-media*.

S5. Beer, R.D., Dinca, A., and MacLeod, D.I.A. (2006). Ideal white can be yellowish and bluish, but not reddish and greenish. Journal of Vision *6*, 417.

S6. Adan, A., Archer, S.N., Hidalgo, M.P., Di Milia, L., Natale, V., and Randler, C. (2012). Circadian typology: a comprehensive review. Chronobiol Int *29*, 1153-1175.

Chapter 3.2: #theDress: Categorical perception of an ambiguous color image

The contents of the chapter have been published:
Lafer-Sousa, R. and Conway, B.R. #theDress: Categorical perception of an ambiguous color image. *JOV* **17**, 25. (2017)

3.2.1. Abstract

We present a full analysis of data from our preliminary report (Lafer-Sousa et al, 2015), and test whether #theDress image is multi-stable. The key attribute of a "multi-stable" image is that it must give rise to more than one mutually exclusive percept. Clustering algorithms of color-matching data showed that the dress was seen categorically, as white/gold (W/G) or blue/black (B/K), with a blue/brown transition state. Multinomial regression predicted categorical labels. Consistent with our prior hypothesis, W/G observers inferred a cool illuminant, whereas B/K observers inferred a warm illuminant; moreover, subjects could use skin color alone to infer the illuminant. The data provide some, albeit weak, support for our hypothesis that day-larks see the dress as W/G and night-owls see it as B/K. About half of observers who were previously familiar with the image reported switching categories at least once. Switching probability increased with professional art experience. Priming with an image that disambiguated the dress as B/K biased reports towards B/K (priming with W/G had negligible impact); furthermore, knowledge of the dress' true colors and any prior exposure to the image shifted the population towards B/K. These results show that some people have switched their perception of the dress. Finally, consistent with a role of attention and local image statistics in determining how multi-stable images are seen, we found that subjects tended to discount as achromatic the dress component that they did not attend to: B/K reporters focused on a blue region; W/G reporters focused on a golden region.

3.2.2. Introduction

Most visual stimuli are under-determined: a given pattern of light can be evidence for many different surfaces or objects. Despite being under-determined, most retinal images are resolved unequivocally. It is not known how the brain resolves such ambiguity, yet this process is fundamental to normal brain function. Multi-stable images are useful tools for investigating the underlying neural mechanisms. The two defining properties of multi-stable stimuli are that they give rise to more than one plausible, stable, percept; and that the alternative percepts are mutually exclusive (Leopold & Logothetis, 1999; Long & Toppino, 2004; Schwartz, Grimault, Hupe, Moore, & Pressnitzer, 2012; Scocchia, Valsecchi, & Triesch, 2014). Multi-stable images are similar to binocular rivalrous stimuli, although in binocular rivalry the competition is between two different images rather than alternative interpretations of a single image. Although the first account of binocular rivalry involved color (Dutour, 1760), to date there are no striking examples of multi-stable color images. Of course, not all colored stimuli are unambiguous. Consider turquoise, which might be called "blue" or "green" by different people. Such ambiguous color stimuli typically retain their ambiguity even when labeled categorically, unlike multi-stable shape images (Klink, van Wezel, & van Ee, 2012). To date, the best example of something approximating a multi-stable color phenomenon is the colored Mach card, in which the color of a bi-colored card folded along the color interface and viewed monocularly can vary depending on whether one perceives the card receding or protruding (Bloj, Kersten, & Hurlbert, 1999). But it is not clear that the color perceptions of the Mach card are categorical. Moreover, the phenomenon is primarily an illusion of 3-D geometry: without stereopsis, the perspective cues are ambiguous; the way the colors are perceived is contingent on how these cues are resolved.

Could #thedress be an elusive multi-stable color image? Initial reports on social media raised the possibility that the image was seen in one of two mutually exclusive ways, as "white and gold" (W/G) or "blue and black" (B/K). But color-matching data (not color names) reported by Gegenfurtner et al (Gegenfurtner, Bloj, & Toscani, 2015) concluded that there were many different ways in which the dress colors could be seen . The tentative conclusion was that reports of two categories arose as an artefact of the two-alternative-forced-choice question posed by social media ("do you see the dress as W/G or B/K?"). The implication was that the true population distribution is unimodal, which is inconsistent with the idea that the image is multi-stable. The Gegenfurtner et al study measured perception of 15 people. It is not known how many subjects would be required to reject the hypothesis that the population distribution is unimodal. We addressed these issues through a full, quantitative analysis of the results we presented in preliminary form shortly after the image was discovered, in which we argued that the dress was seen categorically (Lafer-Sousa, Hermann, & Conway, 2015). A side goal was to evaluate the extent to which tests conducted online replicate results obtained under laboratory conditions. Many studies of perception and cognition are being conducted through online surveys; it remains unclear whether results obtained in lab and online are comparable.

Popular accounts suggest that people are fixed by "one-shot learning" in the way they see the dress image (Drissi Daoudi, Doerig, Parkosadze, Kunchulia, & Herzog, 2017). These observations suggest that the dress is not like a typical multi-stable image, because it is widely thought that most people experience frequent perceptual reversals of multi-stable images—although frequent reversals is not a necessary property of multi-stability (see *Discussion*). Perception of multi-stable images was initially thought to depend only on low-level factors, such as where in the image one looked (Long & Toppino, 2004). It is now recognized that high-level factors, including familiarity with the image, prior knowledge, personality, mood, attention, decision making, and learning, also play a role (Leopold & Logothetis, 1999; Kosegarten & Kose, 2014; Podvigina & Chernigovskaya, 2015). Data in our initial report suggested that some observers experience perceptual reversals of the dress, raising the possibility that the image is not unlike other multi-stable images. Here we determined the extent to which the individual differences in perception of the dress image are fixed. We characterized the conditions that promote perceptual reversals of the dress, and tested five factors known to influence how multi-stable images are perceived: (1) prior knowledge about the image (Rock & Mitchener, 1992); (2)

exposure to disambiguated versions (Fisher, 1967; Long & Toppino, 2004); (3) low-level stimulus properties (e.g. stimulus size) (Chastain & Burnham, 1975) (4) where subjects look (or attend) (Ellis & Stark, 1978; Kawabata & Mori, 1992; Kawabata, Yamagami, & Noaki, 1978); and (5) priors encoded in genes or through lifetime experience (Scocchia et al., 2014). These experiments were afforded because we tested people who varied in terms of both prior exposure to the image and knowledge about the color of the dress in the real world.

By examining the factors that influence perception of the dress image, we hoped to shed light on how the brain resolves under-determined chromatic signals. While low-level sensory mechanisms like adaptation in the retina can account for color constancy under simple viewing conditions (Chichilnisky & Wandell, 1995; D'Zmura & Lennie, 1986; Foster & Nascimento, 1994; Land, 1986; Stiles, 1959; von Kries, 1878; Webster & Mollon, 1995), they fail to explain constancy of natural surfaces (Brainard & Wandell, 1986; Webster & Mollon, 1997) and real scenes (Hedrich, Bloj, & Ruppertsberg, 2009; Khang & Zaidi, 2002; Kraft & Brainard, 1999). We have argued that the competing percepts of the dress are the result of ambiguous lighting information. The colors of the pixels viewed in isolation align with the colors associated with daylight (Brainard & Hurlbert, 2015; Conway, 2015; Lafer-Sousa et al., 2015). The visual system must contend with two plausible interpretations—that the dress is either in cool shadow, or in warm light. In our prior report, we tested the idea that illumination assumptions underlie the individual differences in color perception of the dress, by digitally embedding the dress in scenes containing unambiguous illumination cues to either warm or cool illumination. Most observers conformed to a single categorical percept consistent with the illumination cued (Lafer-Sousa et al., 2015). Here we directly tested the hypothesis by analyzing subjects' judgments about the light shining on the dress. We also tested our hypothesis that the way the dress is seen can be explained by one's chronotype: "night-owls" spend much of their awake time under incandescent light, and we hypothesized they would therefore be more likely to discount the orange component of the dress, and see the dress' colors as B/K; "day larks" spend more of their time under blue daylight, and we surmised that they would discount the blue component of the dress and see the dress colors as W/G (Rogers, 2015).

Finally, we used the dress image as a tool to examine the role of 'memory colors' in color constancy. The spectral bias of the illuminant could, in theory, be determined by comparing the chromatic signals entering the visual system with the object colors stored in memory. The gamut of human skin occupies a distinctive profile in cone-contrast space that is surprisingly stable across skin types, and shifts predictably under varying illuminations (Crichton, Pichat, Mackiewicz, Tian, & Hurlbert, 2012). These statistics, coupled with the fact that skin is viewable in almost every natural glance, make skin a potentially good cue to estimate the spectral bias of the illuminant (Bianco & Schettini, 2012). We used our disambiguation paradigm (Lafer-Sousa et al. 2015), digitally embedding the dress in scenes in which we systematically introduced different cues to the illuminant. Our results provide the first behavioral evidence that skin color is sufficient to recover information about the illuminant for color constancy.

3.2.3 Materials and methods

Experimental setup

Detailed methods are provided in the supplementary material of our prior report (Lafer-Sousa et al, 2015). Raw materials and sample analyses are provided here: https://github.com/rlaferso/-TheDress. The majority of participants (N=2200) were recruited and tested online through Amazon's Mechanical Turk using a combination of template (Morris Alper's Turk Suite Template Generator 2014, available online at http://mturk.mit.edu/template.php) and custom HTML and JavaScript code. A smaller number of subjects (N=53) were recruited from the MIT University and Wellesley College campus through word of mouth and social media, and tested using the M-Turk platform on a calibrated display in the laboratory. We adhered to the MIT Committee on the Use of Humans as Experimental Subjects policies on using Amazon's Mechanical Turk for research purposes. Informed consent was obtained for those subjects who performing the study in the laboratory study. Procedures were approved by the institutional review board of Wellesley College. Subjects were between 18 and 69 years of age. To control for subject quality among the Mechanical Turk participants, we required that subjects have Mechanical Turk approval ratings of 95% or higher, and have completed at least 1000 human intelligence tasks (HITs) on Mechanical Turk previously.

In -laboratory subjects viewed the display at 40 cm. Subjects used a chin rest to control viewing angle and distance. The experiment was performed on a calibrated 21.5-inch iMac computer with 1920 x 1080 pixel resolution in a windowless room with LED overhead lighting (CIExyY: 0.4814 0.4290, 4.3587 cd/m^2), measured off the MacBeth color checker standard white, held at the same location and viewing angle on the monitor at which we presented the dress image). Normal color vision was confirmed with Ishihara plates (Ishihara, 1977).

To ensure that stimuli were the same size across displays for on-line subjects, we specified the sizes of stimulus images in absolute pixels in the HTML experiment code. There is some variability from display to display in terms of the actual physical size of a pixel. We measured the images on a typical monitor in the laboratory, to provide a reasonable estimate for how the pixel values correspond to degrees of visual angle. We estimate that among different displays the variance in actual display size was ~+/-10% of the size measured in lab.

Three experiments were conducted: 1, a main experiment; 2, a follow-up experiment to assess the role of image size in determining what colors people report; 3, an in-laboratory, controlled experiment. Data were pooled from the various experiments depending on the analysis performed.

In each experiment dress color-percepts were queried using two tasks: a free-response color-naming task and a color-matching task. In Experiments 1 and 3 subjects were also asked to report on their impressions of the lighting conditions in the image by providing temperature ratings and verbal descriptors, and to estimate where in the image they thought they spent most of their time looking. As well, Experiments 1 and 3 queried color percepts for a set of digitally synthesized test stimuli featuring the dress (cropped from the original image) embedded in scenes with unambiguous simulated lighting conditions, and a version of the original image that had been spatially blurred. Experiment 3 included an additional set of synthetic test images not presented in Experiment 1. Each experiment contained questions about subjects' demographics, viewing conditions, and past-viewing experiences, which were distributed throughout the experiment (the full questionnaire is reproduced in Lafer-Sousa et al. 2015, SI).

Stimuli

Stimuli reproductions are provided in the Supplemental Image Appendix; can also be viewed online: https://youtu.be/U6c4au-Wu-E.

'Original Image' (used in Experiments 1, 2, and 3): The original dress image that circulated on the internet (courtesy of Cecilia Bleasdale; Figure 1). In Experiments 1 and 3, the original dress photograph was presented at 36% of its original size so that the entire image would be visible on the display. The image was presented at an absolute size of 226 x 337 pixels. This corresponded to 7.2° of visual angle width. In Experiment 2, the original dress photograph was presented at one of four sizes, defined as a % of the original: 10% (63 x 94 pixels, or 2.0° of visual angle on a 21.5-inch iMac), 36% (the image size presented in Experiment 1; 226 x 337 pixels, or 7.2° of visual angle), 100% (628 x 937 pixels, or 20.0° of visual angle; and 150% (942 x 1405 pixels, or 30.0° of visual angle). The stimuli in the 10% and 36% conditions fit fully in all browser windows. For the 100% and 150% images, only part of the image was visible at once in the height direction, but was completely visible in the horizontal dimension. As a result, subjects had to scroll over the stimulus image, from top to bottom, to view it. Scrolling to see the entire image was required in order for subjects to access the buttons to move through the study, ensuring that all subjects saw the entire image even in the 100% and 150% displays. Subjects were randomly assigned to one of the four scale conditions.

'Blurry dress' Stimulus: A blurry version of the original image was presented at 41% of the source size, with a Gaussian blur radius of 3.3 pixels (0.11°). The image was 8.3° of visual angle along the horizontal axis.

Disambiguating Stimuli ('warm' and 'cool' illumination simulations):

- 'Cue-rich' (used in Experiments 1 and 3). To create the cue-rich test stimuli, we digitally dressed a female model in the garment and embedded her in a scene depicting a Rubik's cube under either a simulated warm ('yellowish') or cool ('blueish') illuminant (cube reproduced with permission from Beau Lotto; Lotto & Purves, 2002) (see Supplementary Image Appendix for full size reproductions). In the cool-illumination scene, the woman was positioned in the shadow cast by the cube. The color of the woman's skin and hair was tinted to reflect the color bias of the simulated illuminant using a semi-transparent color overlay. The chromaticity of the overlay was defined on the basis of the chromaticity of the white component of the scene's checkered floor, which provides a quantitative white point for the scene (for the cool-illumination scene we used the white checkers that were cast in shadow, corresponding to our placement of the model in shadow). In the warm scene, the white point was 0.352, 0.394, 66 cd/m^2 (CIE$_{xyY}$ 1931); in the cool scene the white point was 0.249, 0.271, 23 cd/m^2. Note that the pixels making up the dress were never manipulated. The dress portion of the stimulus was presented at 76% of the size of the dress in the original image; the complete picture was 518x429 pixels (~16.5° of visual angle on the horizontal axis).

- 'Uniform Surround' Test Stimuli (used in Experiment 3). To test whether a low-level sensory mechanism like receptor adaptation or local color contrast is sufficient to resolve the dress' colors, we presented we superimposed the isolated dress on uniform fields matched to the mean chromaticity of the cue-rich scenes (CIE$_{xyY:}$ warm field, 0.363, 0.414, 51 cd/m^2; cool field, 0.276, 0.293, 29 cd/m^2). The dress portion of the stimulus was presented at 76% of the size of the dress in the original image; the complete picture was 518x429 pixels (~16.5° of visual angle on the horizontal axis). The pixels that make up the dress were not manipulated.

- 'Skin Only' Test Stimuli (Experiment 3). The rich 3-D simulations we used to convey unambiguous illumination conditions consisted of several cues to the illumination: the average color in the surround, 3-D structure, shadows, and skin coloration consistent with a bias brought about by either warm or cool illumination. To create these cue-rich stimuli, we digitally dressed a female model in the garment and embedded her in a scene depicting a Rubik's cube under either a simulated warm ('yellowish') or cool ('blueish') illuminant (cube reproduced with permission from Beau Lotto; Lotto & Purves, 2002) (**Figure 18** insets). In the cool-illumination scene, the woman was positioned in the shadow cast by the cube. The color of the woman's skin and hair was tinted to reflect the color bias of the simulated illuminant using a semi-transparent color overlay. The chromaticity of the overlay was defined on the basis of the chromaticity of the white component of the scene's checkered floor, which provides a quantitative white point for the scene (for the cool-illumination scene we used the white checkers that were cast in shadow, corresponding to our placement of the model in shadow). In the warm scene, the white point was 0.352, 0.394, 66 cd/m^2 (CIE$_{xyY}$ 1931); in the cool scene the white point was 0.249, 0.271, 23 cd/m^2. Note that the pixels making up the dress were never manipulated. To test whether skin chromaticity is by itself a sufficient cue to achieve good color constancy, we presented the dress superimposed on the woman on a white (achromatic) background (CIExyY: 0.322, 0.352, 75 cd/m^2), and tinted her skin according to the spectral bias of the illuminants simulated in the cue-rich scenes. The dress portion of the stimulus was presented at 76% of the size of the dress in the original image; the complete picture was 518x429 pixels (~16.5° of visual angle on the horizontal axis). The pixels that make up the dress were not manipulated. We considered color constancy 'good' if the majority of subjects conformed to the percept predicted by the lighting cues in each condition, and 'bad' when individual subjects perceptions were unaffected by the changes in simulated lighting conditions: under the cool-illuminant, good color constancy predicts that subjects should discount a cool ('blue') component, and see the dress as WG, while under the warm illuminant, subjects should discount a warm ('yellow') component, and see the dress as BK. Note that we use the terms blue, yellow, white, warm and cool as short-hand for accurate colorimetric descriptions.

McNemar's chi-square tests were used to compare goodness of constancy achieved in different stimulus conditions. McNemar's test is a within-subjects z-test of equality of proportions for repeated measures. Each test compares the proportion of subjects that did or did not conform to the percept cued in stimulus condition X vs. the proportion of subjects that did or did not conform to the percept cued in stimulus condition Y. Six tests were performed: cue rich scene (warm) vs uniform background (warm); skin tint (warm) vs uniform background (warm); skin tint (warm) vs cue rich scene (warm); cue rich scene (cool) vs uniform background (cool); skin tint (cool) vs uniform background (cool); skin tint (cool) vs cue rich scene (cool). Results of these tests are reported in **Table 1**.

Tasks

Color-naming

Each image was shown for fifteen seconds and then subjects were prompted to report the apparent color of the dress via a free response verbal task (two textboxes were provided): *"Please look carefully at the dress. A "continue" button will appear just below this text after 15 seconds. This is a picture of a ____ and ____ dress. (Fill in the colors that you see)."* The image was on the screen continuously while the subjects responded (this was not designed as a test of color memory). Color descriptions of the dress were binned into categories: Blue/Black, White/Gold, Blue/Brown, Other (following the methods outlined in Lafer-Sousa et al., 2015).

Lighting Judgments

After performing the color-naming task, subjects were prompted to rate the apparent quality of the light illuminating the background of the image (*"On a scale from 1 to 5, where 1 is cool and 5 is warm, please rank the lighting conditions **in the background**"*) and the light illuminating the dress (*"On a scale from 1 to 5, where 1 is cool and 5 is warm, please rank the light **illuminating the dress**"*). Subjects were then asked to characterize the light in the background, and the light illuminating the dress, by checking off any of a number of possible verbal descriptors from a list (*"The **lighting in the background** is...Check all that apply/ The light **illuminating the dress** is...Check all that apply":* 'dim', 'dark', 'cool', 'blueish', 'bright', 'warm', 'yellowish', 'glaring', 'blown out', 'washed out', 'reddish', 'greenish', 'purplish', 'iridescent'*).*

Color-matching

Each image was presented a second time (again, for 15 seconds) and this time subjects were prompted to make color matches to four regions of the dress (Figure 1 inset, arrows i,ii,iii,iv), using a color-picker tool comprising a complete color gamut: "*Please adjust the hue (color circle) and brightness (slider bar) to match the pixels you see in the image.*"

For Experiments 1 and 3, in the first half of the experiment, each image was shown for fifteen seconds and then subjects were prompted to perform the color-naming task and the lighting judgment task. In Experiment 2, subjects were not asked to perform the lighting judgment task. The image was on the screen continuously while the subjects responded (this was not designed as a test of color memory). In the second half of the experiment, each image was shown a second time, and after 15 seconds, subjects were prompted to perform the color-matching task (the image remained on the screen continuously while the subjects performed the task). Between the presentation of the first and second images, we collected basic demographics information. Between the presentations of the last two images, we asked subjects about the environment in which they were completing the study. We also asked subjects whether they had viewed the original dress image prior to this study, and if so, whether they had experienced multiple percepts of the dress (i.e. "switching").

The first image shown was always the original dress photograph (courtesy of Cecilia Bleasdale.), but the order of the subsequent two images differed between the two conditions. In addition, all subjects were queried on their perception of a blurry version of the original image.

Experiment 1:

Order A: *Report Colors and Lighting for* 'original dress image', 'cue-rich scene, cool', 'cue-rich scene, warm'; *Make Color matches for* 'original dress image', 'cue-rich scene, cool', 'cue-rich scene, warm'; *Report Colors of the* 'blurry dress']. **Order B:** *Report Colors and Lighting for* 'original dress image', 'cue-rich scene, warm', 'cue-rich scene, cool'; *Make Color matches for* 'original dress image', 'cue-rich scene, warm', 'cue-rich scene, cool'; *Report Colors of the* 'blurry dress'].

Experiment 2:

Order: Report colors for the original dress image; make color matches for original dress image.

Experiment 3:

- **Order A:** *Report Colors and Lighting for* 'original dress image', 'cue-rich scene, cool', 'cue-rich scene, warm', 'uniform surround, cool', 'uniform surround, warm', 'skin-tint only, cool', 'skin-tint only, warm; *Make Color matches for* 'original dress image', 'cue-rich scene, cool', 'cue-rich scene, warm', 'uniform surround, cool', 'uniform surround, warm', 'skin-tint only, cool', 'skin-tint only, warm; *Report Colors of the* 'blurry dress'].
- **Order B:** *Report Colors and Lighting for* 'original dress image', 'cue-rich scene, warm', 'cue-rich scene, cool', 'uniform surround, warm, 'uniform surround, cool, 'skin-tint only,

warm, 'skin-tint only, cool; *Make Color matches for* 'original dress image', 'cue-rich scene, warm', 'cue-rich scene, cool', 'uniform surround, warm, 'uniform surround, cool, 'skin-tint only, warm, 'skin-tint only, cool; *Report Colors of the* 'blurry dress'].

Data Analysis

Analyses are described in the legends and Results, if not here. All statistical analyses were conducted using Matlab. A value of $p < 0.05$ was considered statistically significant.

3.2.4. Results

#TheDress (**Figure 1, left**) is a rare image that elicits striking individual differences in color perception (Gegenfurtner et al., 2015; Lafer-Sousa et al., 2015). Although the pixels that make up the dress are (in isolation) light blue and brown, most observers queried through social media reported seeing the dress as either Blue/Black (B/K) or White/Gold (W/G) (Rogers, 2015). A minority of subjects (~10%) reported seeing the dress as Blue/Brown (B/B).

Categorical perception of the dress: true or false?

Subjects were asked to identify the colors of four regions of the dress (i-iv, **Figure 1**, left). The three-dimensional (RGB) coordinates of the color-matching data were compressed to one dimension using principal component analysis (Lafer-Sousa et al, 2015). Subjects' color matches for the brown region of the dress (i, iv) are plotted against their matches for the blue region of the dress (ii, iii), and are highly correlated. Moreover, the correlation shows two peaks, suggestive of two underlying categories. This pattern of results was consistent for both the naïve subjects and those who had seen the image previously (**Figure 1B,C**). The peaks in the population density plots corresponded well with the categorical color descriptions provided by the participants: Figure 1D shows the color matches made by subjects who reported a "blue/black" (B/K) percept (left panel) or a "white/gold" (W/G) percept (middle panel). Some subjects reported "blue/brown". These subjects showed color matches that were intermediate to the two main categories (Figure 1D, right panel). To quantitatively test the hypothesis that the dress is viewed categorically, we performed a *k*-means clustering assessment on the color-matching data.

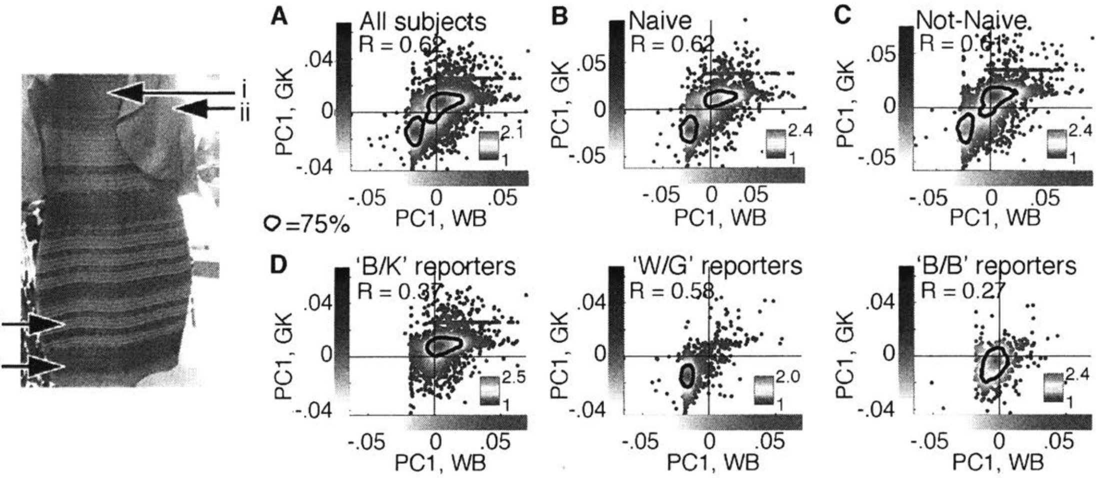


Figure 1. Population distributions of subjects' color-matches show categorical perception of the dress. Subjects used a digital color picker to match their perception of four regions of the dress (i, ii, iii, iv); the dress image was shown throughout the color-matching procedure. A, Matches for regions (i) and (iv) of the dress plotted against matches for regions (ii) and (iii), for all online subjects ($N = 2200$; $R = 0.62$, $p < 0.001$). Contours contain the highest density (75%) of matches. The first principal component of the population matches (computed from CIELUV values) to (i,iv) defined the y axis (gold/black, 'GK'); the first PC of the population matches to (ii,iii) defined the x axis (white/blue, 'WB'). Each subject's (x,y) values are the PC weights for their matches; each has two (x,y) pairs, corresponding to (i,ii) and (iii,iv). Color scale is number of matches (smoothed). B, Color matches for regions (i,iii) of the dress plotted against matches for regions (ii,iv) for subjects who had never seen the dress before the experiment ('Naïve') ($N= 1017$; $R = 0.62$, $p < 0.001$). Axes and contours were defined using data from only the Naïve subjects. C, Color matches for regions (i,iv) of the dress plotted against matches for regions (ii,iii) for subjects who had seen the dress before the experiment ($N= 1183$; $R = 0.61$, $p < 0.001$). Axes and contours were defined using data from only the subset of subjects who were Not-Naïve. D, Color matches for all subjects (from A) were sorted by subjects' verbal color descriptions ("blue/black" = B/K, $N = 1184$; "white/gold"=W/G, $N = 686$; "blue/brown"=B/B, $N = 272$) and plotted separately. Axes defined as in A. In all panels, contours contain the highest density (75%) of the matches shown in each plot. Dress image reproduced with permission from Cecilia Bleasdale.

There are several methods for estimating the optimal number of *k* clusters (groups) in a distribution, but most are constrained to assessing solutions of two or more clusters. To test our hypothesis, we need some way of assessing the relative goodness of clustering for a single-component ($k = 1$ cluster) versus a $k > 1$ component model. To do so, we clustered the data, varying the number of clusters ($k = 1, 2, \ldots 13$), then used the Gap method to assess the outcomes and identify the optimal *K* clusters (Tibshirani, Walther, & Hastie, 2001). The Gap statistic is estimated by comparing the within-cluster dispersion of a *k*-component model to its expectation under a single-component null, and seeks to identify the smallest *k* satisfying: $\mathrm{Gap}(k) \geq \mathrm{Gap}(k+1) - \mathrm{SE}(\mathrm{Gap}(k+1))$. For color matches made to both upper (**Figure 2A-C**) and lower (**Figure 2D-F**) regions of the dress, the single-component solution was rejected in favor of 2 or 3 clusters, confirming the suspected non-unimodality of the underlying population distribution.

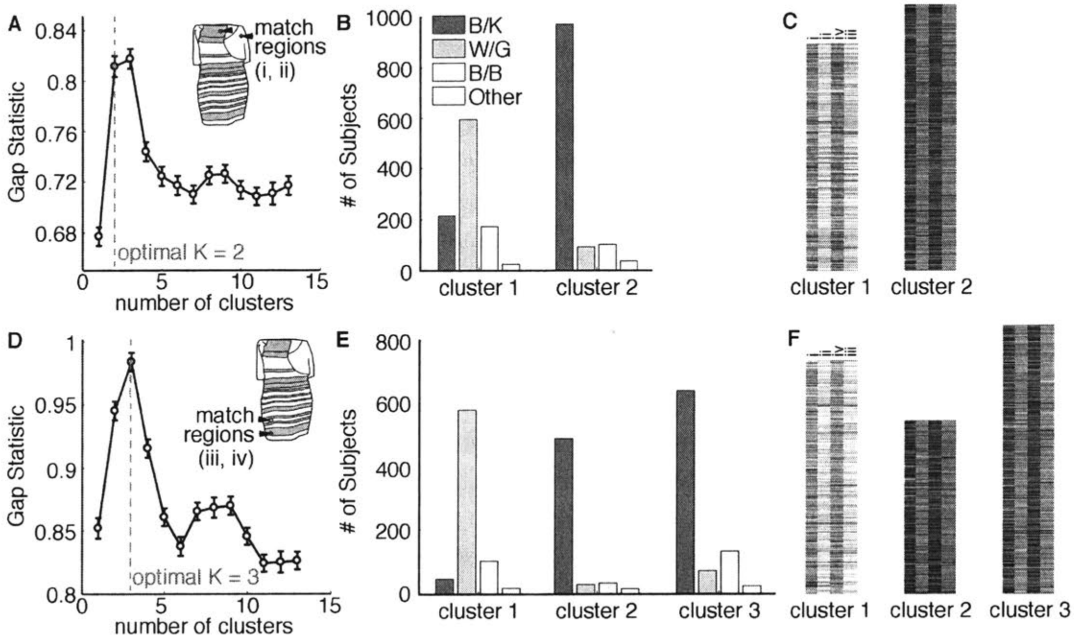


Figure 2. K-means clustering of color-matching data favor a 2 or 3 component model over a single-component distribution. Plots summarize the results from k-means clustering assessment (via the Gap method; Tibshirani et al., 2001) of the color matching data presented in Figure 1A (N=2200 subjects). A, The Gap statistic computed as a function of the number of clusters, for color matching data (PCA weights) obtained for the upper regions of the dress, using all the data from the online population. Dashed line indicates the optimal k clusters. B, Bar plot showing the cluster assignments of the color matches, binned by the color terms used by the subjects to describe the dress. C, RGB values of the color matches, sorted by cluster assignment from panel B; each thin horizontal bar shows the color matches for a single subject. D, E, F, as for panels A-C, but for color matches made to the bottom regions of the dress. The Gap analysis compares the within-cluster dispersion of a k-component model to its expectation under a null model of a single component. The algorithm seeks to identify the smallest number of clusters satisfying: Gap(k)≥GAPMAX−SE(GAPMAX), where k is the number of clusters, Gap(k) is the gap value for the clustering solution with k clusters, GAPMAX is the largest gap value, and SE(GAPMAX) is the standard error corresponding to the largest gap value. The optimal k solution for the distribution of upper dress regions color-matches is 2 clusters, and for the lower dress regions it is 3; confirming the suspected non-unimodality of the underlying population distribution.

The bar plots (**Figure 2B, E**) show the distribution of the different color terms assigned to each cluster, for the optimal *k* solution returned by the Gap analysis. The clustering algorithm assigned the majority of W/G reporters' matches to Cluster 1 (in both the upper and lower region analyses) and the majority of B/K reporters' matches to Cluster 2 (upper region analysis) or Clusters 2 and 3 (lower region analysis). Matches made by subjects who described the dress as Blue/Brown (B/B) or other colors outside of the main categories ('Other') were distributed more evenly across the clusters, even when 3 clusters were returned. Each thin band in the tapestries (**Figure 2C, F**) corresponds to the color matches made by a single subject, providing a visual snapshot of the success of the clustering algorithms in separating W/G and B/K reporters. These results show that perception of the dress in the population is categorical, not continuous, and reject the idea that reports of categorical perception are an artifact caused by a forced choice.

Categorical perception of the dress: How many categories?

The results of Figure 2 suggest that the underlying population may comprise two or three distinct categories. In our initial report, we argued that subjects who report the dress as "blue/brown" constituted a distinct third category, intermediate between the two main categories. **Figure 3, top row,** shows the spatial relationship of the optimal *k* cluster centroids ('x's) and the color-matching distributions (contours) for subjects grouped by their verbal report, for an analysis of the upper match regions (contours were created in MATLAB using *scatplot*[1]). The cluster centroids coincide with the center of the color-matching data for W/G and B/K subjects ('x's fall inside the contours; contours and centroids obtained with different halves of the data). Color matches made by subjects who reported B/B fell between these centroids (**Figure 3, top right panel**). These results suggest that the blue/brown report does not reflect a distinct category. Figure 3 bottom row shows the relationship between the optimal *k* cluster centroids for the bottom match regions, which returned three clusters. But none of the centroids fell within the contour capturing color matches made by B/B subjects (Figure 3, bottom right panel). These results show that the third category, when evident, is a subgroup of the population of observers who describe the dress as blue/black. In addition to the Gap analysis, we applied the Silhouette clustering criterion and the Calinski-Harabasz clustering criterion; these methods do not allow for a single-component solution. They returned an optimal solution of 2 clusters, for both the upper and lower match regions.

---

[1] http://www.mathworks.com/matlabcentral/fileexchange/8577-scatplot

[2] At first blush, there are some obvious similarities: 1) all the images in the set are composed of two low saturation, opponent colors; 2) one color component is usually perceived as achromatic, the other as chromatic (consistent with observers attributing the chromatic bias of one component to the illuminant); 3) all appear to have non-uniform light fields and poor exposure quality; and 4) all are tightly cropped, making judgments about scene geometry and

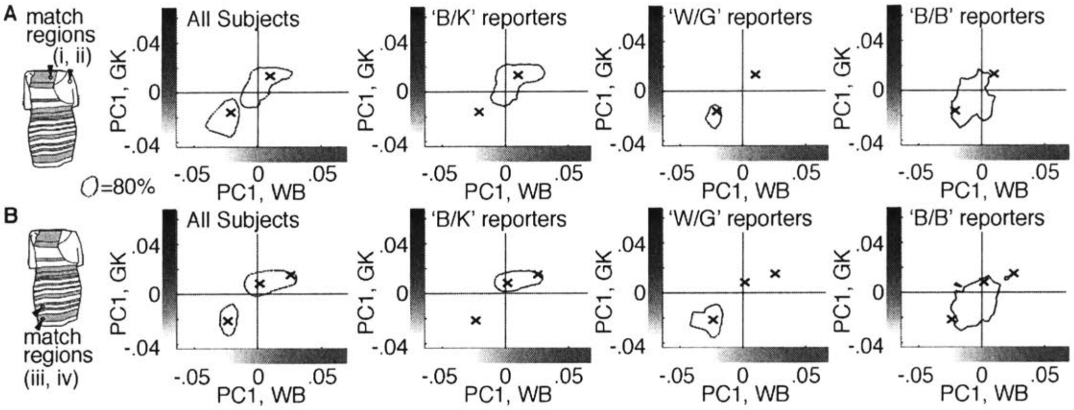


Figure 3. Comparison of color matching data (contours) with predictions from the k-means clustering solutions ('x's), sorted by subjects' verbal reports. Color-matching distribution contours and k-means cluster centroids derived from independent data sets for A, the color-matches made to upper regions of the dress (region i is plotted against region ii; PCA weights and PC axes from Figure 1A), and B, lower regions of the dress (iv, iii). Distribution contours were determined using one half of the data set (randomly sampled from the online subject pool); cluster centroids were determined using the left-out data (clustered using the optimal k identified in the Gap analysis from Figure 2). Individual plots show the contours encompassing the top 80% of color-matches for, from left to right, half the data from each group: all subjects; subjects who described the dress as blue and black (B/K); subjects who described the dress as white and gold (W/G); and subjects who described the dress as blue and brown (B/B). Within each row the same cluster centroids are re-plotted across the panels, reflecting the outcome of clustering the unsorted data set (i.e. independent of verbal reports).

Does prior exposure to the image change the number of categories manifest in the population? The color matching data for naïve versus not-naïve subjects were similar (Figure 1), suggesting that prior exposure had no impact on the number of categories in the population. **Figure 4** shows Gap statistical tests of the color-matching data to establish this conclusion. The optimal number of clusters for naïve subjects (N=1017), for either the top or bottom regions of the dress, produced an optimal cluster value of two (**Figure 4A**); the only color-matching data that produced more than two optimal clusters were those obtained on non-naïve subjects tasked with matching the lower part of the dress (N= 1183, three clusters, **Figure 4B**). The results of the analyses carried out using the Silhouette clustering criterion and the Calinski-Harabasz clustering criterion returned 2 clusters, regardless of whether the matches came from subjects with prior experience.

Together with the qualitative evaluation of the color-matching data we reported previously (Lafer-Sousa et al., 2015), these results show that the categories reported in the social media reflect true categories in the population and are not a result of the way the question was posed.

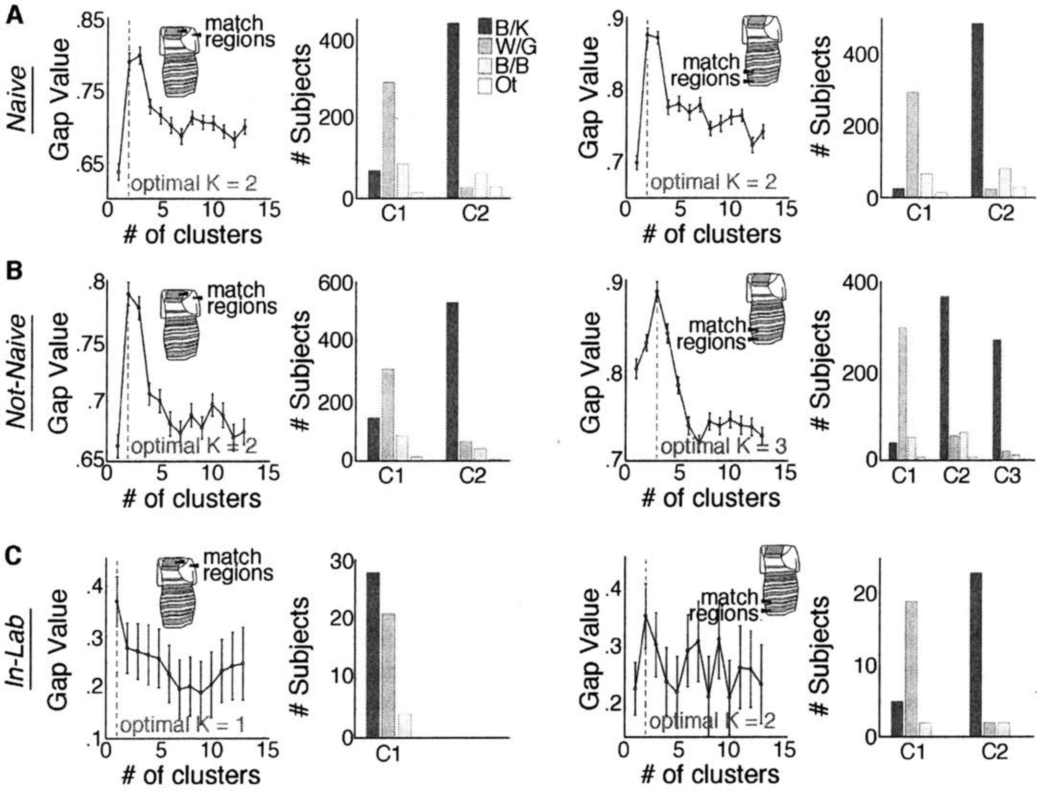


Figure 4. K-means clustering solutions for naïve, non-naïve, and in-lab subjects. The k-means clustering assessment was performed on the subset of subjects who A, had never seen the dress before participating in the study (N = 1017; top row), B, who had seen the dress before (N = 1183; middle row), and C, subjects who were tested under controlled conditions in the laboratory (N = 53; bottom row). Conventions as in Figure 2.

Categorical perception of the dress: Power analysis

The categorical nature of the population distribution appears, on first inspection, to contradict reports that the true population distribution is continuous. The discrepancy may be resolved by considering the large differences in the number of subjects in the different studies: conclusions of a continuous distribution were made using data from a less than two dozen subjects, whereas the present analysis depends on data from several thousand subjects. How many subjects is necessary to uncover the true population distribution? To address this question, we performed a power analysis by computing the optimal *k* using sub-samples of the data we collected, and bootstrapping (**Figure 5**). The variance around the predicted *k* decreases and the predicted *k* increases with increasing numbers of samples. That the optimal *k*=2, and not 1, becomes significant with about 125-180 subjects.

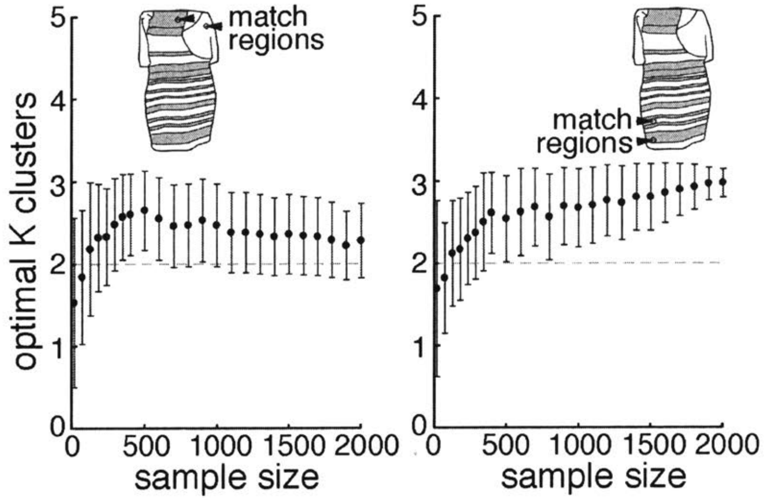


Figure 5. Power analysis. Plots show the results of k-means clustering (as in Figure 2) for a range of sample sizes (color-matching data randomly sampled from the online subject pool). For each sample size tested, the average Gap value derived from 100 bootstraps is plotted; error bars show the standard deviation. Left panel corresponds to upper dress region matches; right panel, to lower dress region matches. The variance around the predicted k decreases and the predicted k increases with increasing numbers of samples. That the optimal k=2, and not 1, becomes significant with about 125-180 subjects, and plateaus around 500 subjects.

The gap statistic obtained on all subjects (N=2200), for the upper match regions, yielded comparable values for 2 and 3 clusters (Figure 2); in this case, the optimal gap value is considered 2, because the gap method favors the smallest number of clusters satisfying the method's criterion. The gap statistic for the lower match regions was clearly distinguished as 3 clusters. This difference between the upper and lower match regions accounts for the difference in the results of the power analysis: for the upper regions, the optimal k converges between 2 and 3 clusters (and the error bars remain large even at big sample sizes); for the lower regions, the optimal k converges on 3 clusters, and the error bars get very small at big sample sizes. The additional cluster identified using color matches for the bottom region of the dress do not correspond to a discrete "blue/brown" category, but rather a subdivision of the "blue/black" category (Figure 3B, two centroids fall within the B/K reporters).

Our estimate of the number of subjects required to adequately assess the underlying population distribution was made using data obtained online; there is much higher variability in the viewing conditions and subject pool for experiments conducted online versus in lab. It is possible that the number of subjects needed to determine the true underlying population distribution would be lower for data obtained in lab, where the viewing conditions can be better controlled. In our prior study, we collected data in 53 subjects under controlled lab conditions. *K*-means clustering analysis of these data set neither rejected a single-component model nor confirmed a single-component model (**Figure 4C**). These results suggest that conducting the experiments in-lab confers no benefit in uncovering the population distribution.

Categorical perception of the dress: Comparing results obtained online and in lab

The results in Figure 4C show that even under controlled viewing conditions, samples of more than 53 participants are needed to reliably uncover the true population distribution. Nonetheless, the data collected in lab showed trends consistent with two underlying categories of observers. First, qualitative assessment of the color-matching plots (SI in (Lafer-Sousa et al., 2015)) show evidence of two clumps. Second, the optimal *k* for one of the two sets of regions (iii, iv) was two, even if this optimal value is not strongly distinguished from other values of *k* (**Figure 4C** left panel). That the optimal *k* is 2 becomes clearer when the analysis is run on data combining color matches for all regions tested, essentially doubling the data set, which returns an optimal *k* of 2 (data not shown). Third, the relative distribution of W/G to B/K observers among naïve participants was about the same for subjects tested on line versus in lab (Lafer-Sousa et al., 2015). And fourth, the average chromaticity of the color matches made by subjects in lab was consistent with those made by subjects online (**Figure 6**). Regardless of whether the data were obtained online or in-lab, the results showed the same pattern: compared to B/K subjects, W/G subjects reported not only higher luminance, but also higher values of u* (redness) and v* (yellowness), for all four regions tested. The strongest changes were in the luminance and v* dimensions. The comparability of data collected on-line versus in-lab is consistent with the idea that the factors that determine how one sees the dress are relatively high-level, divorced from the specific low-level conditions of viewing (such as the white balance, mean luminance, and size of the display).

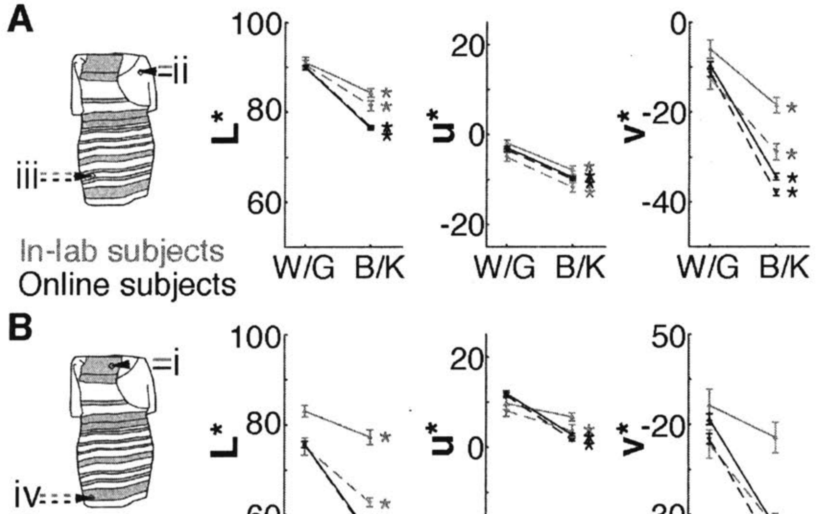


Figure 6. Color matches made by 'B/K' reporters and 'W/G' reporters differ in lightness and hue, under controlled conditions and online. A. Matches for the blue/white regions. B. Matches for the black/gold regions. Plots show the average lightness and hue components (CIELUV 1976: L*, u*, v*) of color-matches made by subjects reporting 'B/K' or 'W/G', tested online (black lines; N =1174; 770

‘B/K’, 404 ‘W/G’), and under controlled viewing conditions (red lines; N = 49; 28 ‘B/K’, 21 ‘W’/G’). Solid lines show data from the upper regions of the dress, and dashed lines show data from the lower regions; error bars show 95% C.I. Asterisks show cases for which ‘B/K’ and ‘W/G’ matches differed significantly (paired t-tests): in the online data, matches differed in all three color dimensions (u*, or “red-greenness”; v*, or “blue-yellowness”; and Luminance) (p-values <0.001); and in the in-lab data, matches differed in all but the u* and v* dimensions of one region (i) (p-values < 0.001 for all paired t-tests, except u* and v* of region i: p = 0.06, 0.2).

Categorical perception of the dress: Predicting color terms from color matches

The results in Figures 1-6 support the conclusion that the color of the dress photograph is resolved as one of two dominant categories, consistent with the initial social media reports. As an additional test of the hypothesis, we performed a multinomial logistic regression analysis to test the extent to which we could predict the colors a person would say the dress was, given the color matches a person makes. If, for example, B/B reporters represent a distinct/stable perceptual category, then a classifier (trained on independent data) should be able to distinguish B/B reporters from B/K and W/G, on the basis of color-matches alone. A multinomial logit model was fit using responses from a subset of the online subjects ($N_{train} = 549$) and tested on responses from left-out subjects ($N_{test} = 547$). **Figure** 7 shows the prediction outcomes of applying the model to the left-out color-matching data (N= 547). The results are sorted according to the verbal report given by the subjects. Across all participants who reported B/K, the average predicted probability of B/K category membership was 0.85 for naïve subjects (**Figure 7A**, top left) and 0.81 for not-naïve subjects (**Figure 7A**, top right). These probabilities were such that 92% of B/K reporters were classified correctly (**Figure 7B**, 93% of naïve, 89% of not-naïve); and 98% of W/G reporters were classified correctly (100% of naïve, 91% of not-naïve). The model was less successful in its classification of B/B reporters: 51% of B/B reporters were classified correctly (54% of naive, 45% of not-naive); and it failed to classify all ‘Other’ reporters as ‘Other’. Overall the model was accurate 85% of the time. This accuracy is very high and may underestimate the true accuracy: there was a delay of a few minutes between when subjects provided verbal reports about the dress color and when they did the color-matching experiment; the perception of the dress may have switched during this time.

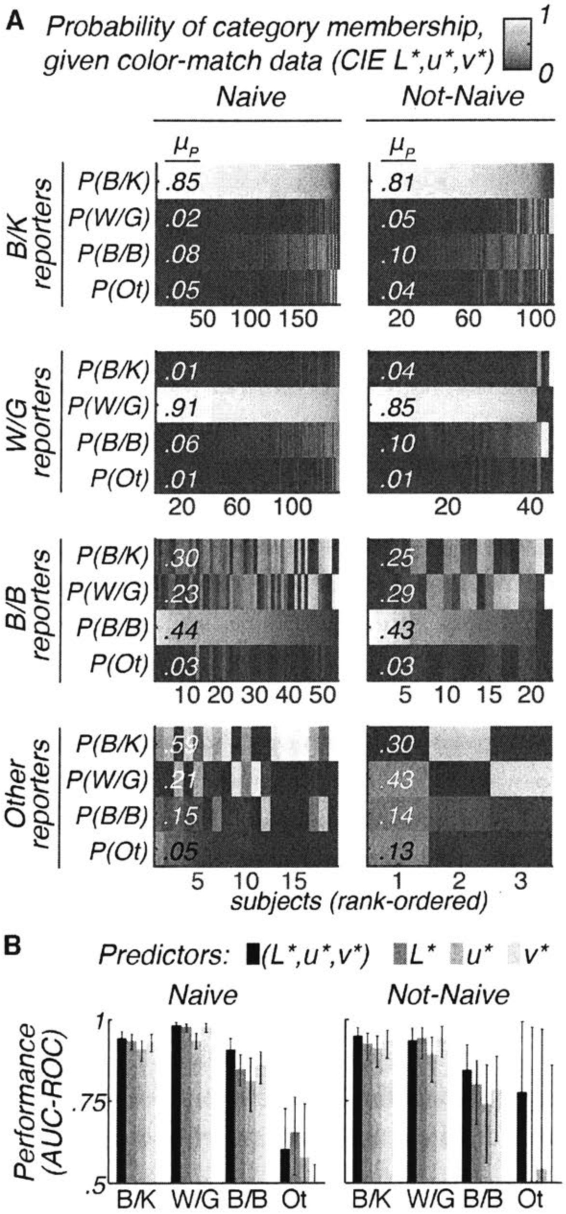


Figure 7. Categorical verbal reports can be predicted from color matches. Multinomial logistic regression was used to build nominal response models (classifiers). Four models were generated: the 'full' model was fit using the L*, u*, and v* components of subjects' color matches (to all 4 match regions) as predictors; three additional models were fit using either the L* or u* or v* component of subjects' matches as predictors. Models were fit with responses from a subset of the online subjects (half the subjects from Experiment 2, Ntrain = 549) and tested on responses from the left-out subjects (Ntest = 547). A, Predicted probability of category membership for the 'full' model. Each panel contains the results for data from individual (left-out) subjects, grouped by the verbal label they used to describe the dress (ground truth), and whether they had seen the dress prior to the study ('Naïve'). Each thin vertical column within a panel shows the results for a single subject: the colors of each row in the column represent the predicted probability that the subject used each of the categorical labels ('B/K', 'W/G','B/B','Other'); each column sums to 1. Subjects are rank-ordered by the predicted probability for

the ground-truth class. The average predicted probabilities for each response category are denoted by (μP). B, Bar plots quantifying classification performance (the area under the receiver operating characteristic curves, computed using the true positive rates and false positive rates), by category, for each of the 4 models. Error bars indicate 95% C.I. Values greater than 0.90 indicate 'excellent' performance; values between 0.75-0.90 ≈ 'fair to good' performance; values below 0.5-0.75 ≈ 'poor' performance. We compared the accuracy of the various models against each other using Matlab's testcholdout function: the 'L* only' model performed no better than the 'u* only' model (Naïve: $p = 0.5$; Not-Naïve: $p = 0.8$), or the 'v* only' model (Naïve: $p = 0.5$; Not-Naïve: $p = 0.3$). The 'full' model was more accurate than the 'L* only' model, but only among 'Naïve' subjects (Naïve: $p<0.001$; Not-Naive: $p = 0.09$). True positive rates (Sensitivity) for all four models are provided in **Table 1**.

		True positive rate				
Model predictor	Subjects	All	B/K	W/G	B/B	Other
(L*, u*, v*)	Naïve	86%	93%	100%	54%	0%
	Not-Naïve	83%	89%	91%	45%	0%
L*	Naïve	79%	96%	96%	2%	0%
	Not-Naïve	78%	91%	89%	5%	0%
u*	Naïve	78%	92%	89%	24%	0%
	Not-Naïve	79%	89%	89%	18%	0%
v*	Naïve	80%	89%	99%	30%	0%
	Not-Naïve	81%	86%	96%	36%	0%

Table 1. True positive rate of multinomial classifiers trained to predict verbal color reports from color matches (see Figure 7). Four models were generated: The full model was fitted using the L*, u*, and v* components of subjects' color matches as predictors; three additional models were fitted using either the L*, u*, or v* component of subjects' matches as predictors. Models were fitted with responses from a subset of the online subjects (half the subjects from Experiment 2, Ntrain = 549) and tested on responses from the left-out subjects (Ntest = 547). The table shows the true positive rates—(# of true positives)/(# of true positives + # of false negatives)—for each model, broken down by subjects' prior experience with the image (Naïve, Not-Naïve) and their verbal label (B/K, W/G, B/B, and Other).

Do the model results contradict the conclusions drawn in Figures 2 and 3 by providing evidence for three categories? While model accuracy of 51% for the B/B observers is above chance (chance = 25%), only about 25% of the correctly classified B/B reporters were classified by a high predicted probability for category membership, and many people who reported B/B were misclassified with high predicted probability as B/K or W/G (**Figure 7A**, third row). Most people who reported "other" were classified by high predicted probability as either B/K or W/G. Together with the results in Figures 2 and 3, these results suggest that the B/B designation is not a distinct category; it may constitute a transient state between B/K and W/G, which would be consistent with the properties of a bi-stable phenomenon.

The model results in Figure 7 were obtained using all color dimensions. When the model was fit using only the L, or u*, or v* components of subjects' color matches as the predictors, the classification accuracy remained high, between 78% and 82% depending on which dimension was used. This provides additional support for the contention B/K and W/G reporters are differentiated by both the lightness and hue of the matches they select.

Categorical perception of the dress: Switching perception from one category to another

The evidence presented in Figures 1-7 strongly suggests that the dress image is an ambiguous image that the visual system can interpret as one of two mutually exclusive categorical percepts. When viewing ambiguous shape images such as the Necker cube, subjects often report a change in their perception of the image, from one stable state to the other. But it has also been shown that knowledge of the fact that the image can be perceived in different ways can have a profound impact on whether subjects see the image flip. The on-line data we obtained included a diverse subject pool. Unlike most other on-line surveys of the phenomenon, participants were not recruited with links attached to media reports describing the dress. As a result, many of the participants in our study were entirely naïve to the image. Moreover, of those who were not-naïve, many had no knowledge of the actual color of the real dress, enabling us to test the extent to which subjects can change their perception of the dress, and if so, the impact on flip rates conferred by knowledge of the image's multi-stability.

We asked subjects who were not naïve to the image about their first encounter with the image. The distribution of verbal color reports corresponding to the first time non-naïve subjects viewed the dress (recalled from memory, **Figure 8, filled symbol**) showed a greater proportion of W/G than reported by the same subjects in response to our presentation of the image ($X^2 = 731$, $p < 0.001$). Some of these subjects must have switched their perception of the dress' colors since their first viewing; the results show that experience with the image biased the population towards the B/K percept. The relative proportion of B/K first-encounter reports is comparable to findings of other surveys and matches the proportion recovered in experiments of naïve subjects shown the image at 100% size (the solid symbol overlaps the open red circle, **Figure 8**). As we showed previously, reducing the size of the image from its native size on the internet also biased the population towards the B/K percept. These two factors (image size and experience) interact (ANOVA performed with the data from the Figure 8, open circles confirmed a main effect of scale, $p < 0.001$; a main effect of experience, $p < 0.001$; and their interaction, $p < 0.001$).

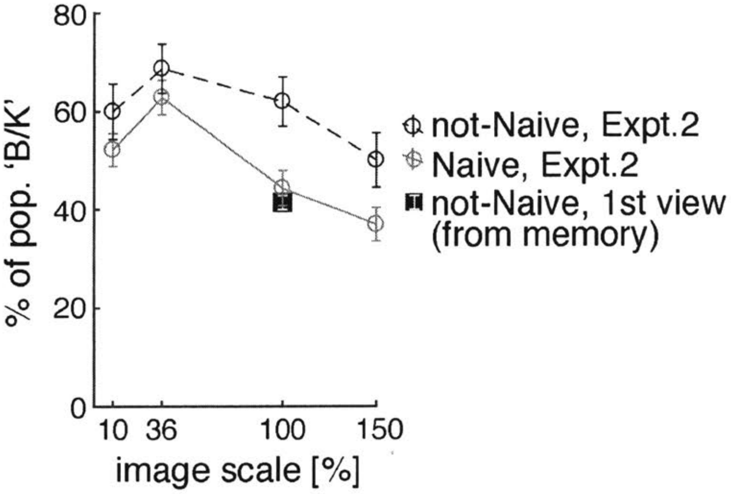


Figure 8. Image scale and prior exposure impact perception of the dress' colors. In Experiment 2 we varied the presentation scale of the image (different subjects saw different scales; 10%, 36%, 100%, 150%). And unlike Experiment 1, subjects were never shown an image in which the lighting cues were disambiguated. Open circles show the percent of subjects who reported 'B/K' at each scale. The red line shows data from subjects who had not seen the dress prior to participating in the study (10%, N=207; 36%, N=194; 100%, N=192; 150%, N=195). The black line shows data from subjects who had seen the dress prior to the experiment (10%, N=80; 36%, N=80; 100%, N=100; 150%, N=78). Viewing the image at a reduced scale and prior experience with the image both increased the proportion of subjects who reported B/K, and the two factors interacted (ANOVA performed with bootstrapped data from the open circles showed a main effect of scale, $p < 0.001$ main effect of experience, $p < 0.001$, and interaction, $p < 0.001$). Not-naïve subjects also reported on the colors they perceived when they first saw the image, recalled from memory (solid black square; data from not-naïve subjects in Experiments 1 and 2, N = 1037). The distribution of verbal color reports corresponding to the first time not-naïve subjects viewed the dress differed from the distribution of reports the same subjects provided in response to our presentation of the image, confirming that many had "flipped" (Chi-squared test of independence: X2 = 731, $p < 0.001$). Error bars show SEs (bootstrapping, sampling with replacement).

Among the 845 non-naïve subjects, 577 knew that the true dress color was blue and black, allowing us to ask how knowledge alters perception. Knowledge of the dress' color in real life dramatically altered the ratio of B/K to W/G responses in the population in a direction that favored B/K (**Figure 9A, B**). Moreover, knowledge of the true dress colors increased reports of flipping between B/K and W/G (**Figure 9C**). Among people who first saw W/G, yet who reported knowing the true colors of the dress, 42% had switched to B/K (we confirmed that these individuals were not merely reporting the dress' true colors while continuing to perceive W/G, by running their color matches through our classifier: only 9% of them were classified as W/G; 76% were classified as B/K, 13% as B/B). Among people who first saw W/G, and reported not knowing the true colors of the dress, only 14% had switched to B/K. These results show that knowledge of the dress' true colors impacted whether people could see the dress flip.

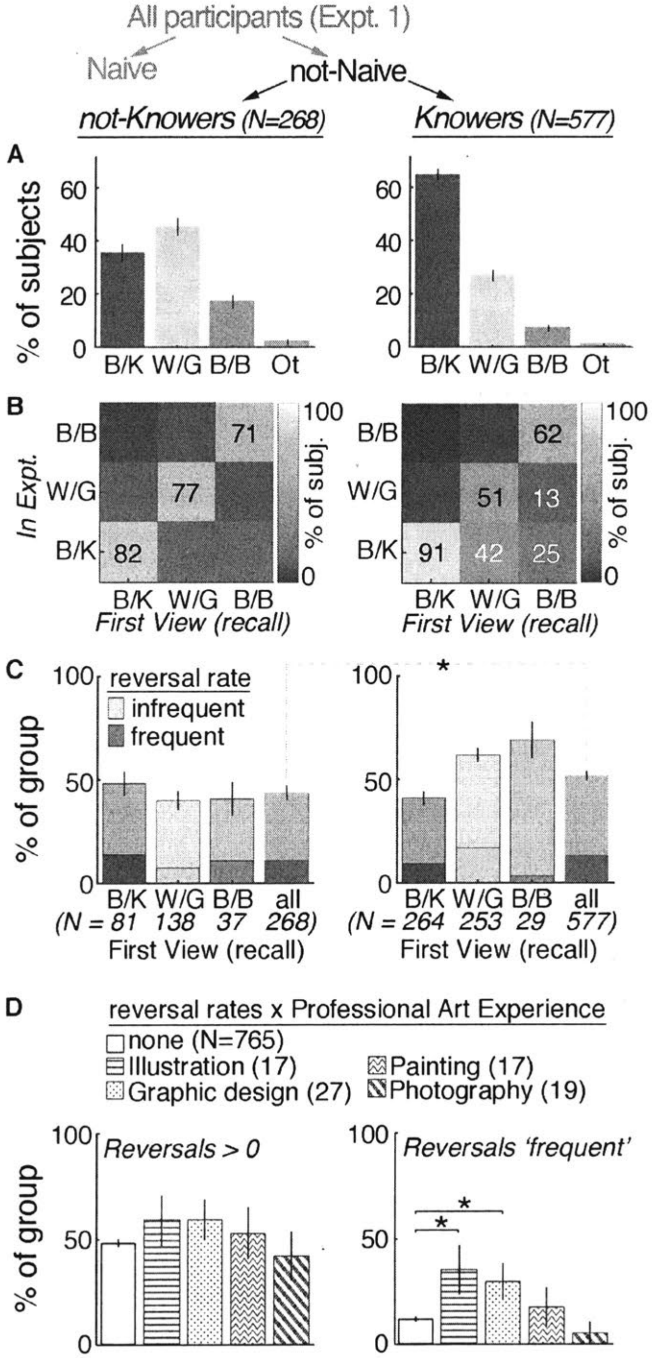


Figure 9. Familiarity with the dress image impacts subsequent viewings. A, The distribution of color reports (Experiment 1; image scale always 36%) from not-naïve subjects (N = 845) who had knowledge of the dress' colors in real life ('knowers', N = 577), differed from the distribution of reports from not-naïve subjects who did not ('not knowers', N = 268) (Chi-squared test for difference of proportions: $p < 0.001$; X2 = 66). B, Color reports of the dress during the experiment as a function of how subjects first perceived the dress (recalled from memory). Results for 'not-knowers' are predominantly along the x=y diagonal, reflecting dominance of initial stable state (though the presence of deviations was still significant: Chi-squared test of independence: X2 = 277 $p < 0.001$). Results for 'Knowers' showed more

substantial deviations from the diagonal, particularly for subjects who first saw W/G or B/B, reflecting a weakening of initial state and an increased dominance of B/K state (the color of the dress in real life) (Chi-squared test of independence: X2 = 394, $p < 0.001$). C, Quantification of self-reported reversal rates for 'knowers' and 'not knowers'; percentages computed within class (B/K, W/G, B/B, Other; class corresponded to 'first-view' percepts (from memory). Subjects were asked: "In your viewings prior to this study, did your perception of the dress colors ever change? Y/N" and "How often did you see it change? Frequently/Infrequently/Never." The proportion of subjects who reported changes was higher among knowers than not-knowers (gray asterisk, Chi-squared test for difference of proportions: p = 0.036; X2 = 4.4). Error bars indicate SEs (bootstrapping, sampling with replacement). D, Quantification of self-reported reversals as a function of professional art experience. Left panel shows the proportion of subjects who reported at least one reversal, sorted by professional art experience. Right panel shows the proportion of those subjects who reported that reversals were 'frequent.' Compared to non-artists, reports of 'frequent reversals' were significantly higher among subjects who indicated having professional art experience in the fields of Illustration (p=0.004, two-proportion z-test) and Graphic Design (p=0.005). Error bars indicate SEs (bootstrapping, sampling with replacement).

The distribution of observers across categories was different between participants with knowledge of the true dress color and participants who were naïve (compare Figure 1 of Lafer-Sousa et al with **Figure 9A**, right panel; $X^2 = 21$, $p < 0.001$; all data obtained with the 36% image). Specifically, the proportion B/K among participants with knowledge of the true colors (65%) was higher than the proportion of B/K among naïve participants (54%). Curiously, among the population who had seen the dress previously, but did not know its colors, we found roughly the same proportion of B/K observers as W/G observers (**Figure 9B**, left panel); this distribution was different than we observed for naïve participants, who were much more likely to see B/K ($X^2 = 23$, $p < 0.001$). Non-naïve no-knowledge individuals likely first saw the image in social media, where it was shown at a larger scale than we used in this set of experiments. We attribute the relatively higher levels of W/G reports among the non-naïve no-knowledge group, even though the image we showed was at the smaller size, to the fact that their first view likely established a prior about the colors of the dress that had not been updated with any subsequent knowledge of the dress' true colors.

On average, half of subjects reported experiencing the dress reverse at least once, while only 12% reported 'frequent' reversals (Figure 9C). Given that reversal rates in multi-stable perception can be influenced by cognitive factors like personality, creativity, and attention, we examined the proportion of subjects reporting reversals as a function of their professional art experience (subjects could indicate professional experience with *graphic design*, *illustration*, *photography*, *painting*, and *art history*). Among individuals who indicated professional *illustration* experience (N=17) 59% reported at least one reversal, with 35% reporting 'frequent' reversals (3x the proportion observed in the general population). Among individuals with *graphic design* experience (N=27), 60% reported at least one reversal, with 30% reporting 'frequent' reversals (2.5x the general population). Among painters (N=17), 53% reported at least one reversal, with18% reporting 'frequent' reversals (slightly more than the general population). Among photographers (N=19), 42% reported at least one reversal, with just 5% reporting 'frequent' reversals (less than the general population).

Together the analysis of the population responses of different categories of observers shows that (1) How you first saw the dress establishes a prior; (2) Knowledge of the colors of the dress in real life updates this prior, biasing it toward B/K; (3) Varying the image size systematically biases the percept (increasing image size increases W/G reporting; reducing image size increases B/K reporting); (4) Experience with the image over time, independent of knowledge of the dress's true colors, biases the population towards B/K; and (5) Reversal frequencies vary with professional art experience. These results uncover the important role played by both low-level perceptual features (such as image size) and high-level features (such as knowledge), in shaping how people perceive the colors of #theDress, and add to a growing body of evidence that exposure to social media can change the colors we see.

What accounts for the different ways in which the dress colors are seen?

We have argued that the multi-stability of the image derives specifically because the colors of the image align with the daylight locus (Brainard & Hurlbert, 2015; B.R. Conway, 2015; Lafer-Sousa et al., 2015). We hypothesized that in this context, multiple percepts become possible because the illumination cues in the image are ambiguous: subjects may infer either a warm or cool illuminant, and discount it accordingly. Consistent with this notion, color matches made by B/K and W/G reporters systematically shifted along the daylight locus, with B/K matches being shifted away from the warm end of the locus (consistent with discounting a warm illuminant) and W/G matches being shifted away from the cool end (consistent with discounting a cool illuminant) (**Figure 10A,B, left plots**). W/G matches were also lighter on average than B/K matches, consistent with the idea that subjects are not only discounting chromatic biases in the illuminant, but also lightness biases expected if one thought the dress were in shadow (**Figure 10A,B, right plots**). In our prior report, we tested the idea that illumination assumptions underlie the individual differences in color perception of the dress by determining how perception changes when the dress is embedded in a scene with disambiguated lighting (Lafer-Sousa et al, 2015). The results support the idea that observers who see B/K assume the dress is illuminated by a warm illuminant, while observers who see W/G assume that the dress is illuminated by a cool light source. Further evidence for this idea has been provided by others (Toscani et al, 2017; Witzel et al, 2017).

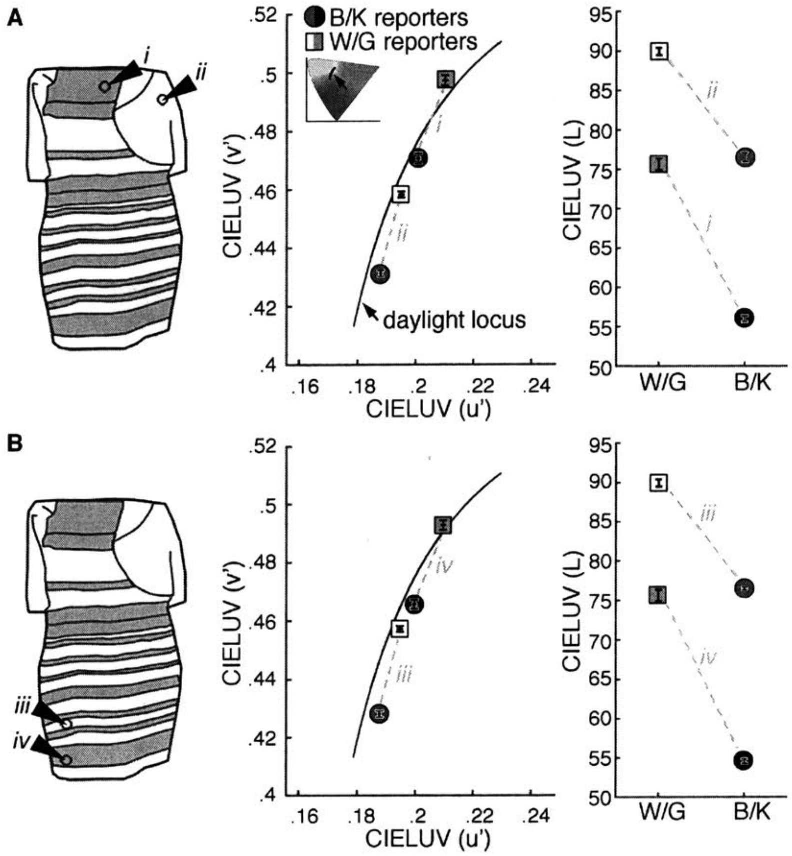


Figure 10. Color matches made by 'B/K' and 'W/G' reporters track the daylight locus. Chromaticity matches for the top regions of the dress (panel A) or bottom regions of the dress (panel B). Left graphs, mean hue (CIE u',v') of color-matches made by online subjects (B/K N =1174; 770; W/G N=404). Right plot, mean lightness of matches (CIE L). Error bars show the 95% C.I. of the mean. Data were grouped by the verbal report made by the subjects: W/G reporters (squares); B/K reporters (circles). The color of the symbol corresponds to the color term used by the subjects (blue, black, white, or gold). Black line shows the daylight locus. Inset = CIE 1976 u'v' color space with the daylight locus (arrow).

To further test the hypothesis, we applied a classification algorithm to data collected in our initial survey (1074 subjects online; 53 subjects in-lab) in which we asked people to explicitly report on the lighting conditions in the image (Supplementary Material of Lafer-Sousa et al, 2015 provides the full questionnaire). We ran two experiments to assess subjective experience of the lighting conditions. First, we asked subjects to rate the illumination on a scale of 1-5 for "cool" versus "warm"; second, we asked subjects to check off any of a number of possible verbal descriptors, including "dim", "dark", "bright", "warm", "cool", "blueish", "yellowish", etc (see Methods). Most subjects, regardless of their perception of the dress, reported the background illumination in the image to be warm (**Figure 11**, right panels; B/K and W/G reporter ratings did not differ, t-test: online subjects, p = 0.2; in-lab subjects, p = 0.5). But subjects who saw W/G differed from those who saw B/K in terms of their inference about the light on the dress itself: W/G percepts were associated with cool illumination, as if the dress were backlit and cast in shadow, while B/K percepts were associated with a warmer illumination, as if the dress were lit by the same global light as the rest of the room (**Figure 11**, left panels; t-tests, online subjects, p < 0.001; in-lab subjects, p < 0.001). These analyses quantify results in our initial report and are consistent with other findings (Chetverikov & Ivanchei, 2016; Toscani, Gegenfurtner, & Doerschner, 2017; Witzel, Racey, & O'Regan, 2017).

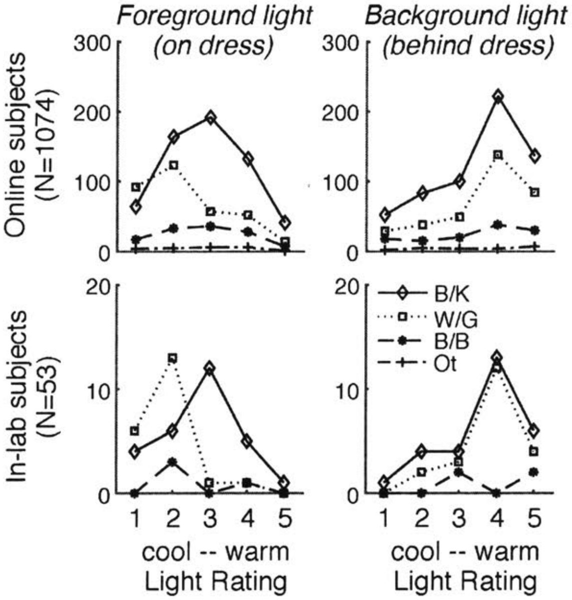


Figure 11. Temperature ratings of the light shining on the dress systematically differ as a function of percept. Subjects (1074 online; 53 in-lab) were asked to rate their impression of the light shining on the dress, and the light illuminating the background, on a scale of 1-5 for "cool" to "warm". Plots show illumination ratings for the light on the dress (left panels) and the light in the background (right panels), grouped by verbal report of the dress' colors (B/K, W/G, B/B, Other) (top row, online subjects; bottom row, in-lab subjects). Subjects' ratings of the light on the dress systematically differed as a function of percept (two-sample t-test comparing the ratings provided by subjects who saw the dress as B/K versus those who saw W/G, online subjects p < 0.001; in-lab subjects p = 0.002). Subjects' ratings of the background light did not differ as a function of percept (p = 0.2, online subjects; *p* = 0.5, in-lab subjects).

The results obtained using data on the warm/cool ratings were confirmed by an analysis of the descriptors that subjects used to characterize the lighting. W/G and B/K subjects were indistinguishable in the words they used to report on the illumination of the background (in **Figure 12A** the descriptors form one cluster in the right-hand bubble plots, where the most common descriptor was 'bright'), but showed strikingly different word choices when reporting on the illumination over the dress itself (descriptors form two clusters, **Figure 12A** left-hand bubble plots), with words like 'dim' and 'cool' corresponding to higher likelihoods of W/G reporting, and words like 'warm' and 'bright' to higher likelihoods of B/K reporting. The binary logistic regression using the lighting descriptors as predictors reliably classified B/K and W/G reporters (**Figure 12B**, classification histograms; online subjects: correct rate = 84%, d' = 1.83, $R^2 = 0.52$; in-lab: 100%, d' = +∞, $R^2 = 1$), and outperformed a constant only model (online subjects: $X^2 = 583$, $p < 0.001$; in-lab subjects: $X^2 = 66.9$, $p = 0.04$). Note that, once again, the results obtained on-line were consistent with those obtained in-lab.

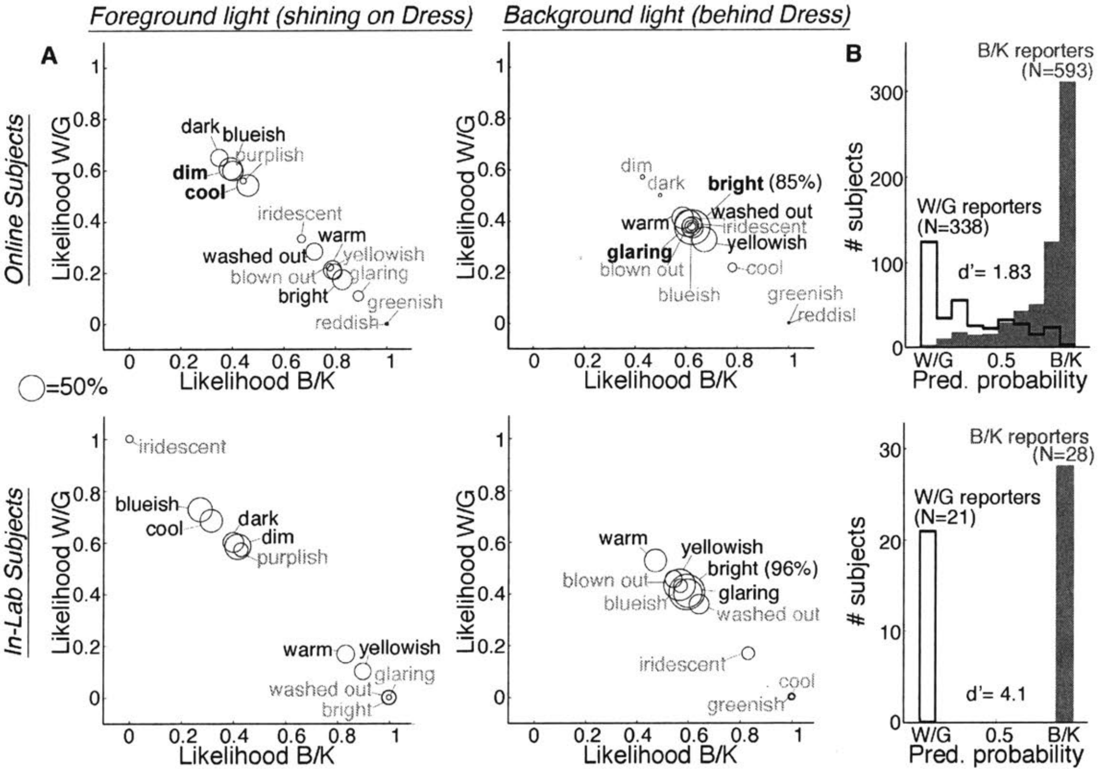


Figure 12. Subjects' color percepts of the dress are predicted by their inference of the lighting conditions. Subjects (1074 online, top row; 49 in-lab, bottom row; only subjects who reported either W/G or B/K are included in this analysis) were asked to characterize the light shining on the dress, and the light illuminating the background, by checking off any of a number of possible verbal descriptors from a list ('dim', 'dark', 'cool', 'blueish', 'bright', 'warm', 'yellowish', 'glaring', 'blown out', 'washed out', 'reddish', 'greenish', 'purplish', 'iridescent'). A, For each word in the list, the likelihood of being B/K = (#B/K reporters who used the term) / (#B/K + #W/G who used the term). The diameter of the bubble reflects the proportion of people in the population who used the term (#B/K who used it + #W/G who used it) / (#B/K reporters + #W/G reporters) (inset key = 50%). Bubble plots for the light shining on the dress ('foreground') (left panels) and the light in the background (right panels) (top row: online subjects; bottom row: in-lab subjects). B, Classification histograms for a binary logistic regression where the lighting descriptors were used as predictors to distinguish 'B/K' from 'W/G' reporters (online subjects: correct rate = 84% (85% for B/K, 80% for W/G), d' = 1.83, R2 = 0.52; in-lab: 100%, d= +∞, R2 = 1). A test of the full model against a constant model only was statistically significant, indicating that the predictors (the verbal descriptors of the lighting conditions) as a set reliably distinguish between B/K reporters and W/G reporters (online subjects: X2 = 583, p < 0.001; in-lab subjects: X2 = 66.9, p = 0.037).

Quantification of color-matching data obtained in 1127 people conducted mostly online using 'cue-rich' 'disambiguating' stimuli show that for both of the 'disambiguating' scenes ('cool' and 'warm' illumination simulations, **Figure 13A**), the optimal *K*-means clustering solution is a single component model (one cluster), consistent with the unimodal distribution of color-terms reported under these conditions (**Figure 13B-D**). When asked to rate the lighting conditions in the simulated scenes, subjects also conformed to ratings consistent with the lighting conditions cued (**Figure 14**). The results confirm that embedding the dress in scenes with unambiguous illumination cues resolves the individual differences in perception of the dress. The results also prove that the individual differences measured online, for the original dress image, are not simply the result of variability in viewing conditions.

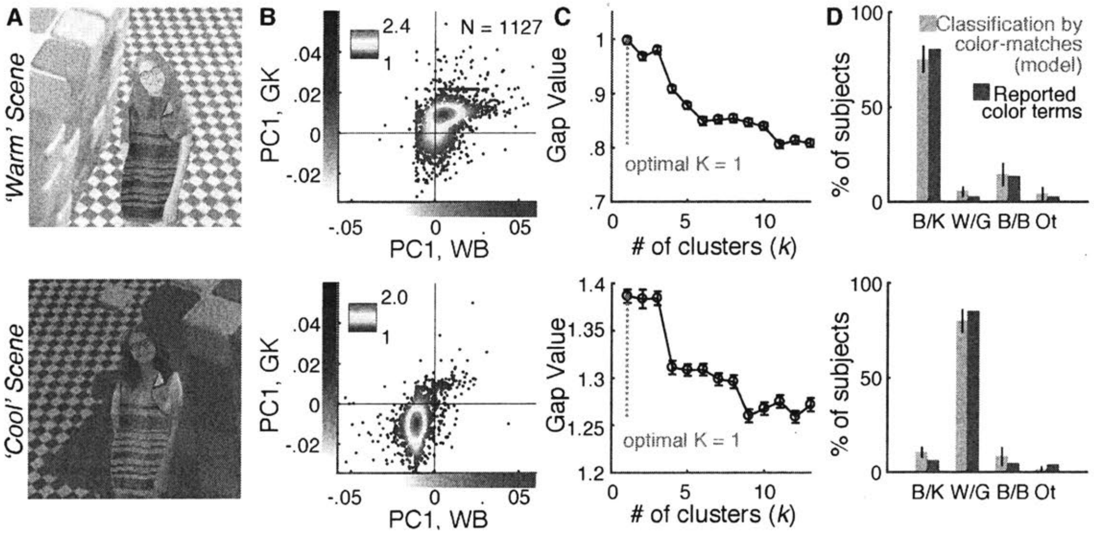


Figure 13. When the dress is embedded in scenes with overt cues to the illumination, the dress is perceived in a way that can be predicted from the overt cues: W/G when cool illuminant is cued; B/K when warm illuminant is cued. A, The dress was digitally embedded in simulated contexts designed to convey warm illumination (top row) or cool illumination (bottom). Cues to the illuminant are provided by a global color tint applied to the whole scene, including the skin of the model, but not the pixels of the dress. B, Distribution of subjects' (N=1127; 1074 online, 53 in-lab) color-matches. Conventions as for Figure 1. C, Results of k-means clustering assessments of the matching data from B (conventions as in Figure 2). The analysis favored a single component model (optimal k=1 cluster) for both the 'warm' and 'cool' scene distributions. D, Distribution of categorical percepts observed (dark gray bars) and the distribution predicted from the multinomial classifier (light gray bars; see Figure 7). Error bars show 95% CI. Photograph of the dress used with permission. Copyright Cecilia Bleasdale.

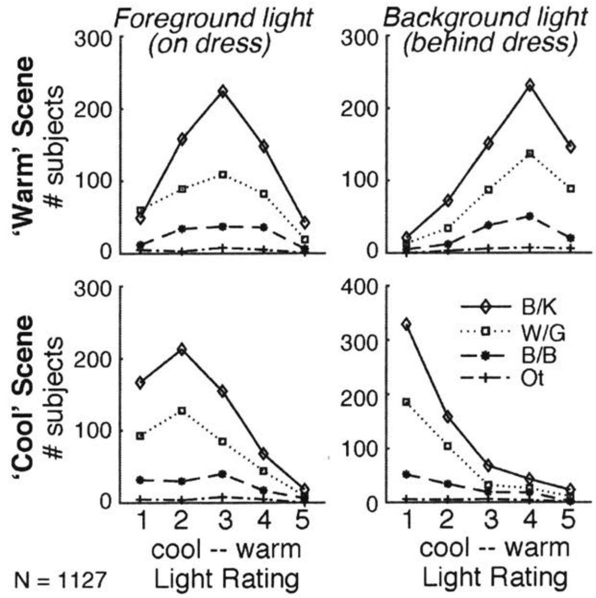


Figure 14. When the dress is embedded in scenes with overt cues to the illumination, the lighting is perceived in a predictable way. Subjects' ratings of the illumination in the simulated scenes from Figure 13 (online N=1127; 1074; in-lab N=53); warm illumination (top row) and cool illumination (bottom row) (conventions as in Figure 11; data is grouped by the color terms that subjects used to describe the original dress image). Subjects' ratings of the foreground light did not differ as a function of initial percept (p=0.9), nor did their ratings of the background light (p =0.8); two-sample t-tests comparing the ratings provided by subjects who originally reported B/K vs. those who originally reported W/G). Rating variance was higher for the original image (Figure 11) than for either test (cool, $p < 0.001$; warm, $p < 0.001$, F-test), but similar for the tests (p = 0.08).

What informs people's priors? Immediate prior exposure

The results above show that experience acquired over the medium term (knowledge of the true colors of the dress) impacts perception. What about experience in the very short term? Presumably priors on lighting are updating constantly, weighted by the reliability of the data. We sought to test whether exposure to a disambiguated version of the image that was digitally manipulated to provide clear information about the illuminant impacts how subjects see the original image.

We were able to address this hypothesis because we carried out two versions of Experiment 1 (on different participants) (**Figure 15**). In one version, subjects provided color matches for the original dress image after being exposed to the image depicting a warm illuminant; and in the second version, subjects provided color matches for the original dress image after being exposed to the image depicting a cool illuminant. We also conducted a separate experiment (Experiment 2) in which subjects were never exposed to the disambiguating stimuli. At the beginning of all experiments, subjects provided color terms for the original image. In our analysis we leveraged the discovery described above (**Figure 7**), that color matches reliably predict verbal reports. We compared the verbal reports made by subjects at the beginning of the experiment with the verbal reports we predicted they would make given their color matches at the end of the experiment. If exposure to a disambiguated stimulus updates a prior about the lighting condition, the predicted verbal reports made on the basis of color matching data should differ from the verbal reports made by the subjects for Experiment 1, but not Experiment 2; specifically, the predicted reports in Experiment 2 should be biased towards B/K when subjects were exposed to the warm scene, and to W/G when exposed to the cool scene.

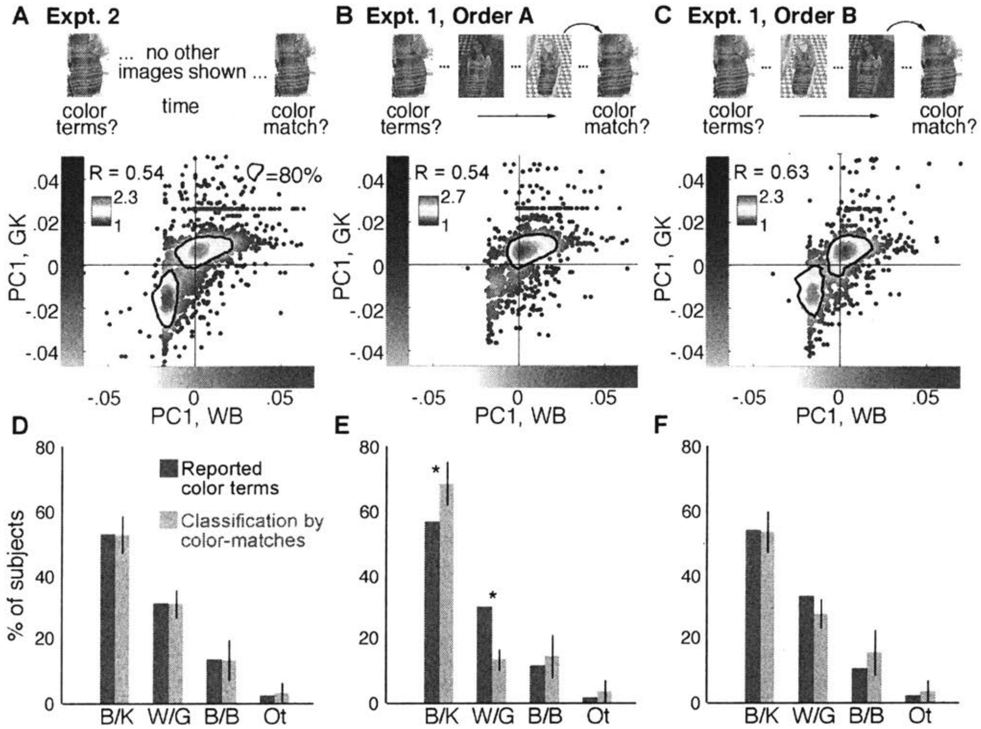


Figure 15. Priming with disambiguated scenes of the dress impacts subsequent viewings, providing evidence that priors on lighting conditions can be updated by short-term experience. Two experiments were conducted in separate groups of subjects. In Experiment 1, subjects gave color matches for the original dress image immediately after viewing the disambiguating stimuli from Figure 13; in Experiment 2, subjects were never exposed to the disambiguating stimuli. In Experiment 1, two groups of subjects saw the disambiguating stimuli in one of two orders that differed depending on which disambiguating stimulus immediately preceded the color-matching task. All subjects provided verbal reports prior to viewing the disambiguating stimuli. A, Analysis of data from Experiment 2 (online subjects N = 1126). Scatter plot (conventions as in Figure 1A) shows two peaks. B, Analysis of color-matching data from Experiment 1, Order A (N = 553; subjects viewed the simulated 'warm' scene, the 'B/K primer', directly before performing the color-matching task on the original image). K-means clustering returns 2 clusters, but there is one dominant peak. C, Analysis of color-matching data from Experiment 1, Order B (N = 523; subjects viewed the simulated 'cool scene, the 'W/G primer', directly before color-matching on the original image). Scatter plot has two strong peaks. D, Histograms comparing the distribution of categorical percepts recovered in Experiment 2 (control, no-primer) during the verbal task and the color-matching task. Dark gray bars show data from the verbal reporting task;

light gray bars show the distribution of verbal reports predicted from the color matches using the category response model (see Figure 7A; the classifier was trained on half the data and the plot shows classification for the other half). Error bars show 95% CI. Distributions do not differ. E, Bar plots showing the distribution of categorical percepts for Experiment 1, Order A. Dark bars show verbal reports collected prior to priming. Light bars show verbal labels predicted from color matches made following exposure to the B/K-primer. Asterisks indicate a significant shift from W/G to B/K reporting after B/K priming (dark bar falls outside of the 95% CI of the light bar). F, Histograms comparing distributions for WG-primer (Experiment 1, Order B). Photograph of the dress used with permission. Copyright Cecilia Bleasdale.

Figure 15A shows the results for the control case—no exposure to a disambiguating stimulus—and replicates the findings in **Figure 1A**: the density plot shows two strong peaks, corresponding to B/K and W/G reporters (Figure 15A is a subset of the data shown in Figure 1A). We deployed our classifier trained on independent data (**Figure 7**) to categorize observers on the basis of the color-matching data they provided. The distribution of verbal reports predicted by the classifier (light gray bars) is almost identical to the distribution of verbal reports that subjects provided (dark gray bars) (**Figure 15D**). **Figure15B** shows the results of Experiment 1 Order A, in which a separate set of subjects (N= 553) viewed the simulated warmly-lit scene immediately before they gave color matches for the original image. The distribution of verbal reports (obtained prior to color matching) is essentially indistinguishable from the distribution obtained in the control experiment (compare dark bars in **Figure 15E** with dark bars in **Figure 15D**), providing reassurance that we sampled a sufficient number of subjects to recover an accurate estimate of the population distribution. But compared to **Figure 15A**, the density plot in **Figure 15B** shows only one strong peak, which aligns with the color matches made by B/K subjects in the control experiment (the contour contains 80% of the data). These results show that more subjects reported B/K, and fewer reported W/G, than expected on the basis of the verbal reports that the subjects provided. K-means clustering returned two optimal clusters, showing that the W/G peak was still present, albeit diminished. The bar plots (**Figure 15E**) quantify the shift. **Figure 15C,F** shows the results for an independent set of subjects (N=523) who participated in Experiment 1, Order B. Unlike Order A, the data look similar to the control case—the density plot shows two strong peaks, and the distribution of verbal reports predicted from the color matches is not different from the distribution of verbal reports that subjects actually provided. These results show that priming subjects to see B/K influences people to see B/K, whereas priming people to see W/G has no effect.

What informs people's priors? Long-term exposure (chronotype)

The initial media reports, and many preliminary scientific studies, implied that a given observer's experience of the image was stable over time. The results described above support an alternative conclusion: that people can switch their perception of the image. In light of these results, we sought to revisit our first hypothesis about the origin of the individual differences in how people see the dress—that they reflect ingrained differences in the priors people hold about the spectrum of the illuminant. Most typical illuminants have chromaticities that fall along the daylight locus, which from a neutral point extends either towards a warm (orange) bias or a cool (blue) bias. We had hypothesized that cumulative life experience shaped the prior one holds about the spectrum of the illuminant, and that in the absence of strong cues to the illuminant, the visual system relies on these priors to discount the spectral bias in the illuminant. We speculated that the way in which the dress was seen could therefore be predicted by one's chronotype, with day-larks being more like to report W/G (having a prior biased toward blue-skylight illumination), and night-owls being more likely to report B/K (having a prior biased toward warm artificial light). We tested this idea by analyzing subjects' self-reported wake and bed times (hour of the day); these data were acquired at the time of our initial survey, and we queried many people who had no previous experience with the dress image. **Figure 16A** shows subjects' probability of being awake at any given hour of the day as a function of dress color report. The W/G distribution is very slightly phase shifted (to the left), relative to the B/K distribution, indicating earlier wake and bed times for W/G reporters ('day-lark' chronotype). To test the explanatory power of chronotype, we ran a binomial regression analysis, using the set of subjects' waking and sleeping hours as predictors (each hour of the day was assigned a value of 1 or 0) to distinguish 'B/K' from 'W/G' reporters (**Figure 16B**). The regression performed above chance for the data set collected in-lab, but did poorly for the online population (in-lab: correct rate = 81% overall, 79% for B/K, 86% for W/G; d' = 1.86; $R^2 = 0.54$; online: correct rate = 67% overall, 93% for B/K, 20% for W/G; d' = 0.68; $R^2 = 0.1$). We also ran a stepwise regression on the in-lab dataset to determine which predictors offered the most explanatory power. In each step, a given predictive variable was considered for addition or subtraction to the set of explanatory variables, evaluated on the basis of how well the new predictive variable improved the model. The set of explanatory variables included 6 specific hours of the day (7AM, 8AM, 9AM, 10PM, 11PM, 1AM) and two interactions terms (7AM with 11PM; 9AM with 1AM).

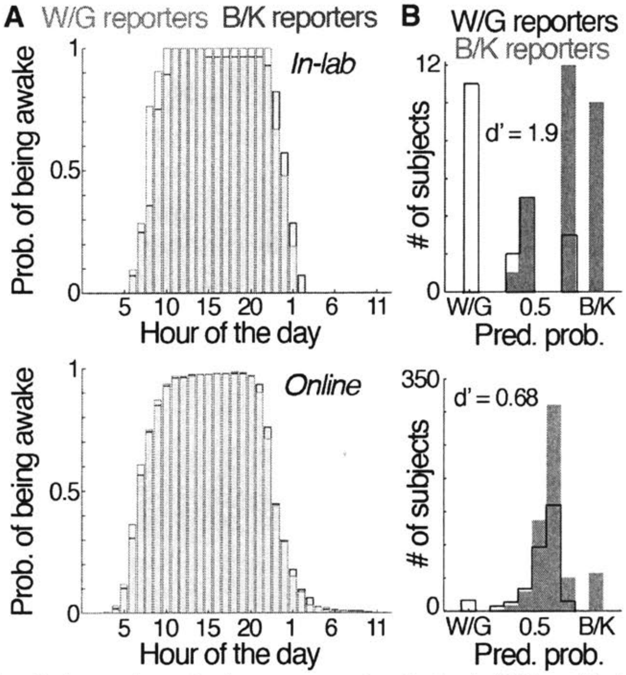


Figure 16. Relationship of chronotype to dress percept. A, Probability of being awake at any given hour of the day as a function of dress percept (B/K reporters, blue bars; W/G reporters, gold bars) (top row: in-lab, N = 49; bottom: online, N = 931, subjects from Experiment 1). For a given hour of the day, P(awake)B/K = (# B/K reporters awake) / (# B/K reporters); P(awake)W/G = (# W/G reporters awake) / (# W/G reporters). Computed from subjects' reported wake and bed hours. B, Classification histograms for binomial logistic regression over the set of subjects' waking and sleeping hours (where each hour of the day was assigned a value of 1 or 0 and served as a predictor) to distinguish 'B/K' from 'W/G' reporters. In-lab: overall hit rate = 81%; hit rate B/K, 79%; hit rate W/G, 86% (d' = 1.86; R2 = 0.54). Online: overall hit rate = 67%; hit rate for B/K, 93%; hit rate for W/G, 20% (d' = 0.68; R2 = 0.1).

The predictive power of wake/sleep cycle on dress percept shown in Figure 16 is weaker than we might have expected given the correlations reported by Wallisch (Wallisch, 2017). There are four differences between the two studies. First, Wallisch queried subjects using links tied to news reports of the dress image; we queried subjects using surveys that were independent of news reports of the phenomenon (and many of our subjects were entirely naïve to the image). Second, Wallisch asked people how they saw the dress when they first encountered it (a memory test); we asked people not only how they saw it initially, but also how they saw it during the study (a perception test). Third, Wallisch asked people to self-report the chronotype category (day-lark versus night owl); we asked people to report the hour that they woke up and went to bed. Fourth, Wallisch surveyed about 13000 subjects; we surveyed about 1000.

Could these differences account for the different results in the two studies? First, a person's report of their own behavior can differ depending on the circumstances of questioning, as political polling shows: people can report favoring different candidates depending on who is asking the question. Although the news stories used to recruit participants in the Wallisch study did not themselves describe the chronotype theory, many contemporary news stories did describe this theory (e.g., Rogers, 2015). It is conceivable that people who read the stories linked by Wallisch would have read more widely on the topic, and come across explicit descriptions of the chronotype theory. If so, could this information have biased the participants to report a chronotype that lined up with their dress percept? This explanation reverses the direction of causality (dress percept influencing reports of chronotype rather than chronotype influencing dress percept), and is plausible: the way we saw the dress powerfully shaped our identity. People formed "camps" defined by the different dress percepts that transcended race, geography and culture; it is not unreasonable to think that this new identity influenced how we report a subjective chronotype. Second, people's memories are fallible, and not always an accurate indication of perception (exit polls differ from actual election results). Third, it is not clear that self-reports of chronotype category are accurate indications of actual sleep-wake cycles. Querying the actual hour provides a direct test of the specific hypothesis that differential exposure to daylight influences the perceptual state. Querying category (day-lark versus night-owl) is indirect, and potentially inaccurate. One person's idea of a day-lark might be very different from another person's idea of a day-lark; the same category might correspond to very different amounts of daylight exposure. We are not aware of any studies showing that people would be more likely to misreport their wake/sleep cycle if queried as to the specific hour rather than the chronotype category. To the contrary, we suspect that people can accurately remember the hour they wake up and go to bed because they have several explicit markers of the time of day (alarm clocks, radio programs, TV shows). Finally, is sample size to blame? The large sample size we used was sufficient to demonstrate categorical perceptions, and should be sufficient to uncover phenomena that are as fundamental to the way we see the dress. Nonetheless it remains possible that our study lacked sufficient power to address the chronotype theory. One could address the question by re-analyzing the data from Wallisch, to determine how much data are required, conducting an analogous power analysis to the one we present in Figure 5. Taken together, the most parsimonious conclusion of the two studies is threefold: (1) chronotype has a modest impact on dress percept; (2) the way someone sees the dress can bias how they report their chronotype; and (3) as with other multi-stable images, many factors influence how the dress is perceived, and no single predictor has complete predictive power.

Do differences in where subjects report looking impact how they see the dress?

How a subject resolves the ambiguity in a multi-stable image such as the Necker cube can depend on where in the image a subject (first) looks (Chastain & Burnham, 1975; Kawabata et al., 1978) or attends (Peterson & Gibson, 1991), an idea proposed by Necker himself (Long & Toppino, 2004). Towards our goal of assessing the extent to which the dress is analogous to multi-stable shape images, we sought to address whether subjective reports of looking behavior differed between subjects who perceived W/G versus subjects who saw B/K. Subjects were asked to identify where in the image they spent most of their time looking, using a grid overlaying the image as a guide (**Figure 17A, left panel**). We appreciate that subjective reports of eye movements are not a good indicator of actual eye movements (Vo, Aizenman, & Wolfe, 2016; Wu & Cavanagh, 2016), and interpret these results as indications of what part of the image a subject considered most important (i.e. what part of the image they attended to). We focus our analysis on naïve subjects who viewed the dress as either W/G or B/K, and sorted the responses on the basis of their verbal reports (**Figure 17A, right panel**). We draw three conclusions. First, the patterns of responses for both groups of observers are not random: both W/G and B/K observers tend to identify regions in the center/top half of the image. This observation is consistent with other findings showing that, on average, subjects generally have a center bias (Tatler, 2007), and a top bias for visual search (Durgin, Doyle, & Egan, 2008). Second, the patterns of responses for the two groups of observers were different: B/K observers were more biased towards the upper right (shoulder) region of the dress, whereas W/G observers were more inclined to identify the center of the dress (within the fat horizontal brown stripe). Peak locations for each group are indicated with a bold line around the cell (**Figure 17A**). We performed a binomial regression on the reported looking locations and then did a receiver operating characteristic (ROC) analysis to compare the differences in perceived looking behavior between the two groups (**Figure 17C**); the analysis shows that the differences between the groups are significant. Third, both sets of observers identify a component in the dress image that corresponds to the chromatic element that defines their perceptual state: B/K observers tend to identify a blue region, whereas W/G observers identify a brown region. These results suggest that attention to different local components (spatial frequency and color statistics) within the image play a role in determining what colors are seen in the image.

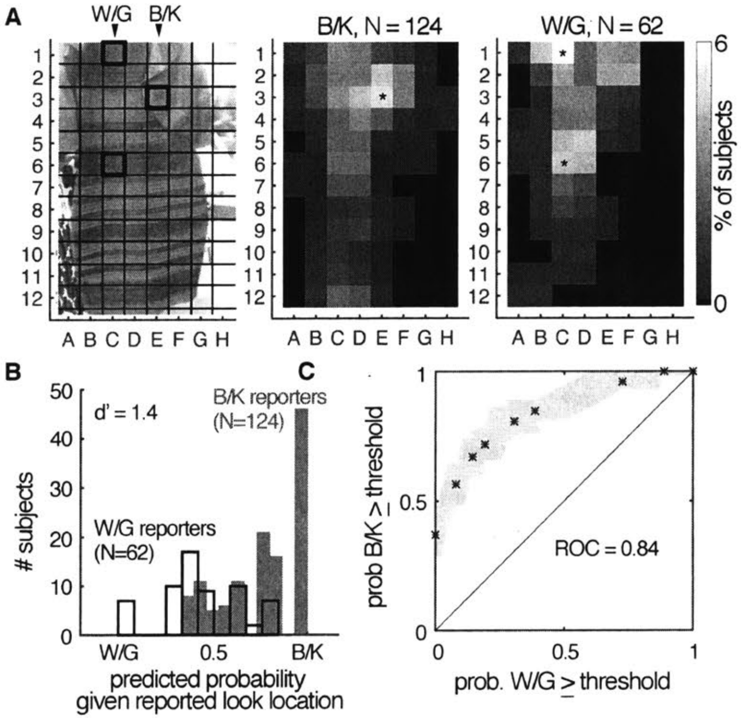


Figure 17. Self-reported looking behavior differs among B/K and W/G percievers. At the end of Experiment 1, subjects were asked to indicate the region of the image they felt they spent the most time looking at, using a grid overlaying the image as a guide. A, Heat maps show data sorted by verbal report (left, B/K reporters; right, W/G reporters). The analysis was restricted to naïve subjects (B/K =124 ; W/G N =62). The maps were smoothed with a moving average of 2x2 grid squares. Peak locations for each group are shown with a bold outline. B, Classification histograms from a binary logistic regression fit using self-reported looking locations. The looking predictor variable was encoded as a binarized vector corresponding to all 96 grid locations, where reported 'peak look location' was assigned '1', all other locations were assigned '0'). Histograms quanfity the probability of being B/K or W/G, given the looking reports. A test of the full model against a constant only model was significant (X2 = 76; p = 0.008; R2 = 0.33), showing that the looking report distinguishes B/K from W/G. C, ROC analysis confirms that the distributions in B are separable (ROC = 0.84; optimal threshold to assess classification accuracy is ~0.6). With a threshold of 0.6, the model hit rate was 76% for B/K, and 76% for W/G (d' = 1.44). Photograph of the dress used with permission. Copyright Cecilia Bleasdale.

Is skin-tint a sufficient cue to the illuminant?

To test the possible role of memory colors, in particular skin color, in color constancy we deployed a variation of our disambiguation paradigm (Experiment 1 **Figure 13**): we asked subjects to report on the colors they saw for versions of the image that were manipulated to include only a low-level cue (an illuminant-biased uniform surround), or skin with an illuminant-biased tint (**Figure 18**). The experiment was carried out under controlled viewing conditions in the laboratory (N = 53).

When asked to rate the lighting conditions in the cue-rich scenes, subjects conformed to ratings consistent with the lighting conditions cued (data not shown). Replicating our prior report (Lafer-Sousa et al., 2015), and the findings in **Figure 13**, and consistent with color-constancy predictions, regardless of how subjects initially perceived the dress' colors, most subjects' percepts conformed to the percept predicted by the color of the illuminant cued. The distribution of subject percepts changed dramatically when we changed the context from warm to cool ($p < 0.001$, paired t-test): when cool light was cued, most subjects reported WG (**Figure 18A**, left panel); when warm light was cued, most subjects reported BK (**Figure 18A,** right panel). To test whether a low-level sensory mechanism like receptor adaptation or local color contrast, long argued to play a powerful role in color constancy (Brainard & Wandell, 1992; Hurlbert & Wolf, 2004; Land, 1986), could disambiguate the dress' colors, we superimposed the isolated dress on uniform fields matched to the mean chromaticity of the cue-rich scenes. If low-level sensory adaptation or local color contrast is sufficient to achieve color constancy, surrounding the dress by a yellowish field should induce a uniform BK percept, while surrounding the dress by a bluish field should induce a WG percept. Contrary to this prediction, when the dress was surrounded by the uniform chromatic backgrounds, subjects' percepts did not conform to the percept predicted by the color of the background; the distribution of subject percepts was the same for both backgrounds ($p = 0.71$, paired t-test) (**Figure 18B**). The cue-rich scenes provided significantly better color constancy than the uniform surrounds (summary statistics provided in **Table 2**). Taken together, these results suggest that a strictly low-level mechanism is insufficient to elicit a stable color percept of the dress.

To test whether skin chromaticity is by itself a sufficient cue to achieve good color constancy, we presented the dress superimposed on the woman on a white background and tinted her skin according to the spectral bias of the illuminants simulated in the cue-rich scenes. Consistent with the hypothesized role of skin memory color in discounting the illuminant from a scene, we found that subjects' color percepts were predicted by the color bias of the skin tint. The distribution of subject percepts changed dramatically between the warm and cool conditions ($p = 0.002$, paired t-test): when the tint cued cool light, most subjects reported WG (**Figure 18C**, left panel); when the tint cued warm light, most subjects reported BK (**Figure 18C,** right panel). The presence of tinted skin alone was significantly more effective than the uniform illuminant-biased background in achieving good color constancy, and was as effective as the cue-rich scene (**Table 2**). These results suggest people can use skin chromaticity to recover information about the spectral composition of the illuminant to achieve color constancy

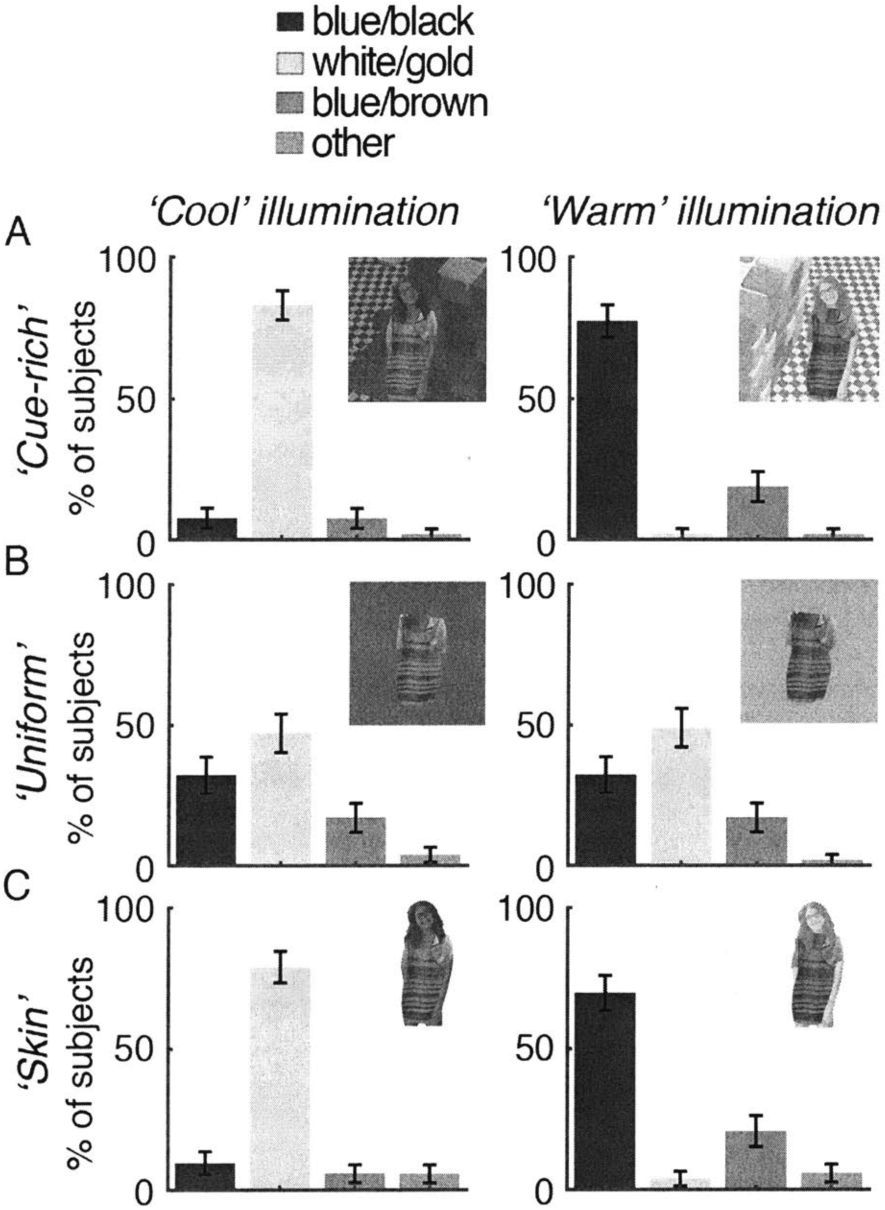


Figure 18. Behavioral evidence that people can achieve color constancy by using human skin as a reference to ascertain the spectral bias of the light source. Histograms showing subjects' reports of the dress' color when it was embedded in simulated contexts designed to convey cool illumination (left column) or warm illumination (right): A, 'cue-rich' scenes, B, uniform illuminant-biased chromatic backgrounds, C, skin with illuminant-biased tints, in isolation. The color of the bar corresponds to the verbal reports (key at top). Insets show stimulus thumbnail (see SI Image Appendix for full size reproductions). Error bars were calculated using bootstrapping (10,000 bootstrapped samples). N = 53 subjects. See Table 2 for McNemar's chi-square tests comparing goodness of constancy achieved across stimulus conditions. Photograph of the dress used with permission. Copyright Cecilia Bleasdale.

Table 2. McNemar's chi-square tests comparing goodness of constancy achieved in different stimulus conditions (see Figure 18). McNemar's test is a within-subjects z-test of equality of proportions for repeated measures. Each test compares the proportion of subjects that did or did not conform to the percept cued in stimulus condition X vs. the proportion of subjects that did or did not conform to the percept cued in stimulus condition Y.

Test Comparison	X^2	p
cue-rich scene (warm) VS. uniform background (warm)	**22.04**	**$3*10^{-6}$**
cue-rich scene (cool) VS. uniform background (cool)	**15.43**	**$8*10^{-5}$**
skin tint (warm) VS. uniform background (warm)	**15.04**	**$1*10^{-4}$**
skin tint (cool) VS. uniform background (cool)	**15.06**	**$1*10^{-4}$**
skin tint (warm) VS. cue-rich scene (warm)	2.25	0.1
skin tint (cool) VS. cue-rich scene (cool)	0.25	0.6

3.2.5. Discussion

This study addresses the extent to which the dress image is multi-stable. To be "multi-stable", it is widely assumed that the image must appear to change state ("flip" or reverse) on a short timescale (the Necker cube is typically seen to pop out, and then a few seconds later, to recede). But a review of the literature, discussed below, suggests that rapid flipping is not a necessary feature of multi-stability. Rather, the key properties of multi-stable stimuli are that they have more than one plausible percept; and that the alternative percepts are mutually exclusive. By these criteria, the results presented here provide the first evidence for a multi-stable color image. Quantitative analysis of color-matching data from a large, diverse population shows that the dress image is experienced categorically as either 'Blue and Black' (B/K) or 'White and Gold' (W/G). A power analysis shows that several hundred observers are needed to uncover the categorical nature of the population distribution. The two perceptual states were distinguished not only by differences in lightness, as suggested previously (Gegenfurtner et al., 2015), but also by differences in hue. This discovery is important: a multi-stable image with states defined only by differences in lightness might not, arguably, be considered a multi-stable *color* image. Classifiers trained on the color-matching data performed very well: given your color matching data, our algorithm can very likely report how you describe the dress. These results not only support the idea that the dress is multi-stable, but also provide an important example whereby subjective color experiences can be predicted from objective colorimetric data.

Color matches made by the population of B/K observers covered a larger region of color space than the matches made by W/G observers. This result may reflect the asymmetry in blue/yellow perception: changes in yellowness are more detectable than changes in blueness (Winkler, Spillmann, Werner, & Webster, 2015). Color matches to the bottom region of the dress gave rise to three clusters, instead of two. One of us (BRC) initially saw the dress as blue/brown, and so we were initially inclined to interpret the third cluster as demarking a stable "blue/brown" category. But the cluster peak falls outside of the color matches made by observers who report blue/brown. Seeing the dress as blue/brown appears to be a transient state. Our classifier accurately classified only half of B/B reporters, and only with relatively low probability. Among all groups of subjects, those reporting B/B (and "other") also reported the highest rates of changing their minds about the color of the dress. Rather than identifying a stable B/B category, the additional cluster subdivides the people who see the dress as B/K into separate dark-blue reporters and light-blue reporters. This subdivision of "blue" is reminiscent of some languages that carve blue into distinct light and dark categories (Winawer et al., 2007).

The results we obtained in-lab were generally consistent with those obtained on-line, suggesting that the neural operations that give rise to different perceptions of the dress are largely invariant to low-level stimulus features such as mean luminance and white-point, which vary among different displays. These results contribute to a growing body of work that exploits the greater power and participant diversity afforded by online experiments (Wilmer et al., 2012).

#theDress: Evidence that it is multi-stable

When the image first went viral on the internet, many people stated that they had a stable perception of the dress and were surprised that others saw it differently (Rogers, 2015). The apparent lack of reversibility is evidence against the notion that the image is analogous to multi-stable shape images such as the Necker cube. But is a person's experience of the dress really fixed? We found that about 50% of subjects reported a reversal prior to participating in our experiment; about 12% reported that their experience of reversals was 'frequent'. We did not ask subjects if their perception of the dress flipped while they performed our experiment, but several subjects tested in-lab spontaneously reported that it did. Perceptual reversals of the dress are probably not as rare as initially suggested in the popular press. The general assumption that one's perception of the image is fixed may partly be accounted for by the fact that people do not look at the image for very long. Given our own experience, we suspect that requiring subjects to maintain engagement with the image for longer periods of time would promote reversals; extended, continuous exposure promotes reversals for other multi-stable images (Leopold, Wilke, Maier, & Logothetis, 2002). Consistent with the role of "top-down" factors in determining how multi-stable shape images are seen (Scocchia et al., 2014), we found that knowledge of the dress' color in real life increased the likelihood that subjects experienced a reversal, as did having professional illustration and graphic design experience. Over time the population has gravitated towards seeing the dress as B/K, the colors of the dress in real life. That perception has drifted towards the "true" colors of the dress is reminiscent of some other ambiguous images, such as the Dalmatian image (Gregory, 1970) and Mooney faces, where knowledge of what is being depicted has a profound (and stable) impact on how the ambiguous image is subsequently seen (Lupyan, 2017). But knowledge of the dress' true colors is not by itself decisive; many people who know the dress' true colors, maintain that the image appears W/G, suggesting the dress is more similar to a multi-stable image than a simple ambiguous image.

Although we find some evidence that people can change their mind about the color of the dress, most observers say that the image does not flip spontaneously or frequently, which appears to contradict the idea that the dress is multi-stable. But is frequent perceptual flipping a necessary feature of multi-stability? Perhaps not. The reversal frequency of multi-stable images varies from individual to individual and stimulus to stimulus; the ease of reversibility depends on (among other factors) ambiguity type and prior knowledge. For example, content reversals and figure-ground reversals occur with higher frequency than perspective reversals (Kosegarten & Kose, 2014). Some multi-stable images, such as the rotating dancer, are notoriously "sticky" (some people never spontaneously see the dancer change direction). Moreover, subjects rarely report spontaneous reversals unless they are told that the images can be seen in different states, even for the Necker cube (Rock & Mitchener, 1992). And even knowledge of multi-stability is not always sufficient to induce reversals (Kosegarten & Kose, 2014). That most people's experience of the dress image is relatively stable is not sufficient to reject the idea that the image is multi-stable. The "stickiness" of the dress image may simply indicate that the visual apparatus prioritizes stable and consistent color perception, which may explain why so few examples of multi-stability in color have been discovered.

The perception of multi-stable images is impacted not only by top-down factors such as prior knowledge, but also by bottom-up factors (Long & Toppino, 2004) such as the local image statistics at the point of gaze (Chastain & Burnham, 1975; Kawabata & Mori, 1992; Kawabata et al., 1978). Reducing differences in texture between the dress and the surround (by blurring or shrinking the image) increased reports of B/K (Lafer-Sousa et al., 2015). When asked to identify the region they looked at most, subjects who reported B/K selected a blue region, whereas subjects who reported W/G picked a gold region. Thus all subjects tended to discount as achromatic the component that they thought they were not looking at. Where one looks will bias the interpretation of color statistics across a scene (Toscani, Valsecchi, & Gegenfurtner, 2015) and influence color induction (Brenner, Granzier, & Smeets, 2007; Hansen & Gegenfurtner, 2005), providing an explanation for the looking behavior: B/K reporters focused on the right edge of the dress, where color judgments would reflect stronger induction by the warm background. W/G reporters identified regions in the center of the image, away from the warm background. Curiously the most common region identified across all subjects tended to cover a large area of uniform color (B/K: the broad blue shoulder; W/G: the wide dark stripe), perhaps reflecting the importance of low spatial frequencies in color constancy (Dixon & Shapiro, 2017). Although these analyses suggest that bottom-up factors shape how we see the dress, it is possible that where subjects say they look is caused by how they see the dress' colors.

#theDress: Assumptions about the illumination that are updated over short and long time scales explain the individual differences

In our preliminary report, we argued that the image is consistent with two different interpretations: that the dress is in shadow, illuminated by bluish light (in which case people see the dress as W/G), or that the dress is illuminated by a warm light (in which case people see it as

B/K). Here we tested this hypothesis directly by analyzing the results of the subjects' reports on the lighting conditions. Indeed, the largest factor that determined how the dress was seen is how subjects interpreted the illumination: subjects that thought the dress was in shadow were much more likely to see the dress as W/G. Our results show that the inferences about the lighting correspond to whether subjects segment the scene into one or two 'frameworks' (Zdravkovic, Economou, & Gilchrist, 2012): on the one hand, B/K reporters make similar judgments about the light illuminating the background and the light illuminating the foreground (one 'framework': bright, warm direct/global illumination); on the other hand, W/G reporters make dissimilar judgments about the background and foreground lighting (two 'frameworks': bright, warm background light; but dim/cool foreground light, consistent with cast shadow).

The two different interpretations of the dress' colors are enabled because the chromaticities of the pixels within it are restricted to the colors of the daylight axis (Lafer-Sousa et al., 2015): when viewed in isolation, all the pixels are either bluish or orangeish. There is no other chromatic information available for the visual system to resolve the lighting conditions. We argued that people must rely not only on other cues to sort out the lighting conditions, but also on priors about what they assumed the lighting conditions to be. These priors are presumably established over both a short time frame (what time of day is it now? What illuminant were you recently experiencing?), and over a longer time frame (what sort of light are you most often exposed to?). We tested for priors acquired over both time scales. Priming subjects with an unambiguous version of the image caused subjects to make color matches consistent with the unambiguous version of the image, regardless of how they said they saw the image initially. These results provide evidence that subjects flipped their perception during the experiment, and support for the idea that the brain updates its illumination priors as soon as reliable information about the lighting is obtained. But curiously, we saw these effects only in one direction: priming with a warm illuminant predisposed people to make color matches consistent with B/K, but priming with a cool illuminant had little impact. We do not have an explanation for the asymmetry, although it is consistent with three other observations showing that W/G is less stable over time. The B/K report is increased by: (1) reducing the image size; (2) any prior exposure to the image; and (3) knowledge of the color of the dress in the real world.

In the initial media reports of the dress, we outlined a "chronotype" theory to account for the individual differences (Lafer-Sousa et al., 2015; Rogers, 2015). According to this theory, day-larks see the dress as W/G, while night-owls see it as B/K. The idea was that day-larks are more exposed to a blue spectral bias in the illuminant, and so are more likely to discount the blue component of the dress and see the dress as W/G; night-owls are more exposed to incandescent light with an orange spectral bias, and are therefore more likely to discount the warm component in the image and see it as B/K the (Lafer-Sousa et al., 2015; Rogers, 2015). The explanatory power of chronotype in our study was larger for the in-lab data than the on-line data. All in-lab subjects in our study were residents of the Boston area, and so were matched for regional light cycle and atmospheric conditions. The participants in the online experiments were distributed across the globe, which may account for the noisier results obtained in this population.

Nonetheless, the predictive power of wake/sleep cycle on dress percept that we report here is somewhat weaker than we would have expected given the correlations reported by Wallisch (Wallisch, 2017). One salient difference between the studies is the way in which subjects were recruited: Wallisch recruited subjects through online links attached to news reports of the image. While we favor the idea that different chronotypes underlie the individual differences in perception of the dress, it is possible that the causal link is the reverse. People formed "camps" defined by the different dress percepts, which transcended race, geography and culture; it is not unreasonable to think that this new identity influenced how a person reported their chronotype in the study by Wallisch. Taken together, the most parsimonious conclusion of the two studies is threefold: (1) chronotype has a modest impact on dress percept; (2) the way someone sees the dress can bias how they report their chronotype; and (3) as with other multi-stable images, many factors influence how the dress is perceived, and no single predictor has complete predictive power.

#theDress: A tool for understanding color

Finally, we showcase the dress image as a tool to test the role of memory in color constancy. An estimate of the spectrum of the illuminant can theoretically be achieved by performing gamut correlation between an observed object and a memory gamut for that object (Ling & Hurlbert, 2008). In particular, skin has been proposed as a potential cue for color constancy, due to its stable statistics (Hurlbert, 2007; Bianco & Schettini, 2012; Crichton et al., 2012). Using digitally manipulated versions of the dress, we show that a color tint applied to the skin of a model made to look like she is wearing the dress was sufficient for observers to infer the spectral bias of the illuminant, and achieve a predictable percept of the dress' colors. Placing the dress on a uniform colored background that matched the spectral bias of a light source was insufficient for subjects to disambiguate the colors of the dress. These results provide a striking demonstration that color constancy exploits more than local color contrast. The importance of skin for color constancy might have been predicted not only by the ubiquity of skin in visual experience, but also by the observation that humans have precise memory colors and preferences for skin, and show little tolerance for color deviations in reproductions of skin (Bartleson & Bray, 1962; Chauhan, Xiao, Yates, & Wuerger, 2015; Hunt, Pitt, & Winter, 1974; Sanders, 1959; Smet, Ryckaert, Pointer, Deconinck, & Hanselaer, 2011; Yendrikhovskij, Blommaert, & Ridder, 1999). Moreover, people are most sensitive to skin color adjustments that roughly align with the spectral bias of natural illuminants (the daylight axis) (Wuerger, Chauhan, Sohaib, Yates, & Xiao, 2016).We only tested illumination cues with Caucasian skin, assessed mostly by Caucasian participants. But given computational analysis of the statistics of different skin types (Crichton et al., 2012), we expect our results would generalize across races. We carry skin with us everywhere, so it is perhaps not surprising that the visual system uses it like a white-balance card in photography, to correct the bias in the illuminant.

3.2.6. Data Availability

All data and sample analysis code are available on a public-accessible website (https://github.com/rlaferso/-TheDress).

3.2.7. References

Bartleson, C. J., & Bray, C. P. (1962). On the preferred Reproduction of Flesh, Blue-Sky, and Green-Grass Colors. *Photographic Science and Engineering, 6*, 19-25.

Bianco, S., & Schettini, R. (2012). Color constancy using faces. *IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA. 2012*, 65-72.

Bloj, M. G., Kersten, D., & Hurlbert, A. C. (1999). Perception of three-dimensional shape influences colour perception through mutual illumination. *Nature, 402*(6764), 877-879.

Brainard, D. H., & Hurlbert, A. C. (2015). Colour Vision: Understanding #TheDress. *Curr Biol, 25*(13), R551-554.

Brainard, D. H., Longere, P., Delahunt, P. B., Freeman, W. T., Kraft, J. M., & Xiao, B. (2006). Bayesian model of human color constancy. *J Vis, 6*(11), 1267-1281.

Brainard, D. H., & Wandell, B. A. (1986). Analysis of the retinex theory of color vision. *J Opt Soc Am A, 3*(10), 1651-1661.

Brainard, D. H., & Wandell, B. A. (1992). Asymmetric color matching: how color appearance depends on the illuminant. *J Opt Soc Am A, 9*(9), 1433-1448.

Brenner, E., Granzier, J. J., & Smeets, J. B. (2007). Perceiving colour at a glimpse: the relevance of where one fixates. *Vision Res, 47*(19), 2557-2568.

Chastain, G., & Burnham, C. A. (1975). The first glimpse determiens the perception of an ambigous figure. *Perception and Psychophysics, 17*(3), 221-224.

Chauhan, T., Xiao, K., Yates, J., & Wuerger, S. (2015). Estimating discrimination ellipsoids for skin images. *Journal of Vision 15*( ), 820.

Chetverikov, A., & Ivanchei, I. (2016). Seeing "the Dress" in the Right Light: Perceived Colors and Inferred Light Sources. *Perception, 45*(8), 910-930.

Chichilnisky, E. J., & Wandell, B. A. (1995). Photoreceptor sensitivity changes explain color appearance shifts induced by large uniform backgrounds in dichoptic matching. *Vision Res, 35*(2), 239-254.

Conway, B. R. (2009). Color vision, cones, and color-coding in the cortex. *Neuroscientist, 15*(3), 274-290.

Conway, B. R. (2015). Why do we care about the colour of the dress? *The Guardian*, http://www.theguardian.com/commentisfree/2015/feb/27/colour-dress-optical-illusion-social-media.

Conway, B. R. (2016). Processing. *in Experience: Culture, Cognition, and the Common Sense, Caroline A. Jones, David Mather, and Rebecca Uchill, eds. (Cambridge MA: MIT Press and the MIT Center for Art, Science & Technology, 2016): 86-109.*

Crichton, S., Pichat, J., Mackiewicz, M., Tian, G., & Hurlbert, A. C. (2012). Skin chromaticity gamuts for illumination recovery. *Society for Imaging Science and Technology, Conference on Color in Graphics, Imaging, and Vision, Amsterdam, the Netherlands.*

D'Zmura, M., & Lennie, P. (1986). Mechanisms of color constancy. *J Opt Soc Am A, 3*(10), 1662-1672.

Dixon, E. L., & Shapiro, A. G. (2017). Spatial filtering, color constancy, and the color-changing dress. *J Vis, 17*(3), 7.

Drissi Daoudi, L., Doerig, A., Parkosadze, K., Kunchulia, M., & Herzog, M. H. (2017). The role of one-shot learning in #TheDress. *J Vis, 17*(3), 15.

Durgin, F. H., Doyle, E., & Egan, L. (2008). Upper-left gaze bias reveals competing search strategies in a reverse Stroop task. *Acta Psychol (Amst), 127*(2), 428-448.

Dutour, E.-F. (1760). Discussion d'une question d'optique. *l'Academie des Sciences. Memoires de Mathematique et de physique presentes par Divers Savants, 3*, 514-530.

Ellis, S. R., & Stark, L. (1978). Eye movements during the viewing of Necker cubes. *Perception, 7*(5), 575-581.

Fisher, G. H. (1967). Preparation of ambiguous stimulus materials. *Perception and Psychophysics, 2*, 421-422.

Foster, D. H., Nascimento, S. M. C. (1994). Relational colour constancy from invariant cone-excitation ratios. *Proc. R. Soc. Lond. B Biol. Sci., 257*(115-121).

Gegenfurtner, K. R., Bloj, M., & Toscani, M. (2015). The many colours of 'the dress'. *Curr Biol, 25*(13), R543-R544.

Gegenfurtner, K. R., & Kiper, D. C. (2003). Color vision. *Annu Rev Neurosci, 26*, 181-206.

Gregory, R. (1970). The intelligent eye. *McGraw-Hill, New York.*

Hansen, T., & Gegenfurtner, K. R. (2005). Classification images for chromatic signal detection. *J Opt Soc Am A Opt Image Sci Vis, 22*(10), 2081-2089.

Hedrich, M., Bloj, M., & Ruppertsberg, A. I. (2009). Color constancy improves for real 3D objects. *J Vis, 9*(4), 16 11-16.

Hunt, R. W. G., Pitt, I. T., & Winter, L. M. (1974). The Preferred Reproduction of Blue Sky, Green Grass and Caucasian Skin in Colour Photography. *J. of Photo. Sci., 22*, 144-149.

Hurlbert, A. (2007). Colour constancy. *Curr Biol, 17*(21), R906-907.

Hurlbert, A., & Wolf, K. (2004). Color contrast: a contributory mechanism to color constancy. *Prog Brain Res, 144*, 147-160.

Ishihara, S. (1977). Tests for colour blindness. Tokyo: Kanehara Shuppen Company.

Kawabata, N., & Mori, T. (1992). Disambiguating ambiguous figures by a model of selective attention. *Biol Cybern, 67*(5), 417-425.

Kawabata, N., Yamagami, K., & Noaki, M. (1978). Visual fixation points and depth perception. *Vision Res, 18*(7), 853-854.

Khang, B. G., & Zaidi, Q. (2002). Cues and strategies for color constancy: perceptual scission, image junctions and transformational color matching. *Vision Res, 42*(2), 211-226.

Klink, P. C., van Wezel, R. J., & van Ee, R. (2012). United we sense, divided we fail: context-driven perception of ambiguous visual stimuli. *Philos Trans R Soc Lond B Biol Sci, 367*(1591), 932-941.

Kosegarten, J., & Kose, G. (2014). Seeing reversals in ambiguous images: to know or not to know? *Perceptual and motor skills: perceptoin, 119*(1), 228-236.

Kraft, J. M., & Brainard, D. H. (1999). Mechanisms of color constancy under nearly natural viewing. *Proc Natl Acad Sci U S A, 96*(1), 307-312.

Lafer-Sousa, R., Hermann, K. L., & Conway, B. R. (2015). Striking individual differences in color perception uncovered by 'the dress' photograph. *Curr Biol, 25*(13), R545-546.

Land, E. H. (1986). Recent advances in Retinex theory. *Vision Res, 26*(1), 7-21.

Leopold, D. A., & Logothetis, N. K. (1999). Multistable phenomena: changing views in perception. *Trends Cogn Sci, 3*(7), 254-264.

Leopold, D. A., Wilke, M., Maier, A., & Logothetis, N. K. (2002). Stable perception of visually ambiguous patterns. *Nat Neurosci, 5*(6), 605-609.

Ling, Y., & Hurlbert, A. (2008). Role of color memory in successive color constancy. *J Opt Soc Am A Opt Image Sci Vis, 25*(6), 1215-1226.

Long, G. M., & Toppino, T. C. (2004). Enduring interest in perceptual ambiguity: alternating views of reversible figures. *Psychol Bull, 130*(5), 748-768.

Lotto, R., B., & Purves, D. (2002). The empirical basis of color perception. Consciousness and Cognition 11(4): 609-629

Lupyan, G. (2017). Changing What You See by Changing What You Know: The Role of Attention. *Front Psychol, 8*, 553.

Peterson, M. A., & Gibson, B. S. (1991). Directing spatial attention within an object: altering the functional equivalence of shape descriptions. *Journal of Experimental Psychology, 17*(1), 170-182.

Podvigina, D., N., & Chernigovskaya, T., V. (2015). Top-Down Influences to Multistable Perception: Evidence from Temporal Dynamics. International Scholarly and Scientific Research & Innovation 9(11): 3849-3852

Rock, I., & Mitchener, K. (1992). Further evidence of failure of reversal of ambiguous figures by uninformed subjects. *Perception, 21*(1), 39-45.

Rogers, A. (2015). The Science of Why No One Agrees on the Color of This Dress. *Wired,* http://www.wired.com/2015/02/science-one-agrees-color-dress/.

Sanders, C. L. (1959). Color preferences for natural objects. *Illuminating Engineering, 54*, 452-456.

Schwartz, J. L., Grimault, N., Hupe, J. M., Moore, B. C., & Pressnitzer, D. (2012). Multistability in perception: binding sensory modalities, an overview. *Philos Trans R Soc Lond B Biol Sci, 367*(1591), 896-905.

Scocchia, L., Valsecchi, M., & Triesch, J. (2014). Top-down influences on ambiguous perception: the role of stable and transient states of the observer. *Front Hum Neurosci, 8*, 979.

Smet, K. A. G., Ryckaert, W. R., Pointer, M. R., Deconinck, G., & Hanselaer, P. (2011). Colour appearance rating of familiar real objects. *Color Research and Application 36* 192-200.

Stiles, W. S. (1959). Color vision: the approach through increment threshold sensitivity. *Proc Natl Acad Sci U S A, 45*, 100-114.

Tatler, B. W. (2007). The central fixation bias in scene viewing: selecting an optimal viewing position independently of motor biases and image feature distributions. *J Vis, 7*(14), 4 1-17.

Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of data clusters via the Gap statistic. *Journal of the Royal Statistical Society B, 63*, 411-423.

Toscani, M., Gegenfurtner, K. R., & Doerschner, K. (2017). Differences in illumination estimation in #thedress. *J Vis, 17*(1), 22.

Toscani, M., Valsecchi, M., & Gegenfurtner, K. R. (2015). Effect of fixation positions on perception of lightness. *Human Vision and Electronic Imaging Xx, 9394.*

Vo, M. L., Aizenman, A. M., & Wolfe, J. M. (2016). You think you know where you looked? You better look again. *J Exp Psychol Hum Percept Perform, 42*(10), 1477-1481.

von Kries, J. (1878). Beitrag zur Physiologie der Gesichtsempfindugen. *In: MacAdam DL, editor. Sources of color science, 1st edn. Cambridge, MA: MIT press.*, 101–108.

Wallisch, P. (2017). Illumination assumptions account for individual differences in the perceptual interpretation of a profoundly ambiguous stimulus in the color domain: "The dress". *J Vis, 17*(4), 5.

Webster, M. A., & Mollon, J. D. (1995). Colour constancy influenced by contrast adaptation. *Nature, 373*(6516), 694-698.

Webster, M. A., & Mollon, J. D. (1997). Adaptation and the color statistics of natural images. *Vision Res, 37*(23), 3283-3298.

Wilmer, J. B., Germine, L., Chabris, C. F., Chatterjee, G., Gerbasi, M., & Nakayama, K. (2012). Capturing specific abilities as a window into human individuality: the example of face recognition. *Cogn Neuropsychol, 29*(5-6), 360-392.

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proc Natl Acad Sci U S A, 104*(19), 7780-7785.

Winkler, A. D., Spillmann, L., Werner, J. S., & Webster, M. A. (2015). Asymmetries in blue-yellow color perception and in the color of 'the dress'. *Curr Biol, 25*(13), R547-548.

Witzel, C., Racey, C., & O'Regan, J. K. (2017). The most reasonable explanation of "the dress": Implicit assumptions about illumination. *J Vis, 17*(2), 1.

Wu, D. A., & Cavanagh, P. (2016). Where are you looking? Pseudogaze in afterimages. *J Vis, 16*(5), 6.

Wuerger, S., Chauhan, T., Sohaib, A., Yates, J., & Xiao, K. (2016). The sensitivity of the human visual system to subtle skin tone changes *Journal of Vision, 16*(12), 643.

Yendrikhovskij, S. N., Blommaert, F. J. J., & Ridder, H. (1999). Color Reproduction and the Naturalness Constraint. *Color Res. Appl, 24*, 54-67.

Zdravkovic, S., Economou, E., & Gilchrist, A. (2012). Grouping illumination frameworks. *J Exp Psychol Hum Percept Perform, 38*(3), 776-784.

Image (3) - 'cue-rich; cool illumination'

Image (4) - ‘Uniform background; warm illumination’

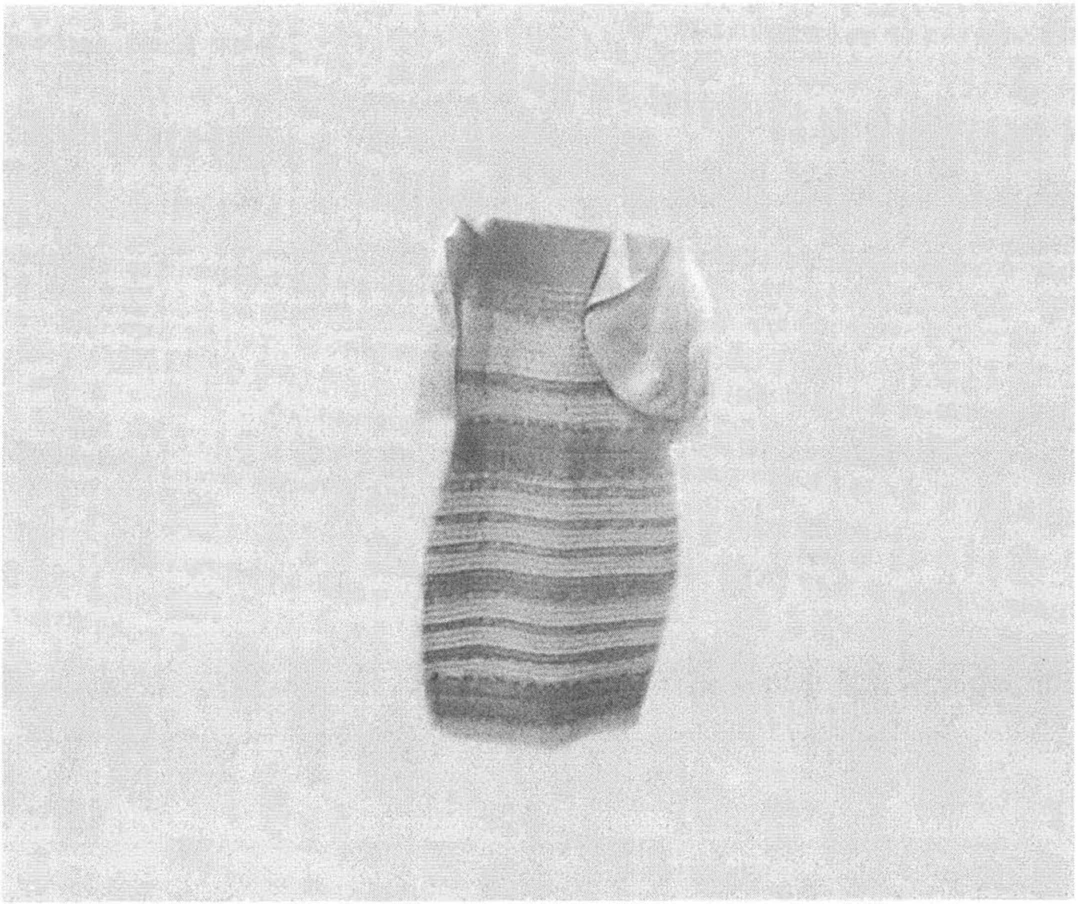


Image (5) - ‘Uniform background; cool illumination’

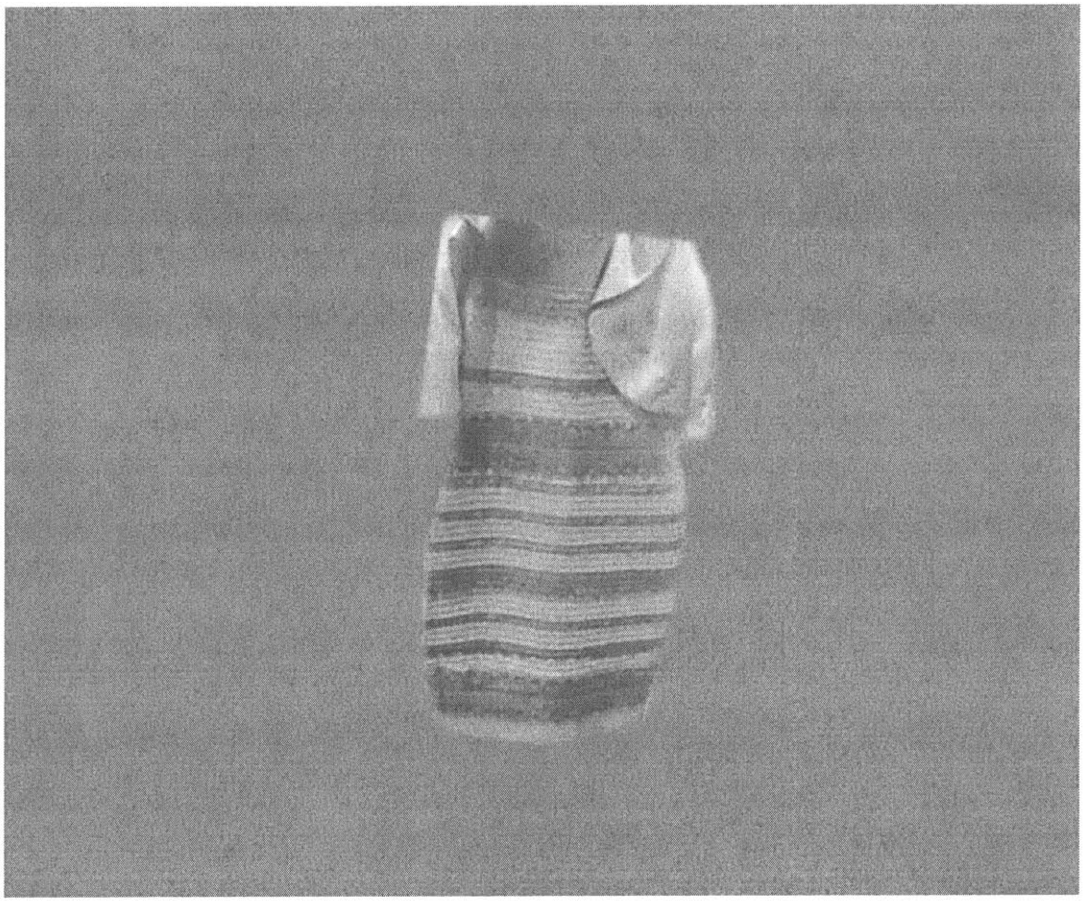
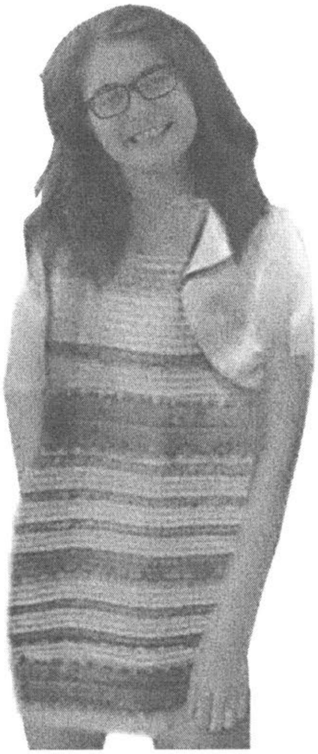


Image (6) - ‘Skin only; warm illumination’

Image (7) - ‘Skin only; cool illumination’



Chapter 4: Paradoxical Impact of memory on color appearance of faces.

This work has been published:
*M. Hassantash, ***Lafer-Sousa, R.,** Afraz, A., Conway, B. Paradoxical Impact of memory on color appearance of faces. *Nature Communications – In Press* (*co-first authors)

4.1. Abstract

What is color vision for? Here we compared the extent to which memory modulates color appearance of objects and faces. Participants matched the colors of stimuli illuminated by low-pressure sodium light, which renders scenes monochromatic. Matches for fruit were not predicted by stimulus identity. In contrast, matches for faces were predictable, but surprising: faces appeared green and looked sick. The paradoxical face-color percept could be reconciled by a model in which skin-color statistics constrain both efficient encoding and Bayesian decoding of chromatic information. The color-matching data suggest that the face-color prior is established by visual signals arising from the recently evolved L-M cone system, not the older S-cone channel. Taken together, the results show that when retinal mechanisms of color vision are impaired, the impact of memory on color perception is greatest for face color, supporting the idea that trichromatic color plays an important role in social communication.

4.2. Introduction

The benefits of color vision have been surprisingly difficult to pin down[1,2]. One popular idea is that color aids detection of ripe fruit. The main support for this hypothesis comes from an analysis of the spectral tuning of the cones and the chromatic signals of fruit[3]. But there are many surfaces besides fruit whose colors would be well discriminated by the cones, including artificial objects and faces [4]. Another idea is that color facilitates social communication about emotion, health, and sex[5,6]. Face color provides important cues to health, emotion, and attractiveness; and face context determines the meaning of the color of a face[7]. But the colors of other objects are similarly informative and determined by shape context. For example, a strawberry's color determines its nutritive (and attractive) value, and its shape provides the context for this determination.

We sought to directly test the relative role of color in object and face perception by measuring the impact of shape on color appearance under viewing conditions that cause a loss-of-function of retinal mechanisms of color. Our approach was inspired by studies that probe memory colors using digital displays in which participants adjust images to appear achromatic [8]. In those studies, a banana that appears achromatic nonetheless retains some color as assessed with a colorimeter. One hypothesis is that the residual color is required to cancel a memory-induced color attributed to the banana shape. This logic implies that objectively achromatic renderings of color-diagnostic objects should appear somewhat tinged with their typical color, an idea that remains contentious [9]. Establishing the role of shape knowledge on color perception may depend on the vividness of the shape cues. For example, the impact of memory on color perception appears to be stronger when shape cues are enhanced [10,11]. To achieve the most vivid shape experience while impairing retinal mechanisms for color, we presented real-world stimuli in a room illuminated by monochromatic low-pressure sodium (LPS) light—such lighting causes a profound failure of color constancy and only variations in lightness can be perceived [12](**Supplementary Figure 1**). Our goal was to measure under these conditions the colors people see in faces and objects, to evaluate the alternative hypotheses regarding the role of color in behavior. If memory modulates color perception, we predicted that objects with a diagnostic color such as fruit and skin should have a subtle color corresponding to the normal colors of the stimuli (arbitrarily colored objects, such as Legos, serve as a control). The results were surprising. We found no clear evidence of the impact of memory on color appearance of the fruit, and a paradoxical impact of memory on the color appearance of faces: instead of appearing in their typical color, faces appeared green.

4.3. Results

Color matches under white and low-pressure sodium light. Twenty participants matched 35 stimuli, first under LPS light, and then under white light (**Figure 1a**). The appearance of the matches (**Figure 1b**) are not necessarily an accurate representation of the color appearance of the stimuli in the experimental conditions: first, the colors will depend on the calibration of the printer or display used to show the figure; second, the colors do not account for differences in the adaptation state under the different illumination conditions. Nonetheless, the figure shows that most participants matched the stimuli under white light as expected. For example, skin samples were pinkish or brownish (depending on the race of the actor); the strawberry and tomato were red; the orange fruit was orange; and the ping-pong ball was white. These color matches are comparable to colors that a separate group of participants gave when asked to match object colors only from memory (data not shown). Under LPS light most of the stimuli were matched as yellowish (varying in lightness, from yellow-orange to brown; **Figure 1b, right panel**). Because the visual system would be adapted to the LPS light, it is likely that the stimuli did not appear as yellow to the participants as suggested by the yellowness of the matches reproduced in Figure 1 (participants reported that most stimuli appeared depleted of color, consistent with prior reports[12]). Nonetheless, the figure shows that the color matches for face stimuli were different from all other stimuli: faces were matched as green (**Figure 1b**, right panel, top 8 rows). Photographs of faces were matched with a slight green tinge, but not as green as the matches to real faces. After participants completed the matches under LPS light and before they proceeded to the tasks under white light, we asked participants to "tell us if you noticed anything about your color experience". All participants stated that their color perception was not normal. Seventy percent of participants (N=10/10 female, 4/10 male) reported that real faces looked green or otherwise sick (significantly more female than male participants, chi-square test, p=0.003, Chi-square 8.57). More female subjects (N=7/10) than male subjects (N=2) reported that faces looked sick (chi-square test, p=0.03, Chi-square 5.05). The paradoxical percept of face color under LPS light cannot be attributed to demand characteristics since the color reports do not correspond to typical face color.

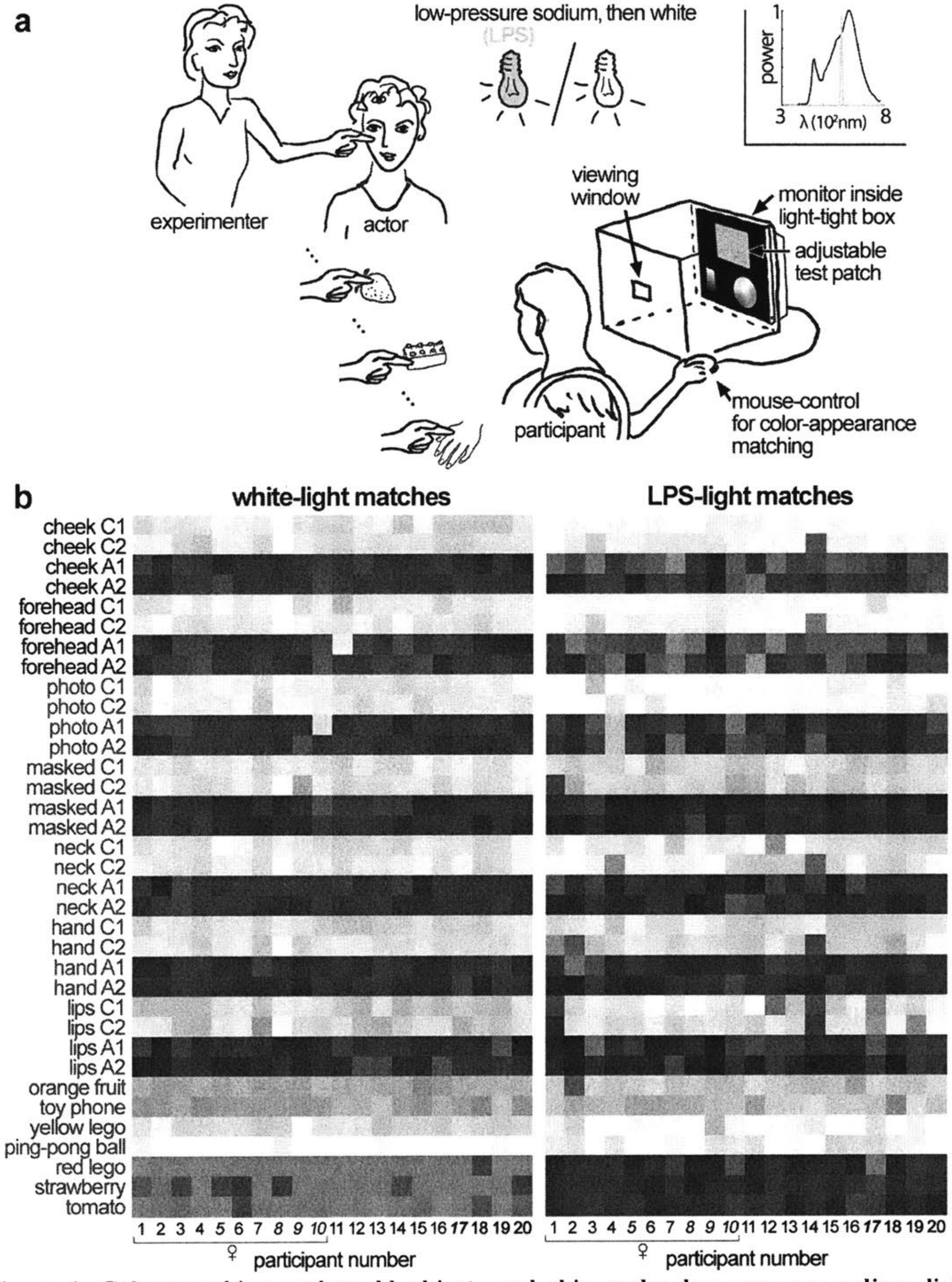


Figure 1. Color-matching real-world objects and skin under low-pressure sodium light, which impairs retinal mechanisms for color vision. A, Participants (N=20) used a computer to match the color of real-world objects (items listed in panel b) and skin (four female actors, 2 Caucasian, 2 African

American), first illuminated by low-pressure sodium light and then broad-band white light (inset shows illuminant spectra). The participant was seated so that they could view the test stimuli shown to them by the experimenter as well as a 2 x 2 cm viewing window in an otherwise light-tight box through which they could see a color-calibrated monitor (21.5-in. iMac computer, pixel resolution 1,920 × 1,080). Participants used a mouse to navigate a color-space disc and lightness strip, setting the hue, brightness, and chroma of the test patch to make the color match. Photographs and spectral measurements of objects are given in **Supplementary Figure 1. B,** Color-appearance matches made by all participants for the 35 test stimuli. The swatches are rendered using the RGB values as matched on a calibrated monitor (white point: XYZ = 87.4, 100, 57.7). Participant information: average age 27.5 (range 19-33);10 female;12 Caucasian; four Asian (italics); three African American (bold); one South-East Asian (italic bold). Source data are provided as a Source Data file.

To quantify the color matches, the RGB values for each match were converted into L* a* b* color space (**Supplementary Dataset 1**). L* a* b* color space is designed to be perceptually uniform: the a* axis shows the red-green perceptual dimension which roughly aligns with the L-M cone-opponent axis; the b* axis shows the blue-yellow dimension which roughly aligns with the S-(L+M) cone-opponent axis. Plotted in these coordinates, the angle indicates hue (red, orange, yellow, etc.), while the vector length indicates saturation (chroma). As predicted, matches under white light to all objects (filled symbols, left panel **Figure 2**) were close to the color values measured with a spectroradiometer (open symbols). Matches under the LPS light to the toys, fruit, and body skin were consistent with the spectrum of LPS light. Matches to face skin under LPS light showed a striking deviation towards negative a* values, corresponding to a greenish shift from the color of the LPS light (bottom right panel, **Figure 2**). Matched lightness values for Caucasian actors were shifted towards higher L* values compared to measured lightness values, while matched lightness values for African-American actors were shifted towards lower L*; these race-dependent shifts in lightness matches were evident under both white light (two-tailed t-test: Caucasian, $p = 10^{-28}$; African-American, $p = 10^{-16}$) and LPS light (Caucasian, $p = 10^{-25}$; African-American, $p = 10^{-3}$) (**Figure 2; Supplementary Figure 3**). Matches under LPS light to all stimuli underestimated the chroma (saturation) values measured with a spectroradiometer, consistent with subjects' visual system's having been adapted to the spectrum of LPS.

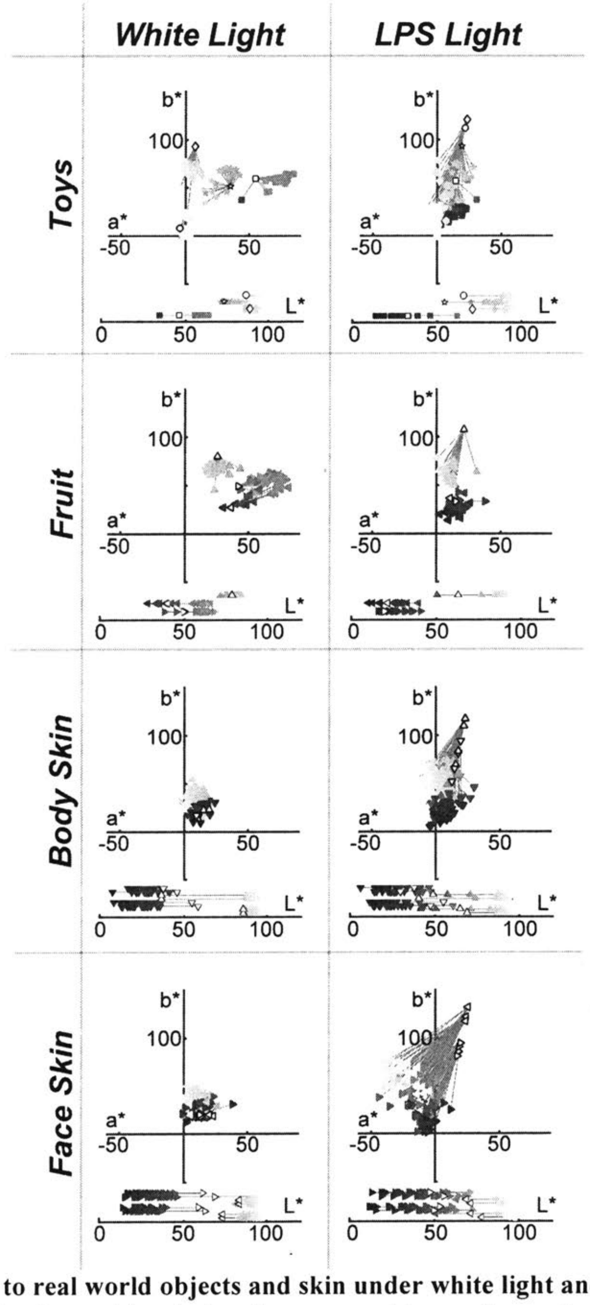


Figure 2. Color matches to real world objects and skin under white light and low-pressure sodium light. Participants' (N=20) color matches to toys (square = red Lego; diamond = yellow Lego; star = orange toy phone; circle = ping-pong ball), fruit (right-facing triangle = tomato; left-facing triangle = strawberry; upward triangle = orange), body skin (hand, neck of four actors; downward triangle = African American; upward triangle = Caucasian), and face skin (check, forehead of four actors; rightward triangle = African American; leftward triangle = Caucasian). Lines connect matched values (colored markers) to the spectroradiometric measurements (open markers). Spectroradiometric measurements were transformed to L* a* b* using the Macbeth white card under white light as the white point, CIE XYZ = [27.9446 25.0000 10.5134]). Matched values were transformed from RGB to CIE L*a*b* using the chromaticity coordinates of the monitor, gamma corrected. Source data are provided as a Source Data file.

Close inspection of Figure 2 shows that the precise hue matches were correlated with measured lightness: Under both lighting conditions, objects of lower lightness were matched subtly redder compared to measured values, while objects with higher lightness were matched subtly yellower (darker data points tend to be to the right of their corresponding spectral measurements; lighter data points tend to be to the left). This interaction of lightness and hue perception has been described previously and reflects perceptual not cognitive mechanisms [13]. It has also been shown that chroma matches can be influenced by lightness[14]. To model the impact of lightness and chroma on matched hue we ran a multivariate linear regression. The model estimated matched hue, lightness, and chroma given measured hue, lightness, and chroma, and was fit using data from objects whose color appearance is unlikely to be influenced by shape context (Legos, toy phone, masked forehead). To reconcile the mixture of circular and linear variables, we fit the model the using the a* and b* values as predictors, rather than the hue angle; the model's predicted a* and b* values were then converted to hue angle. **Figure 3** shows that the discrepancy between the matched hues and the measured hues for these non-color-diagnostic objects (**Figure 3a**) can be attributed in part to variation in the measured lightness of the stimuli (**Figure 3b**). The slope of the correlation in Figure 3b is very similar for matches under both LPS and white light; measured lightness accounted for little of the variance in *measured* hue under white light r^2=0.12). The model provides an excellent fit—the model's estimates are correlated with subjects' matches (**Figure 3c**) and the residual distributions are centered on zero and do not systematically vary with estimated hue (**Figure 3d**). These results show that the model does a good job of accounting for the impact of measured lightness and chroma on hue matches.

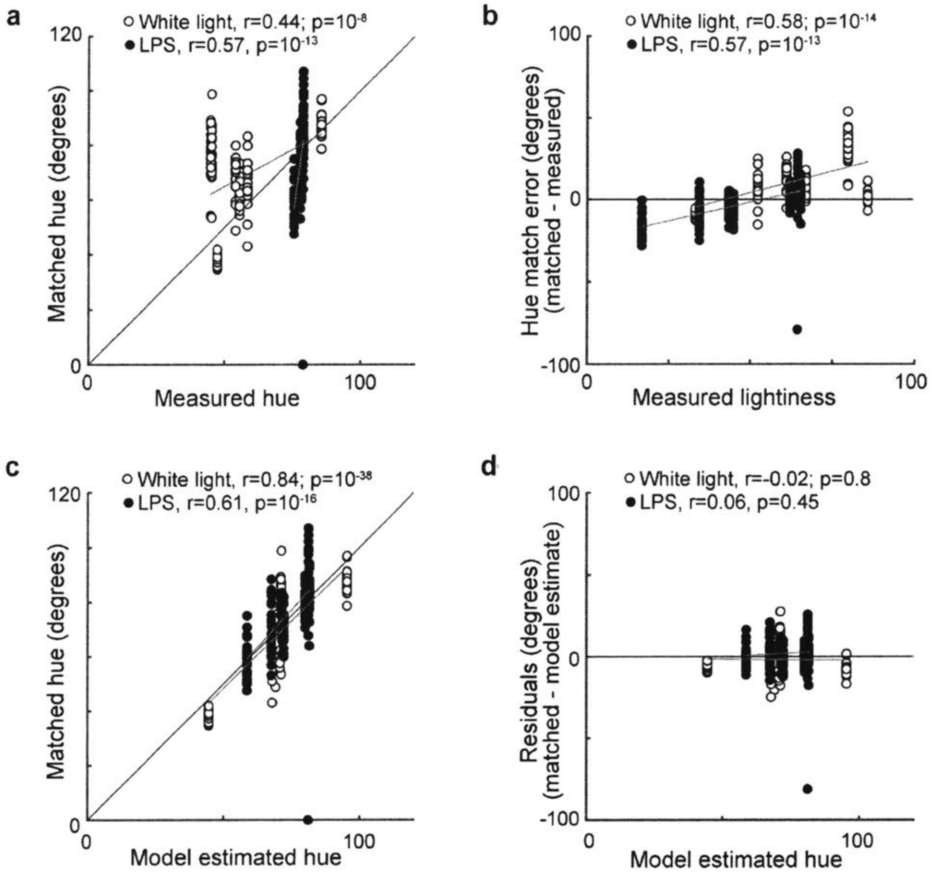


Figure 3. Estimating the impact of lightness and chroma on hue matching using matches to non-color-diagnostic objects. a, Participants' (N=20) hue matches versus the measured hues for non-color-diagnostic stimuli (toys, masked forehead; CIE a*b* angle). Markers show results of matches made under white light (open symbols) and LPS light (solid symbols). **b,** The difference between the angle of the matched hue and the measured hue versus measured lightness (white light, r = 0.58; LPS, r = 0.57). **c,** Matched hue versus the estimated hue from the multivariate linear regression model. The model fit matched hue as predicted by measured hue, lightness, and chroma, for the non-color-diagnostic objects (white light, r = 0.84; LPS, r = 0.61). **d,** Residuals of the model fit (matched – estimated) as a function of estimated hue (white light, r=-0.02; LPS, r=0.06). Source data are provided as a Source Data file.

Residuals obtained by applying the model (trained on non-color-diagnostic objects) to the data obtained for color-diagnostic stimuli (fruit, face skin, body skin) would provide evidence of the impact of object-color knowledge on color appearance. Residuals for the matches made under white light for body and face skin are centered near zero (average [99% CI] for face skin: 5.3 [2.7, 7.8]; body skin: 8.8 [5.9, 11.2]), **Figure 4a**). Residuals of the matches made under LPS light for body skin are also centered near zero (12.6 [10.1 15.4], but residuals of the matches under LPS light for face skin are shifted away from zero (39.8 [36.6, 43.0]). The difference between the residuals for matches to face and body under LPS light are different from the corresponding measurements under white light (paired t-test, $p=10^{-37}$; **Figure 4a**). Unlike the residuals of the matches to face skin and body skin under white light, the residuals for matches to lips under white light had high variance (average [99% CI] for Caucasian lips: 38.15 [32.2, 43.5]; African American lips: -24.6 [-52.3, -0.6]; F-test for equal variances of lip residuals compared to body and face skin residuals, $p = 10^{-78}$). Residuals for matches under LPS light are higher for face stimuli, regardless of race, than for fruit (**Figure 4b**). Residuals for matches to fruit and body skin had a relatively small magnitude but were nonetheless different from zero, which suggests that color matches to these stimuli are modulated by memory. A subtle impact of memory on color appearance is consistent with prior work [8,10,11]. But we cannot rule out alternative explanations, namely that the small-magnitude residuals for fruit and body skin indicate that the model is imperfect and/or color matches to these stimuli reflect a subtle demand characteristic. The analysis of the residuals showed subtle differences in the matches to Caucasian versus African-American skin (**Figure 4b**), which probably reflects the systematic impact of race on lightness matches (**Figure 2, Supplementary Figure 3**). A 3-way ANOVA of lighting condition (white light versus LPS light), skin type (face skin versus body skin), and race (African-American versus Caucasian) uncovered a main effect of lighting ($p=10^{-62}$); a main effect of skin type ($p=10^{-28}$); no main effect of race ($p=0.5$); an interaction of lighting and skin type ($p=10^{-43}$); and a three-way interaction of lighting, skin type, and race ($p = 10^{-9}$). Despite the sex differences in the reports made by participants, there was no sex difference in the residuals for the color matches to faces (two-tailed t-test: LPS, $p = 0.16$; white light, $p = 0.95$)—this shows that male and female participants were equally likely to see faces as green under LPS light but were not equally likely to tell us about it.

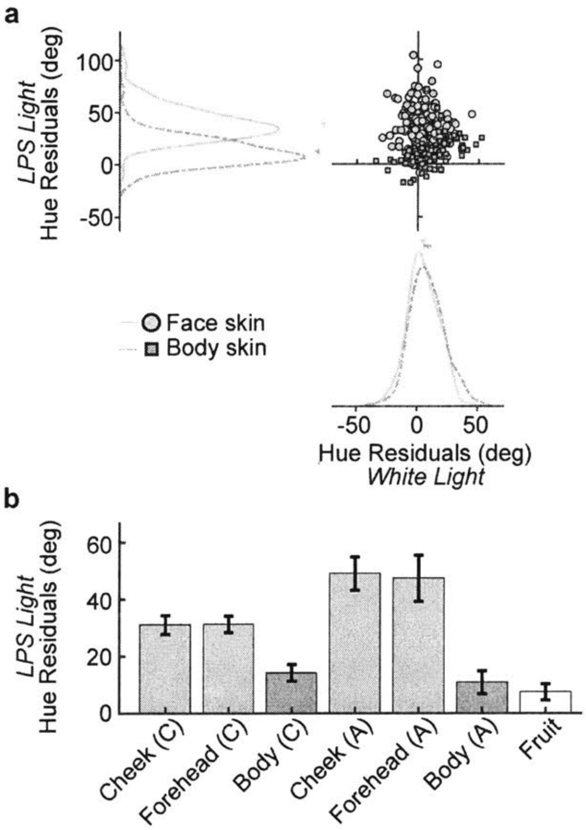


Figure 4. The impact of stimulus category on hue matches under low-pressure sodium light is greater for face skin than body skin or fruit. Hue matches to face skin (cheek, forehead), body skin (neck, hand), and fruit (tomato, strawberry, orange) were estimated from the multivariate linear regression model (see Figure 3; (N=20 participants). **a,** Hue angle not explained by the model (residuals) for skin matches under LPS light versus white light (160 face-skin matches, 160 body-skin matches: 2 face-skin regions; 2 body-skin regions; 4 actors; 20 participants). Markers show averages and 99% C.I. **b,** Residuals for hue matches under LPS light for fruit, and for face skin (cheek, forehead) and body skin (neck and hand, combined) of the two races tested (Caucasian, C; African American, A). Error bars show 99% C.I. Source data are provided as a Source Data file.

The analysis of the residuals shown in **Figure 4a** suggests that the paradoxical color appearance of faces is determined by face context not stimulus material. But body skin can be slightly different from face skin; for example, body skin might have different texture or vascularization. To rule out these possible material differences as an explanation for the paradoxical color matches of faces, we compared the hue matches made to the identical stimulus—the forehead region—with and without face context (**Figure 5**). Matches were made to a patch of forehead with the rest of the face masked and to the same region without the mask. We found no impact of face context on hue matches under white light (**Figure 5a**, top), but a large effect of face context on hue matches under LPS light (**Figure 5a**, bottom). This result was clear for both races tested (**Figure 5b**) and shows that the paradoxical color appearance is caused by face context. Because the paradoxical color matches for faces were only partially evident in photographs of faces (**Figure 1**), we quantified with an ANOVA the impact of face (cheek, unmasked forehead, photo) versus non-face (hand, neck, masked forehead) and three dimensionality (cheek, unmasked forehead, neck, hand) versus two dimensionality (photo and masked forehead) on the hue matches under LPS light. The results showed a main effect of face ($p=10^{-68}$), a main effect of 3-D cues ($p=10^{-23}$), and an interaction of the two factors ($p=10^{-11}$). These results show that the paradoxical color matches for faces are dependent on face-shape context enhanced by having richer shape cues.

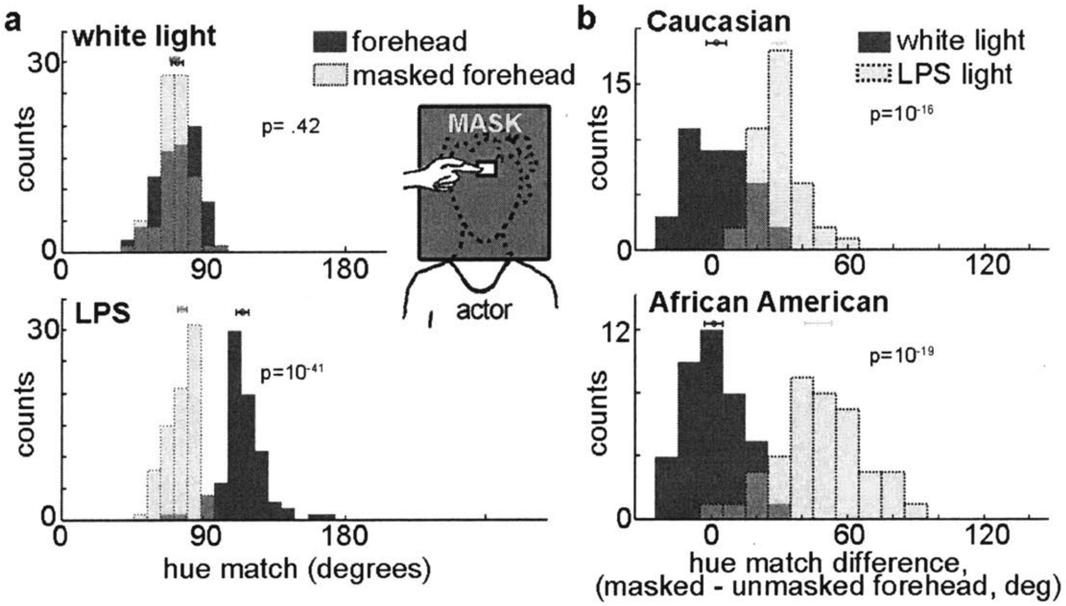


Figure 5. Paradoxical color matches for face skin depends on face-shape information and is not race-dependent. a, Histograms showing the angle of the color matches for forehead and masked forehead under white light (top) and low-pressure sodium light (bottom) (N=20 participants; matches made to four actors: 2 Caucasian, 2 African American). Color matches in the masked condition were made by holding a sheet of black cardboard in front of the actor's face, with a 1.5 cm x 2cm aperture over the forehead. Masking the face context did not change the hue match under white light ($p=0.42$) but had a large impact on the color match under LPS light ($p=10^{-41}$; p values obtained by t-test not assuming equal variance; markers show average values, and error bars are 99% C.I.). **b,** Histograms showing the difference in hue angle for color matches to forehead versus masked forehead, under white light versus LPS light, for Caucasian skin (top) and African American skin (bottom). The two distributions for each race do not have the same mean (Caucasian $p=10^{-16}$; African American $p=10^{-19}$; markers and error bars as for panel a). Source data are provided as a Source Data file.

In three participants we measured color matches to real faces mirror reversed and upside down, and to a doll's face and hand. The main results were replicated in these participants (**Figure 6a**). Moreover, the paradoxical color matches for faces were evident in real faces mirror reversed and upside down (**Figure 6b**); weakly in photographs but not in scrambled photographs (**Figure 6c**); and in a doll's face but not in a doll's hand (**Figure 6d**). These results confirm that the green-face effect is not dependent on a specific reflective or texture property of skin and is enhanced by having more vivid shape cues. The color matches made under LPS light provide information about the signals that the brain uses to form knowledge about skin color.

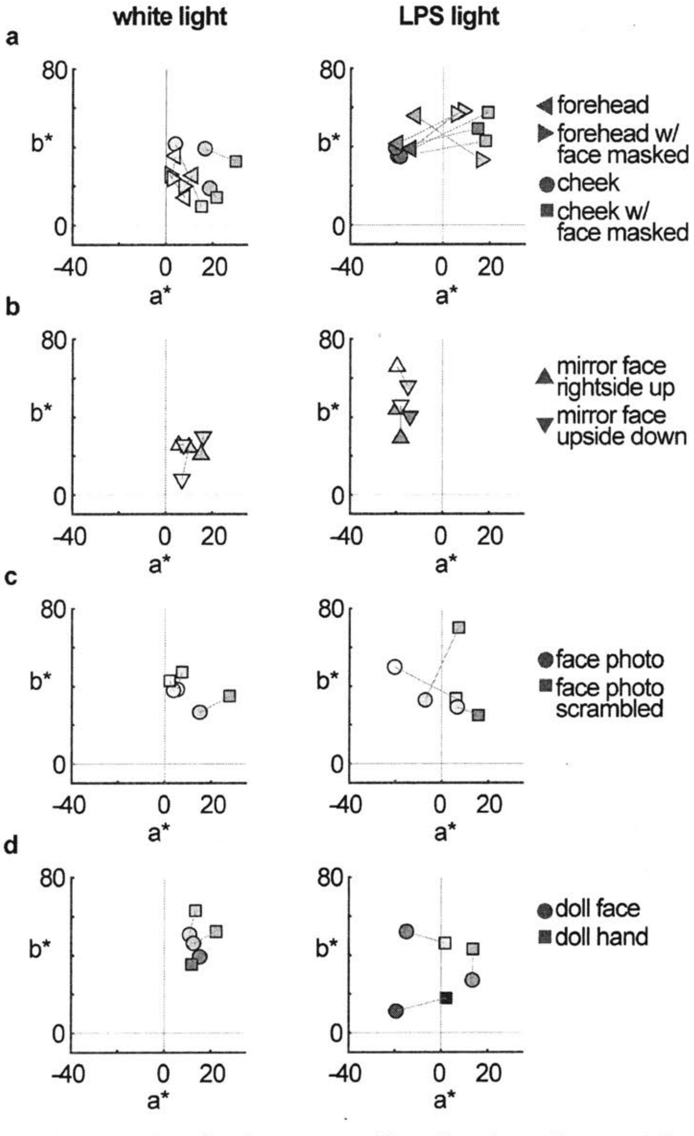


Figure 6. Paradoxical color matches for faces are evident in mirror inverted faces, photos, and a doll face. Color matches under white light and low-pressure sodium light; corresponding stimuli are connected by a line (see key). Data were collected in three participants (#5, 7, 19, in Figure 1). **a,** Forehead and masked-forehead, and cheek and masked cheek. Matches to the masked condition are to the right of the matches to the unmasked condition. **b,** Faces seen upright and upside-down in a mirror. For the upside-down condition, matches were made by looking down at a mirror view of the face. **c,** Photograph of a Caucasian female face and scrambled photograph of the face. **d,** Doll face and doll hand; the doll hand did not show a paradoxical color match for any participant. Symbol color corresponds to the color of the match. Source data are provided as a Source Data file.

Color matches to face skin versus non-face skin were indistinguishable under white light (**Figure 7a**), but they were distinguishable under LPS light by the extent to which they modulated the L-M color axis (**Figure 7b,c**). This suggests that the memory color of faces is encoded by signals that modulate a differential L-cone versus M-cone signal. The L-M system, which defines trichromacy and arose relatively recently in primate evolution (~23 million years ago)[15], relays information about health, sex, emotion, and attractiveness[16,17]—such information is dynamic and independent of face identity [18]. The paradoxical color matches under LPS light imply that perception of face color is weighted towards dynamic features—the emotion or health of a face—rather than stable properties such as identity. Taken together, the results are consistent with the idea that selective pressures related to social cognition fueled the evolution of trichromacy in our primate ancestors.

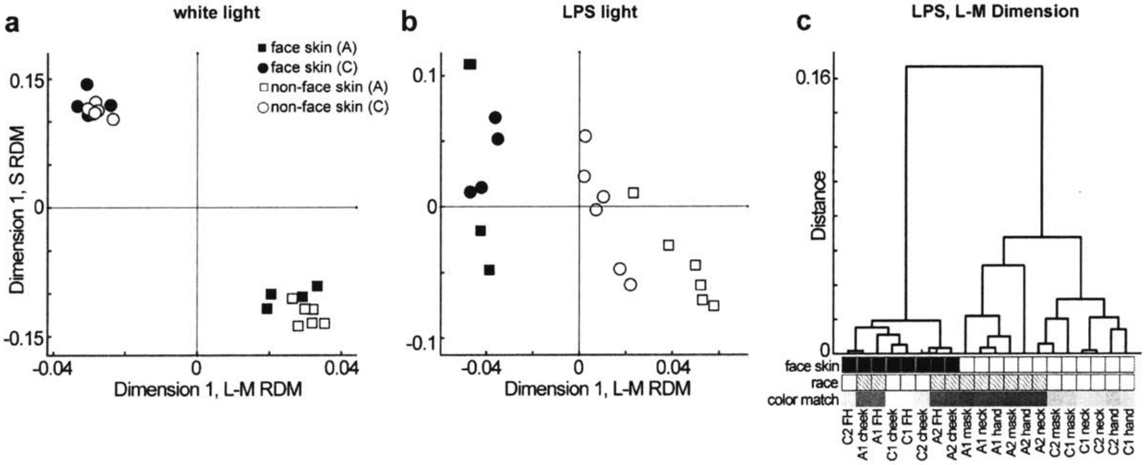


Figure 7. Paradoxical color matches for faces is encoded by the more recently evolved (L-M) system. a, Color matches for skin were evaluated for the extent to which they modulated the L-M versus S cone systems; these values were then subjected to a hierarchical clustering algorithm to determine the relative similarity among matches across the various skin samples; data from 20 participants. The plots show multidimensional scaling (MDS) relating the first dimension of the representational dissimilarity matrix (RDM) for the S component against the first dimension of the RDM for the L-M component, for matches under white light. Matches were made to face skin: cheek and forehead; and non-face skin: masked forehead, neck, hand. Face skin and non-face skin are not separable with either the L-M or the S dimensions. **b,** As for panel a, but for matches made under low-pressure sodium light. Face skin and non-face skin are separable with the L-M dimension but not the S dimension; the L-M dimension also separates body skin by race. **c,** Dendrogram showing the relative dissimilarity distances of the L-M component of the matches under LPS light. The main branch separates face skin from non-face skin; a secondary branch separates race among matches to body skin, but not face skin. Source data are provided as a Source Data file.

4.4. Discussion

The experiments described here probe the impact of memory on color perception and uncover a special role of color in face perception. Consistent with the observation that scenes under low-pressure sodium (LPS) light impair retinal mechanisms for color [12], color matches under LPS light for arbitrarily colored objects were not predicted by colors seen under white light. Color matches for fruit under LPS light were also not predicted by stimulus identity: for example, knowledge that a strawberry is red did not cause participants to match the LPS-illuminated strawberry as red. These results show that cognition does not always hold sway over color appearance. But unexpectedly, color matches under LPS light to one class of stimuli, face skin, were predictable, although surprising: all participants matched faces green. Furthermore, most participants (female>male) reported that faces looked green or appeared sick, showing that the modulation by memory of face color does not remain unconscious. This paradoxical percept was evident for faces of both races tested, was abolished when the face context was masked, and was not observed for matches made to body skin. The results lead to three conclusions. First, the brain has a strong prior specifically for the color of skin, which triggers a prediction-error signal, possibly diagnostic of sickness, when violated in the context of rich face-shape information. Second, trichromatic color plays an especially important role in social communication. Third, cognition can influence perception, refuting notions to the contrary [9]. We found that memory not only modulated perception of face color, but also impacted the lightness matches made to skin reflecting knowledge about race [19](**Supplementary Figure 3**).

Why do faces under LPS light look green? LPS light leaves intact rich shape cues, making it inescapable that the face is real. In the context of a real face, the peculiar spectral signals cannot be discounted with a trivial explanation, unlike the color in a photograph or digital reproduction, which can be attributed to the way the image is generated. Under LPS light, the spectral signals from skin are characterized by a decrement in redness (**Figure 8**); a similar decrement accompanies many illnesses [20,21], caused by sympathetic vasoconstriction of superficial blood vessels or anemia. We suspect that participants attribute the peculiar chromatic signal to sickness as the most likely explanation, which would explain why many participants described LPS-illuminated faces as sick-looking. The chromatic signals of LPS-illuminated skin violate a prior about healthy skin, breaching a naturalness constraint [22]. But why should this breach cause a green appearance?

The decrement in redness for either LPS-illuminated faces or sick faces does not yield negative a* values—colormetrically the signal is still reddish. The green appearance of faces under LPS light is paradoxical: it is an exaggeration of the greenward shift of the received signal compared to the prior. A standard Bayesian account argues that perception is biased towards the prior, not away from it. Yet circumstances in which percepts are biased away from a prior have been described in other domains, for example orientation, biological motion, and size/weight [23-25]. Although these phenomena were originally thought to reflect anti-Bayesian processes, they can be accounted for by a Bayesian observer model constrained by efficient coding, in which the statistics of the natural environment shape both the encoding and decoding of sensory information [26]. In this framework, there is a nonlinear mapping between stimulus and sensory representation that results in repulsive perceptual estimates for stimuli that sufficiently violate sensory expectations. One can think of this repulsion as a form of contrast-enhancement that promotes behaviorally important categorization [27]. According to this interpretation, the exaggeration of prediction errors provides evidence of a computational objective of the visual system: here, the vital distinction between health and sickness. The present results suggest the visual system accomplishes this objective by encoding the statistics of healthy skin. The relative importance of skin color to human behavior is well known: humans have precise color preferences for skin, do not tolerate poor color reproduction of skin [28-31], are especially sensitive to changes in facial redness [17], and can use skin color towards color constancy[32,33]. The results here predict that any circumstance in which the chromatic signals arising from the face deviate from the face-color prior (and cannot be explained away by any more likely explanation) should appear peculiar. Consistent with this prediction, sick faces often appear green [34], evident in emojis (**Figure 8**). Moreover, repulsive biases should scale with prediction error. Faces under LPS light incur larger red decrements than do most sick faces, which may explain why faces under LPS light are matched green by almost all observers.

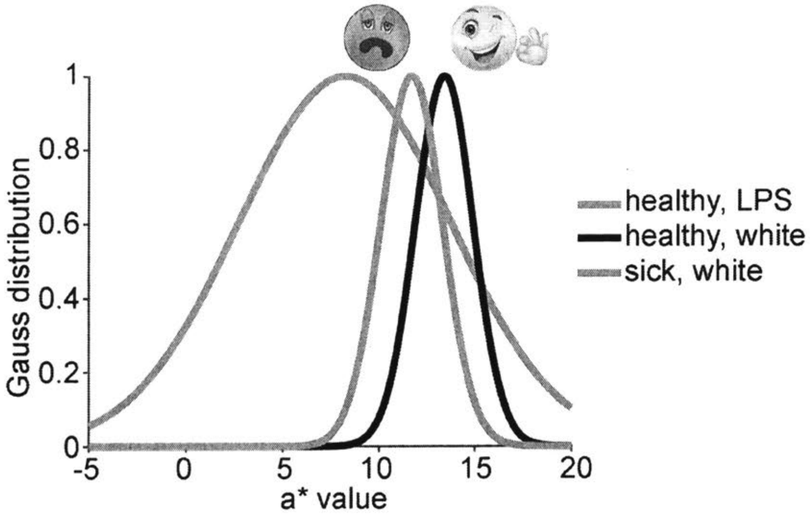


Figure 8. Probability of a* values given health or sickness and lighting condition. The probability distribution for healthy skin under sodium light was generated using participant matches of masked forehead. Gaussian distributions for healthy and sick faces were obtained from the mean and standard deviation of measurements of cheek and forehead provided by Henderson et al (2017). Source data are provided as a Source Data file. Smiley emoticon with OK sign created by Tigatelu with Dreamstime.com. All rights reserved. Sick Emoji created by BRC.

It is widely thought that color and face perception are handled by separate neural circuits[35]. The fact that the paradoxical color percepts reported here depend on face shape implies that color and face shape are processed by some common neural substrate somewhere in the brain. The systematic relationship of face-biased and color-biased regions in the cerebral cortex, and their convergence at the anterior pole of inferior-temporal cortex[18,36], may provide clues to this neural substrate. In addition, the results and interpretation described here predict the existence of neural operations dedicated to encoding skin color, and in particular, neural tuning/populations biased toward skin color priors. This prediction is supported by evidence of a broad bias in the ventral visual pathway towards warm (L>M) colors [4], and preliminary work showing an L>M color-bias specifically in face-selective neurons [37].

Regardless of the mechanism, the results show for the first time that the brain assigns special weight to the color signals from faces compared to color signals from other objects including ripe fruit. This is an important finding because it has been difficult to disentangle competing accounts of the evolutionary pressures that selected for trichromatic color vision. The spectral tuning of L and M cones allows the discrimination of ripe fruit and nutritious leaves[3] and objects from backgrounds more generally[4], but is equally well-suited to encoding color signals associated with health, sex, and emotion[16,38], especially in the face[7]. The results presented here do not refute a role of color in detecting, discriminating, recognizing, and remembering objects, scenes, or fruit [2,6,39], but they help resolve the relative value of color in behavior by showing that trichromatic color signals from the face are especially important for behavior.

4.5. Methods

Color matching procedure

Twenty people with normal color vision (tested with Ishihara plates) participated in the experiment. Each participant matched the color appearance of seven real-world objects including four artificially colored toys (orange phone, red Lego, yellow Lego, white ping-pong ball), three ripe fruits (strawberry, tomato, orange), and various skin regions of four actors (all female; 2 Caucasian, 2 African American; no makeup) **(Figure 1a)**. Test objects were chosen to have the same palette as skin. The Legos and toy phone were included because they should not give rise to shape-dependent color percepts since their colors are arbitrary. All experimental procedures were approved by the Wellesley College Institutional Review Boards, the Massachusetts Institute of Technology Committee on the Use of Humans as Experimental Subjects, and the National Institutes of Health Intramural Institute Clinical Research Review Committee. All participants and actors provided written informed consent and were compensated financially for their involvement. The two actors whose photographs are shown in Supplementary Figure 1 provided written consent to publish their photographs.

Color appearance matches were made using a calibrated computer monitor encased in a light-tight box, inside the testing room. Participants could see the monitor by lifting a small black cloth to reveal a 2cm^2 viewing window. The monitor showed a color-space disc and a lightness bar that the participant could use to adjust a central patch to make their selection with a computer mouse. The monitor was otherwise black, and the color of the test patch was random at the onset of testing. Participants were instructed to match the color of the patch as accurately as possible to the color appearance of the test stimuli, and not to match the color of the stimuli as recalled from memory. We were mindful of the potential for demand characteristics to influence the results: participants might set the test patch to match the typical colors of the stimuli even if they perceived the stimuli to be achromatic. Our initial aim was to quantitatively compare matches for fruit, face skin, and body skin, with the assumption that demand characteristics would equally impact judgments of all stimuli. As described below, the results provide evidence against a role of demand characteristics under our experimental conditions.

The participants did the matching twice: first while the testing room was illuminated with LPS light, and then again after the room light was switched to normal white light (participants adapted to the illumination for ~7 minutes). Participants were tested first under LPS light to prevent short-term recall of color matches made under white light. Test stimuli were presented in a unique order for each participant (**Supplementary Dataset 2**) at ~1m viewing distance. The precise region to be matched was indicated with a lightly drawn circle (for the objects) or by pointing to regions on the actors (forehead, lips, neck, and back of the hand). Participants also color-matched the forehead of each actor in a photograph and the actors' foreheads while masking the rest of the face with black paper (~1 cm diameter aperture). All participants matched the same region of each stimulus, and the lighting on the stimuli was consistent across participants: stimuli were placed at a set location on a table, at a fixed distance from the light sources; actors were seated with their faces at a fixed orientation and gaze angle with respect to the participant. The actors entered the testing room one at a time. The participant's adaptation state was preserved during transitions by having the participant close their eyes and by keeping the anteroom dark. Participants performed the color-matching tasks reliably and consistently; **Supplementary Figure 2** shows test-retest reliability measures for three subjects, tested several months apart. An initial pilot experiment involving separate participants and actors yielded similar conclusions and provided the basis for the experimental approach reported here.

4.6. Data Analysis

The visual system implements color constancy operations that correct for the spectral bias in the illuminant. Color constancy can be almost perfect[40], but fails when the illuminant is monochromatic, as under LPS light[12]. The consequence is that color cards viewed under LPS light have an eerie quality: they are tinged with the color of the light, but as Boynton et al describe (1989) only variations in lightness are perceived. The perception of brightness, depth, perspective, shape, shading, and motion remain intact. The luminance distribution across objects under LPS light is comparable to that under white light, and there is no systematic difference in the luminance distribution under these two illuminants for faces compared to other objects we tested (**Supplementary Figure 1**). Adaptation algorithms that predict color appearance given the illuminant are imperfect and get worse the further an illuminant is from neutral [41]. Because of these constraints, color-correction algorithms are not able to estimate color appearance under the LPS light. Spectral measurements of objects under both white light and LPS light were transformed from CIE XYZ to CIE L*a*b* using spectral measurements of the Macbeth white card under white light as the white point (Supplementary Figure 1). In this report we present two analyses of subjects' color matches. First, we analyzed the raw color matches obtained on the calibrated matching monitor (see **Figure 1**), transformed from RGB to CIE L*a*b* using the measured chromaticity coordinates and luminance curves of the monitor's R, G, B channels (see **Figures 2, 5, 6,** 7). The chromaticity values and luminance curves of the matching monitor were obtained with a spectroradiometer (PR655, Chatsworth, CA) and the monitor was gamma-corrected. Second, we estimated errors in hue matches attributed to lightness and chroma, and analyzed the residuals (see **Figures 3, 4**). These residuals provide an estimate of the impact of color-shape knowledge (priors) on color perception. We also did an analysis in which we empirically estimated the adaptation state for each participant under each experimental condition by having participants match a white ping-pong ball. This approach allowed us to control for variability among our participants in how each person's visual system adjusted to the LPS light; the main results are the same using this approach. But because the white points are slightly different for each participant, and the impact on the gamut is not trivial to compute, we only present the results and analysis using the color space in which the matches were obtained. Throughout the figures, confidence intervals were generated by 1000 bootstraps.

4.7. Data Availability

All data and analysis code are available on a public-accessible website (https://neicommons.nei.nih.gov/#/facecolor).

4.8. References

1. Biederman, I. & Ju, G. Surface versus edge-based determinants of visual recognition. *Cogn Psychol* 20, 38-64, doi:0010-0285(88)90024-2 [pii] (1988).

2. Gegenfurtner, K. R. & Rieger, J. Sensory and cognitive contributions of color to the recognition of natural scenes. *Curr Biol* 10, 805-808 (2000).

3. Regan, B. C. *et al.* Fruits, foliage and the evolution of primate colour vision. *Philosophical Transactions of the Royal Society of London - Series B: Biological Sciences* 356, 229-283 (2001).

4. Rosenthal, I. *et al.* Color statistics of objects, and color tuning of object cortex in macaque monkey. *J Vis* 18, 1, doi:10.1167/18.11.1 (2018).

5. Changizi, M. A., Zhang, Q. & Shimojo, S. Bare skin, blood and the evolution of primate colour vision. *Biol Lett* 2, 217-221 (2006).

6. Conway, B. R. Color vision, cones, and color-coding in the cortex. *Neuroscientist* 15, 274-290 (2009).

7. Lefevre, C. E., Ewbank, M. P., Calder, A. J., von dem Hagen, E. & Perrett, D. I. It is all in the face: carotenoid skin coloration loses attractiveness outside the face. *Biol Lett* 9, 20130633 (2013).

8. Hansen, T., Olkkonen, M., Walter, S. & Gegenfurtner, K. R. Memory modulates color appearance. *Nat Neurosci* 9, 1367-1368 (2006).

9. Firestone, C. & Scholl, B. J. Cognition does not affect perception: Evaluating the evidence for "top-down" effects. *Behav Brain Sci* 39, e229, doi:10.1017/S0140525X15000965 (2016).

10. Olkkonen, M., Hansen, T. & Gegenfurtner, K. R. Color appearance of familiar objects: effects of object shape, texture, and illumination changes. *J Vis* 8, 13 11-16 (2008).

11. Vurro, M., Ling, Y. & Hurlbert, A. C. Memory color of natural familiar objects: effects of surface texture and 3-D shape. *J Vis* 13, 20 (2013).

12. Boynton, R. M. & Purl, K. F. Categorical colour perception under low-pressure sodium lighting with small amounts of added incandescent illumination. *Lighting Res. Technology* 21, 23-27 (1989).

13. Buck, S. L. & DeLawyer, T. Dark versus bright equilibrium hues: rod and cone biases. *J Opt Soc Am A Opt Image Sci Vis* 31, A75-81, doi:10.1364/JOSAA.31.000A75 (2014).

14. Giesel, M. & Gegenfurtner, K. R. Color appearance of real objects varying in material, hue, and shape. *J Vis* 10, 10, doi:10.1167/10.9.10 (2010).

15. Yokoyama, S. & Yokoyama, R. Molecular Evolution of Human Visual Pigment Genes. *Mol Biol Evol* 6, 186-197 (1989).

16. Stephen, I. D., Law Smith, M. J., Stirrat, M. R. & Perrett, D. I. Facial Skin Coloration Affects Perceived Health of Human Faces. *Int J Primatol* 30, 845-857, doi:10.1007/s10764-009-9380-z (2009).

17. Tan, K. W. & Stephen, I. D. Colour detection thresholds in faces and colour patches. *Perception* 42, 733-741, doi:10.1068/p7499 (2013).

18. Conway, B. R. The Organization and Operation of Inferior Temporal Cortex. *Annu Rev Vis Sci* 4, 381-402, doi:10.1146/annurev-vision-091517-034202 (2018).

19. Levin, D. T. & Banaji, M. R. Distortions in the perceived lightness of faces: the role of race categories. *J Exp Psychol Gen* 135, 501-512, doi:10.1037/0096-3445.135.4.501 (2006).

20. Holmes, S. R., King, S., Stott, J. R. R. & Clemes, S. Facial skin pallor increases during motion sickness. *J Psychophysiol* 16, 150-157, doi:10.1027//0269-8803.16.3.150 (2002).

21. Henderson, A. J. *et al.* Skin colour changes during experimentally-induced sickness. *Brain Behav Immun* 60, 312-318, doi:10.1016/j.bbi.2016.11.008 (2017).

22. Yendrikhovskij, S. N., Blommaert, F. J. J. & Ridder, H. Color Reproduction and the Naturalness Constraint. *Color Res. Appl* 24, 54-67 (1999).

23. Campbell, F. W. & Maffei, L. The tilt after-effect: a fresh look. *Vision Res* 11, 833-840 (1971).

24. Sweeny, T. D., Haroz, S. & Whitney, D. Reference repulsion in the categorical perception of biological motion. *Vision Res* 64, 26-34, doi:10.1016/j.visres.2012.05.008 (2012).

25. Brayanov, J. B. & Smith, M. A. Bayesian and "anti-Bayesian" biases in sensory integration for action and perception in the size-weight illusion. *J Neurophysiol* 103, 1518-1531, doi:10.1152/jn.00814.2009 (2010).

26. Wei, X. X. & Stocker, A. A. A Bayesian observer model constrained by efficient coding can explain 'anti-Bayesian' percepts. *Nat Neurosci* 18, 1509-1517, doi:10.1038/nn.4105 (2015).

27. Barlow, H. The exploitation of regularities in the environment by the brain. *Behav Brain Sci* 24, 602-607; discussion 652-671 (2001).

28. Sanders, C. L. Color preferences for natural objects. *Illuminating Engineering* 54, 452-456 (1959).

29. Bartleson, C. J. & Bray, C. P. On the preferred Reproduction of Flesh, Blue-Sky, and Green-Grass Colors. *Photographic Science and Engineering* 6, 19-25 (1962).

30. Hunt, R. W. G., Pitt, I. T. & Winter, L. M. The Preferred Reproduction of Blue Sky, Green Grass and Caucasian Skin in Colour Photography. *J. of Photo. Sci.* 22, 144-149 (1974).

31. Zeng, H. Z. & Luo, R. Colour and tolerance of preferred skin colours on digital photographic images. *Color Res Appl* 38, 30-45, doi:10.1002/col.20696 (2013).

32. Crichton, S., Pichat, J., Mackiewicz, M., Tian, G. & Hurlbert, A. C. Skin chromaticity gamuts for illumination recovery. In *6th European Conf. on Colour in Graph., Imagin, and Vision, CGIV 2012, Amsterdam, the Netherlands, May 6-12, 2012,* 266-271(2012).

33. Lafer-Sousa, R. & Conway, B. R. #TheDress: Categorical perception of an ambiguous color image. *J Vis* 17, 25, doi:10.1167/17.12.25 (2017).

34. Loudon, I. S. Chlorosis, anaemia, and anorexia nervosa. *Br Med J* 281, 1669-1675 (1980).

35. Webster, M. A. & MacLeod, D. I. Visual adaptation and face perception. *Philos Trans R Soc Lond B Biol Sci* 366, 1702-1725, doi:10.1098/rstb.2010.0360 (2011).

36. Lafer-Sousa, R. & Conway, B. R. Parallel, multi-stage processing of colors, faces and shapes in macaque inferior temporal cortex. *Nat Neurosci* 16, 1870-1878 (2013).

37. Rosenthal, I. *et al.* Color responses of face cells in alert macaque monkey. *Society for Neurosciences Abstracts* 307.05/II1 (2018).

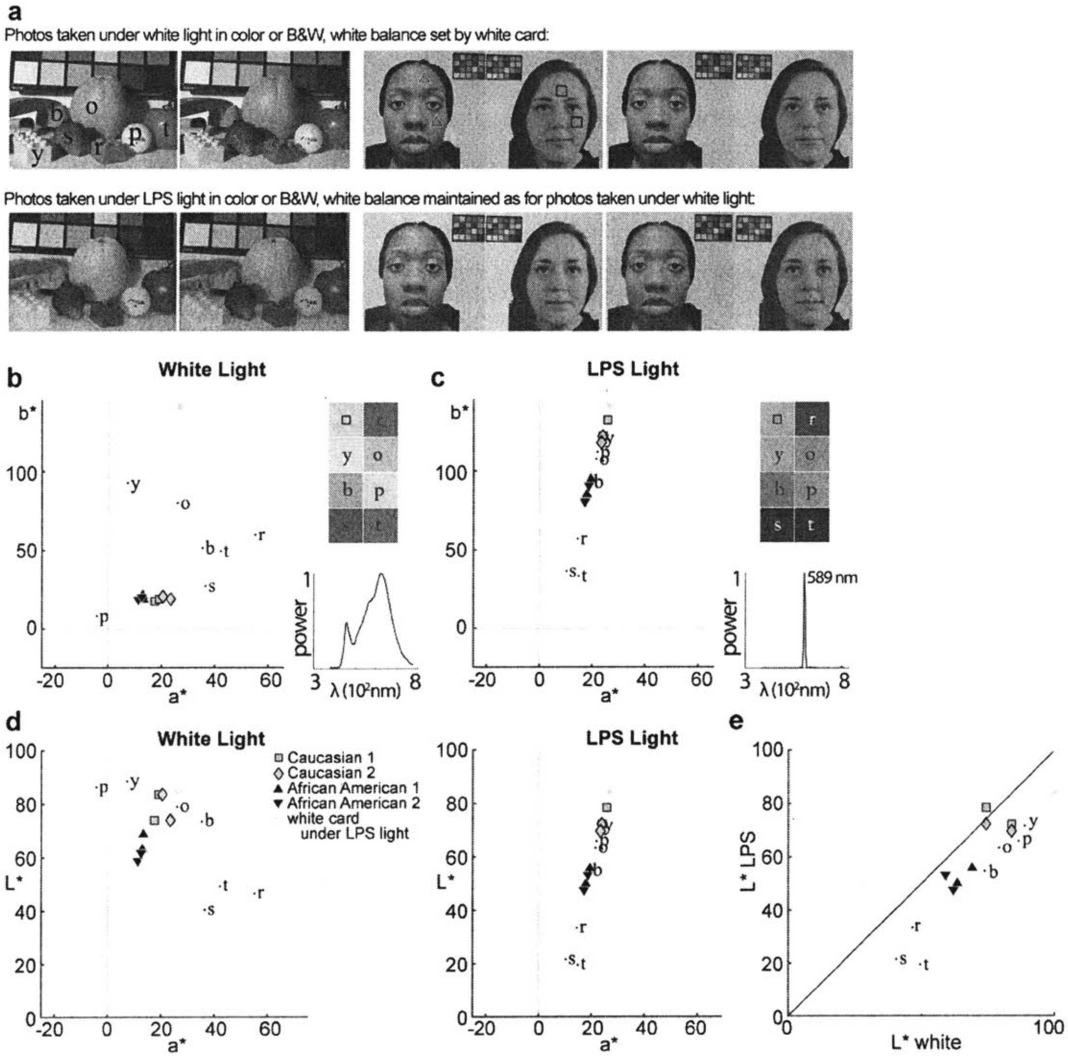
38. Hiramatsu, C., Melin, A. D., Allen, W. L., Dubuc, C. & Higham, J. P. Experimental evidence that primate trichromacy is well suited for detecting primate social colour signals. *Proc Biol Sci* 284, doi:10.1098/rspb.2016.2458 (2017).

39. Tanaka, J., Weiskopf, D. & Williams, P. The role of color in high-level vision. *Trends Cogn Sci* 5, 211-215 (2001).

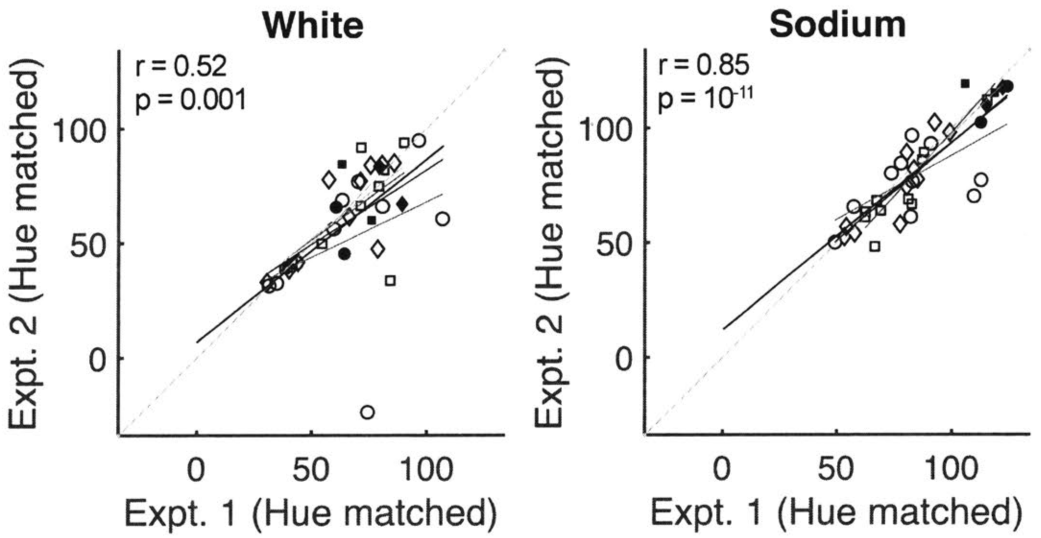
40. Brainard, D. H. & Radonjić, A. Color constancy. *in The New Visual Neurosciences (Werner and M.C. Chalupa, Eds.)* Cambridge, MA: MIT Press, 545-556 (2014).

41. Wyszecki, G. & Stiles, W. S. Color Science: concepts and methods, quantitative data and formulae. *John Wiley and Sons, Inc., New York* (1982).

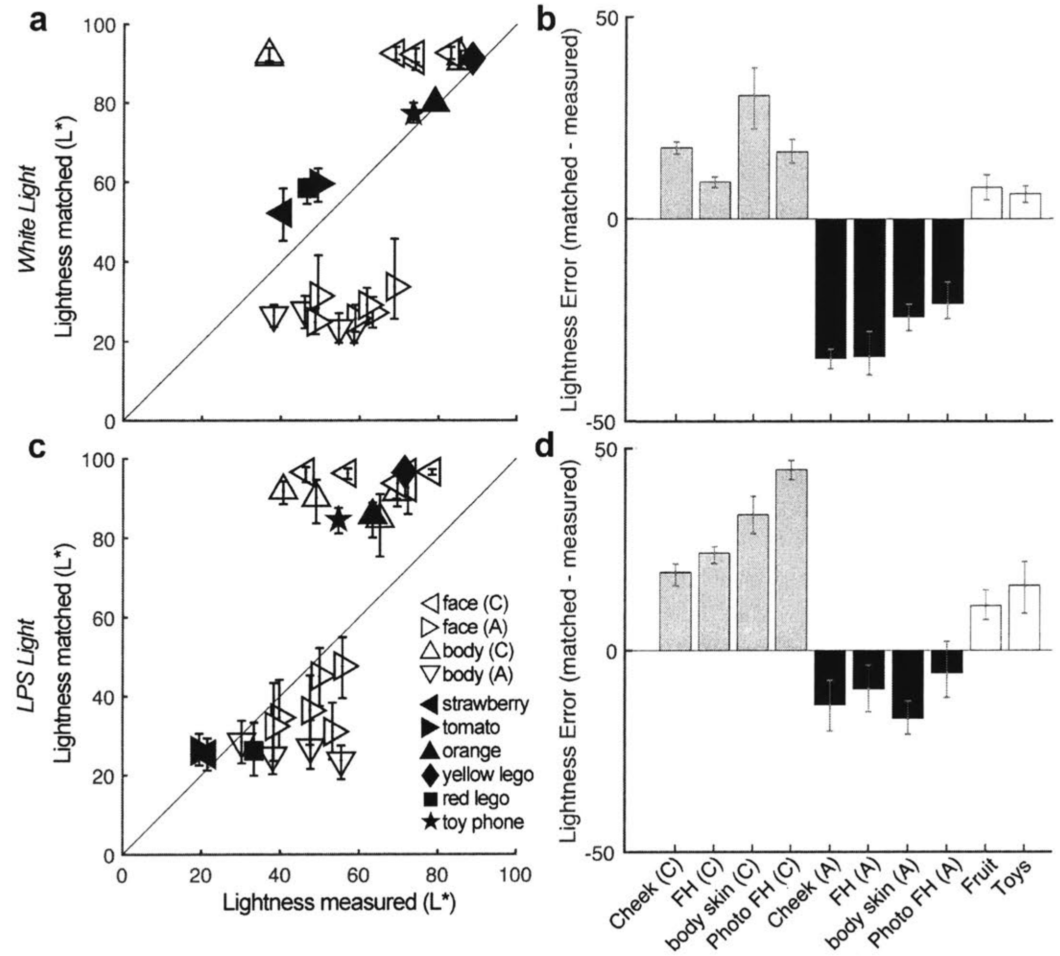
4.9. Supplemental Figures



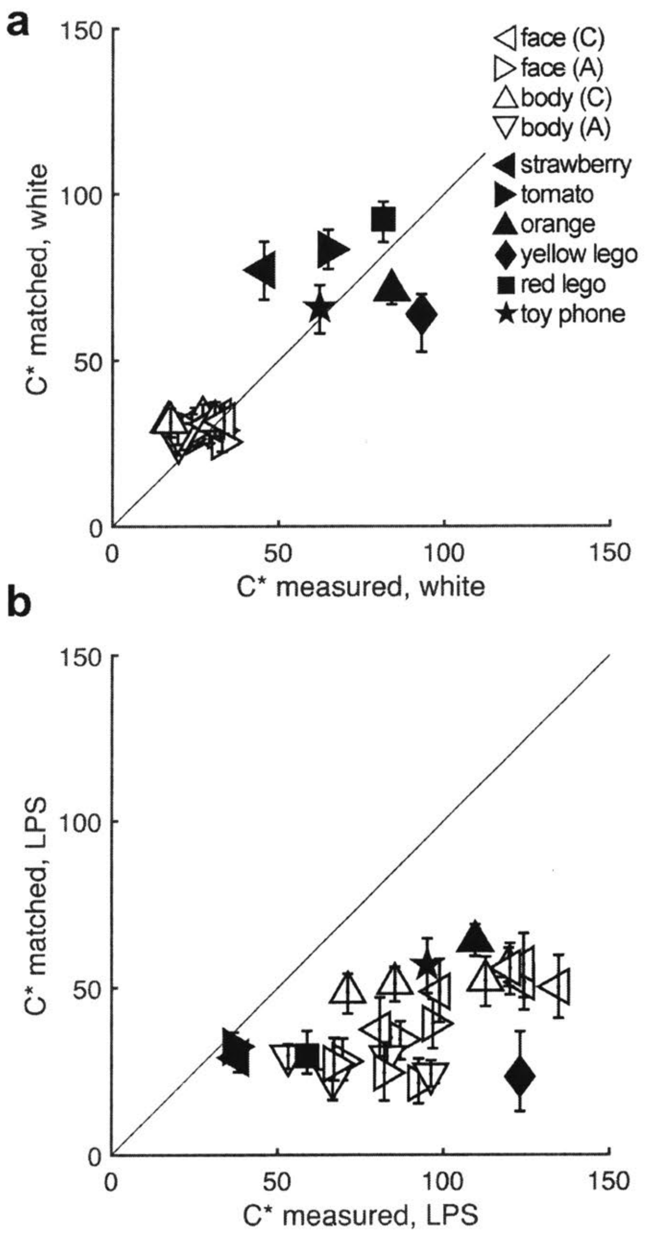
Supplementary Figure 1. Spectral measurements of test stimuli. a) Photographs of the seven objects used, and of two of the actors. The objects are labeled with a letter or a symbol to identify the corresponding spectral measurements given in panels b, c, d, e. The objects were chosen because they occupy the same quadrant as skin in color space. The photographs were taken under white light (top row) and low-pressure sodium light (bottom row), with the white balance set by the white card under white light. The photographs are shown in both color and black and white. b, Hue components of spectral measurements of the objects under white light, plotted in a color space designed to be perceptually uniform (CIE L* a* b*). The yellow asterisk shows the Macbeth white card measured under LPS light. Spectral measured XYZ values were transformed to L* a* b* using the XYZ values measured from the Macbeth white card under white light (CIE XYZ = [27.9446 25.0000 10.5134]). Embedded panels show the spectral power distribution of the illuminant (bottom), and the RGB values of the colors of the objects, white balanced using the spectrum of the light measured from the white Macbeth colorchecker card under white light (top). c, Hue coordinates of spectral measurements of the objects under low-pressure sodium light white balanced using the spectrum of the light measured from the white Macbeth colorchecker card under white light, as in (b). Other conventions as for panel (b). d) Lightness component (L*) of the spectral measurements shown in panels b and c plotted as a function of measured a* component (objects under white light, left; objects under LPS, right). e, Lightness of objects under white light vs LPS light (L* measurements from panel d). Source data are provided as a Source Data file.



Supplementary Figure 2. Color matches are reliable. Test-retest measurements for three participants, measured three months apart; correlation coefficient and p values are for the population. Regression lines for each participant are shown by the colored lines: Under white light: S1, $R^2=0.37$ ($p=0.036$); S2, $R^2=0.13$ ($p=0.24$); S3, $R^2=0.58$ ($p=0.004$); Under sodium light: S1, $R^2=0.88$ ($p=10^{-6}$); S2, $R^2=0.46$ ($p=.002$); S3, $R^2=0.89$ ($p=10^{-6}$). Source data are provided as a Source Data file.



Supplementary Figure 3. Race impacts the lightness of the color matches. a, Matched lightness (CIE L*) versus measured lightness values for objects (filled symbols) and skin (open symbols; key in panel c). b, Lightness error (matched minus measured) for different stimuli (axis label in panel d). Error bars are 99 % C.I. Symbols as for Figure 2. c, d, as for panels a, and b, but for matches made under low-pressure sodium (LPS) light. Source data are provided as a Source Data file.



Supplementary Figure 4. Chroma is underestimated under low-pressure sodium light. a, Matched chroma versus measured chroma for objects (filled symbols) and skin (open symbols) under white light. Symbols as for Figure 2. **b,** Matched chroma versus measured chroma for objects (filled symbols) and skin (open symbols) under low-pressure sodium light. Source data are provided as a Source Data file.

Chapter 5: Conclusions

5.1. Summary

This thesis describes functional MRI and psychophysical studies in humans that provide clues about the perceptual and neural mechanisms underlying our rich experience of a colorful world.

The functional MRI studies presented in Chapter 2 found color-biased cortex is sandwiched between face-selective and place-selective cortex on the bottom surface of the brain in humans. This face/color/place organization mirrors the organization observed on the lateral surface of the temporal lobe in macaques (Lafer-Sousa and Conway, 2013), suggesting that the tripartite system is homologous between species. This result validates the use of macaques as a model for human vision, making possible more powerful investigations into the connectivity, precise neural codes, and development of this part of the brain. Further, this work found that color and shape are segregated in the posterior extent of the ventral pathway and converge anteriorly. These results appear to be consistent with a model of the ventral pathway captured by parallel multi-stage processing of color, shape, faces, and places.

The psychophysical studies presented in Chapter 3 establish the #theDress as an ambiguous multi-stable color stimulus, show that the variable percepts arise because the illumination cues in the image are ambiguous and people make different implicit inferences about the lighting conditions, and demonstrate how the image can be used as a psychophysical tool to test hypotheses about how humans recover illuminant-invariant color. The work provides evidence that color constancy is mediated by sensory, perceptual, and cognitive factors (i.e., low-level features, inferences about 3D scene geometry, prior knowledge, and attention), and provides the first behavioral evidence that human skin is a sufficient cue to recover the illuminant to bring about color constant percepts.

Finally, the psychophysical studies presented in Chapter 4 show that when retinal mechanisms of color are impaired, memory has a paradoxical impact on color appearance that is selective for faces, providing evidence that color contributes to face encoding and social communication. The finding is a compelling example of the ability of cognition to affect perception and sheds light on the selective pressures that brought about trichromatic color vision.

5.2. Discussion

These findings connect to broader questions and inform future directions that I discuss next.

5.2.1. Homology and Evolution

Much is understood about how cone signals from the eye are encoded and transmitted to cortex, but less is known about the mechanisms that decode these signals to bring about perception and guide behavior. Over the past several decades, patient data and neuroimaging have powerfully facilitated a coarse mapping of the anatomical and functional architecture of much of human cortex. And while recent advancements have refined these methods to uncover finer grained neural representations and connectivity patterns, the tools remain limited. Understanding the precise computations conducted within specialized cortical regions, their specific causal role in perception, underlying connectivity, and development requires an animal model. Animal models are most useful when the regions of interest derive from areas present in a common ancestor (i.e. they are homologs). Historically, macaque monkeys have played a central role in advancing our understanding of human vision. The similar functional and anatomical organization of early visual cortex across the two species has been taken as strong evidence for homology. But bolstering the case for homology of the area(s) anterior to early visual cortex, where high-level visual processing culminates, remains key to validating and guiding the use of macaque IT as a model for the human ventral visual pathway (ventral temporal cortex, VTC). This is for two reasons: First, the putatively homologous areas are in grossly different geographic cortical locations (on the lateral surface in macaques and on the ventral surface in humans); and second, these areas are adjacent to V4, an area where the homology between humans and macaques becomes tenuous because of differences in its retinotopic organization. In humans, V4 comprises a single contralateral hemifield representation on the ventral surface, whereas in macaque V4, the upper and lower halves of the hemifield representation are split, located on the ventral and dorsal surfaces of the occipital cortex respectively (Goddard et al. 2011; Winawer and Witthoft 2015).

In Chapter 2 I showed that the functional organization of human ventral temporal cortex (VTC) bears a striking similarity to the organization of macaque inferior temporal cortex (IT) — in both species color-biased regions are systematically sandwiched between face and scene selective regions, forming a highly stereotyped tripartite map. Despite being found in geographically distinct locations of cortex, the suspiciously similar internal organization of these two broad swaths of cortex is characteristic of a 'regional homology'—a neighborhood of cortical regions present in a common ancestor whose geographic location was altered by the non-uniform expansion of cortex as new cortical areas emerged nearby (Orban et al., 2004). The disproportionate expansion of human cortex around the superior temporal sulcus to accommodate the acquisition of language and social processing provides a compelling account for how the tripartite face/color/scene map could have been pushed onto the ventral surface.

While these results provide functional evidence for homology (showing not only similar functional response properties, but similar internal functional organization), the case would be strengthened by anatomical evidence (e.g. similar connectivity, cyto- and myeloarchitecture). Connectivity measures are currently more tractable in humans than measures of cyto- and myeloarchitecture. Accordingly, they have been used to provide evidence for interspecies homology (e.g. Mars et al., 2011), predicated on the notion that if regions are homologous, they should have a similar 'connectivity fingerprint'—their unique pattern of input and output connections with the rest of the brain (Passingham et al., 2002). This method is well suited to establishing homologies because it characterizes regions in an abstract connectivity space, unconstrained by cortical geography. I spent considerable effort trying to use tractography and resting functional correlation to parse out the connectivity of the color-biased regions and their face and scene-selective neighbors, but distinguishing the precise connectivity of nearby patches of cortex is particularly challenging. I have concluded that these methods are not currently adequate for the task at hand. Perhaps future developments will enable such endeavors. Ultimately, a convergence of functional, anatomical, and molecular/genetic evidence will most conclusively establish homology. Nonetheless, the functional evidence presented here provides strong support for the use of monkey IT cortex as a model of the ventral visual pathway in humans.

In addition, the functional organization of color, face, and scene responsive regions in the two species speaks to the ecological significance of these systems across primate species (after all, they appear to have been conserved for at least 30 million years) and informs evolutionary theories about how cortex evolved. Future work would benefit from a similar analysis of the organization of face, color, and scene responsive cortex in a simpler primate species: the marmoset. While the existence of face patches in the marmoset is well established (Hung et al., 2015), little is known about the functional organization of color in this species.

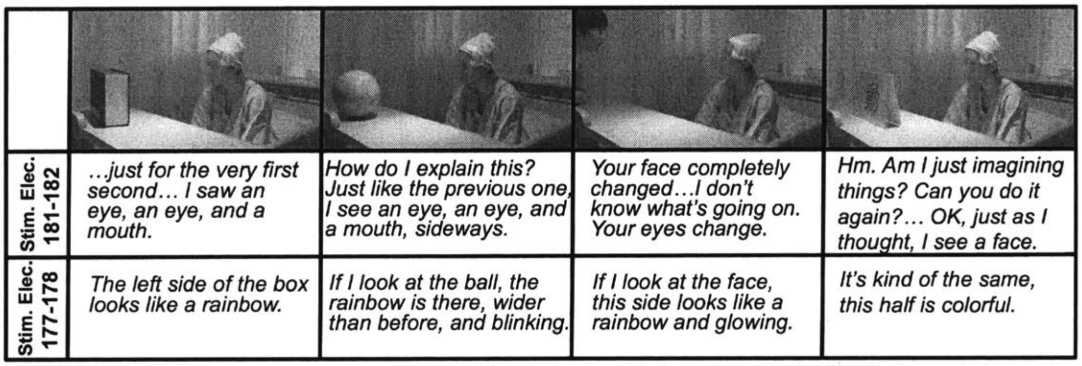
5.2.2. How segregated is the processing of color and shape in ventral temporal cortex?

In Chapter 2 I presented functional imaging evidence showing a strong double dissociation of color and shape in posterior ventral temporal cortex (with some convergence occurring anteriorly) and a weak dissociation of color and category. These dissociations were based on univariate measures, which are coarse. The presence of weak responses to a non-preferred stimulus dimension may nonetheless carry significant information about that dimension in the pattern of responses across voxels. Quantifying these regions' response profiles using a method that is sensitive to response pattern rather than response strength (such as multivariate pattern analysis) could uncover conjoint color and shape/category information in these regions and this remains an important direction for future work. But an especially powerful test of the functional specificity of these regions would be to directly intervene on the regions and ask what happens perceptually.

Causal evidence for the segregation of color and shape processing

Causal studies are nearly impossible to carry out in humans. On occasion we are afforded the opportunity in a clinical setting. In a separate line of work not presented in this thesis, my colleagues and I collaborated with clinicians to use direct electrical stimulation of focal regions along the VTC in a neurosurgery patient to assess the causal specificity of the color/shape dissociation suggested by my imaging results (Schalk et al., 2017). Using a modified version of the stimulus paradigm presented in Chapter 2, the clinicians mapped the selectivity of the cortical tissue underlying electrodes implanted in this area. As expected from my imaging data, this revealed a color-biased region immediately adjacent to a face-selective region. The color-biased site anatomically corresponded to one of the posterior color-biased regions (Pc or Cc). Electrical stimulation was used to ask whether this region is exclusively involved in processing color information, or if it plays some causal role in the perception of shape. Electrical stimulation of the color-biased site did not affect the appearance of face or objects' forms, but instead produced illusory rainbow percepts, apparently on top of them (Figure 1). Similarly, electrical stimulation of the adjacent face-preferring site during viewing of nonface objects produced only illusory face percepts, or facephenes, and did not affect color appearance. This causal double dissociation supports the selective, and perhaps exclusive, causal role of these regions in processing their preferred dimensions only.

The results are consistent with the observation that localized damage to regions of posterior VTC can give rise to selective deficits in color perception, often (though not always) in the absence of form deficits, and vice versa (Heywood et al, 2003; Bouvier and Engel, 2005; Mapelli and Behrmann 1997; Steeves et al., 2004). Taken together, the functional and causal dissociation of color and shape in posterior ventral temporal cortex support the idea that information about color and shape are segregated at this stage of the visual pathway.



Stim. Elec. 181-182	*...just for the very first second... I saw an eye, an eye, and a mouth.*	*How do I explain this? Just like the previous one, I see an eye, an eye, and a mouth, sideways.*	*Your face completely changed...I don't know what's going on. Your eyes change.*	*Hm. Am I just imagining things? Can you do it again?... OK, just as I thought, I see a face.*
Stim. Elec. 177-178	*The left side of the box looks like a rainbow.*	*If I look at the ball, the rainbow is there, wider than before, and blinking.*	*If I look at the face, this side looks like a rainbow and glowing.*	*It's kind of the same, this half is colorful.*

Figure 1. Transcript excerpts (bottom) from patient's report during electrical stimulation of electrodes 181-182 (FFA) and 177-178 (color-preferring site) while viewing a box, a ball, the experimenter's face, or a kanji character (reproduced from Schalk et al., 2017).

Convergence of color and shape processing

At some stage color and shape information must be integrated, after all color is bound to objects in visual experience. Suitably, the anterior color-biased region (Ac) showed a significant preference for shape. Further, this region could not be localized with simple stimuli (drifting gratings), only with naturalistic stimuli (Chapter 2, Figure 7). On the one hand, this could have been a consequence of simple stimuli failing to elicit strong responses across the visual system, making the detection of meaningful difference responses in Ac, where fMRI signal is particularly noisy due to known signal distortions, difficult. On the other hand, it could reflect a transformation of color signals from simple hue representations to high-order real world object/entity representations, integrating object color information with other object properties, including shape. This notion is reinforced by prior neuroimaging work showing that the anterior color-biased region is activated when subjects are in engaged in the retrieval of object color properties, even in the absence of color stimuli, such as accessing knowledge that taxis are typically yellow (Simmons et al., 2007); and patient data showing that lesions in the vicinity of this region can result in a selective behavioral deficit in the ability to associate object color knowledge with object shape (e.g. failing to accurately attribute the color red to a line drawing of a strawberry), while sparing color and shape perception more generally (e.g. Miceli et al., 2001; Stasenko et al., 2014). Given these observations, this region may be particularly important in mediating memory color effects, which bring prior knowledge about typical object colors to bear on perception (discussed in section 5.2.4).

A similar progression from segregated to convergent color and shape responses occurred within the face network: while the posterior face region (OFA) showed no preference for colored stimuli, downstream face-selective regions did so, and this bias increased anteriorly (particularly evident in the anterior face region ATL and the lateral face/social perception region pSTS; though the color-biased regions were still significantly more color-biased than even the most color-biased face region). A similar pattern of results was also observed in monkeys (Lafer-Sousa and Conway, 2013). The role of these weak but significant color signals in face processing remains unknown. Future studies may find they carry important information about social status (health, dominance, emotional state).

Taken together, these results support the idea that ventral temporal cortex is comprised of initially segregated pathways for shape and color, which progressively converge to bring about unified object representations. They also underscore the importance of using naturalistic stimuli to probe high-level color representations and inform directions for future work.

5.2.3. Why is color segregated from shape in posterior VTC?

Differences in the computational constraints and task demands associated with color and shape could be responsible for their apparent segregation VTC. In this section I consider the putative and theoretical functions of the posterior and central color-biased regions with respect to the shape-independent computational challenges and behaviors associated with color.

Detection, categorization, and discrimination of hue

Humans are capable of distinguishing upwards of ~40 million colors (Kuehni, 2015), but employ only a limited set of color categories. That is, perceptual color space is both continuous and categorical. Converging evidence implicates the posterior color-biased region (which extends from V4 into VO-1) in the detection, discrimination, and categorization of hue, independent of shape. Color responses in human and macaque V4 and its anterior neighbor (VO-1 in humans and PIT in monkeys) encode hue, marking a significant transformation from color difference signals (encoded in V1) to representations that correspond well with perceptual color space (Conway et al., 2007; Brouwer and Heeger, 2009, 2013; Bohon et al., 2016). In fact, the "rainbow" percepts elicited in the causal study detailed in the previous section, may have been predicted from the micro-organization of color-tuned cells in macaque V4/PIT, where patches of narrowly hue-tuned cells (dubbed 'globs') are spatially arranged into chromotopic maps that reflect perceptual color space (i.e. nearby cells code nearby colors in perceptual color space), not unlike a rainbow (Conway et al., 2007; Conway and Tsao, 2009).

Moreover, MEG and fMRI results in humans have demonstrated that the neural representation of color in V4 and VO-1 can be modulated by task demands related to flexibly switching between categorical color perception and making fine color discriminations: engaging in a color-naming task shifts the neural color spaces to a categorical representation (Brouwer and Heeger, 2013); and deploying feature-based attention for a particular color, results in a sharpening of the neural population tuning for the attended color, enhancing the ability to make fine discriminations and detect sought after colors during visual search (e.g. when searching for your car in a crowded parking lot, where shape information is not particularly helpful) (Bartsch et al., 2014, 2017).

Higher-order color statistics for material recognition?

Color vision informs us about more than just color. Like texture and gloss, color is helpful for inferring what objects/entities are made of (wood, flesh, stone, fabric, plastic, rubber, water, paste, etc.), what their physical properties are (hard, soft, rough, smooth, viscous, wrinkled, fragile, deformable, etc.), and what state they are in (ripe, rotten, wet, dusty, burnt, rusted, aged, etc.). This information fundamentally informs our interactions with objects/surfaces (e.g. whether to eat something; how cautiously to walk on a surface). The segregation of color and shape observed in posterior VTC may arise due to differences in the computational and behavioral demands associated with processing surface properties in general compared to those for global form, and raises questions about whether the color-biased regions in posterior VTC should be considered strictly involved in color per say or material perception more generally.

Like color, material perception is not trivial (variations in illumination and viewing angles generate substantial ambiguity in the retinal image) and is not always contingent on global shape (for a review see Fleming, 2014; Komatsu and Goda, 2018). Different objects can be made from the same material (e.g. a metal spoon and a metal fork) and the same object can be made from different materials (e.g. a metal spoon and a wooden spoon). Fittingly, people can often (though not always) recognize materials in the absence of global shape information, and such recognition is significantly enhanced by the presence of color (Sharan et al., 2010). Moreover, changes in intrinsic surface reflectance that occur due to changes in material state are not always shape-contingent. For example, when most surfaces get wet, whether a leaf or a patch of pavement, their higher-order color statistics change in a systematic predictable way, which humans are sensitive to (Sawayama et al., 2017). These observations imply some degree of dissociation of surface and shape processing in the brain, along with some linkage of color and other surface properties.

Converging functional imaging evidence suggests that surface properties, particularly natural texture, are processed largely independently of object shape, beginning in early visual cortex and culminating in the medial part of the VTC (Cant and Goodale 2007; Cant and

Goodale 2008; Cavina-Pratesi et al., 2009, 2010). The posterior and central color-biased regions fall in the neighborhood of this putative 'texture pathway', along with regions implicated in processing gloss (Wada et al., 2014) and material category (Hiramatsu et al., 2011), raising the possibility that they may be part of a network that is specifically concerned with material perception. My work did not explicitly address the extent to which color and other surface properties like texture and gloss are represented by common or distinct neural machinery. The strong responses I observed in the posterior and central color-biased regions to scrambled-objects could reflect sensitivity to texture, but this is purely speculative.

Additional work is needed to establish how higher-order color statistics that support material perception are encoded in the brain. To this end, one could use fMRI (or multi unit recordings in monkeys) to ask where in the brain a property like 'wetness' can be decoded (invariant to surface-identity) from neural responses to images of wet and dry surfaces (leaves, rocks, sand, pavement, paper, etc.). A similar paradigm could be deployed to probe how different surface statistics interact to achieve material property representations. Ultimately, a full description of material perception will require accounting for the role of 3D shape information, as the information conveyed by surface reflectance properties is insufficient to capture the full range of material recognition abilities humans display (Sharan et al., 2013). It seems likely that the anterior convergence observed for color and shape may also be accompanied by material quality signals, but this remains to be seen.

Higher-order color statistics for object recognition?

In addition to informing us about material properties, higher-order color statistics like chromatic textures can be informative about object identity. Most natural objects have polychromatic surface textures (while a banana is categorically yellow, it is not uniformly so). Pigment inhomogeneities, variations in micro-surface structure, and surface reflectance gradients that vary as a function of an object's 3D shape result in characteristic color distributions, which form distinctive signatures in cone-contrast space. Some have speculated that these chromatic signatures contribute to object recognition and color constancy (Hurlbert et al., 2007) and have shown that even in the absence of global form information the visual system is able to exploit the polychromaticity of familiar objects in recognition tasks, which suggests color-texture conjunctions contribute to object recognition (Vurro et al., 2010). As well, color-tuning biases in macaque IT are predicted by object color statistics (estimated from an analysis of a massive image database of natural and artificial objects) (Rosenthal et al, 2018), implying color signals in IT play a role in real world object representations. Future work is required to determine whether (and where) object identity can be decoded from neural responses to restricted patches of familiar polychromatic surfaces (like fruits).

Finally, examining how computational models handle color, texture, gloss, 2D and 3D shape may shed light on the extent to which computational and task constraints are responsible for the segregation of shape and surface properties. For example, given a convolutional neural network (CNN) trained on object recognition, one might ask how segregated these dimensions are, at what layer they converge, and whether this organization varies with task (e.g. does this organization differ in a network trained explicitly on material recognition?). Likewise, one might ask how heavily the network relies on a particular feature dimension to carry out its task (e.g. to what extent does the network use diagnostic color for object recognition?) and how readily its features transfer to a new task.

5.2.4. Memory color: the impact of object-color knowledge on perception

In order for the visual system to exploit the regularities of object color statistics (as suggested above) we must learn and store those statistics. This stored knowledge about an object's typical color, as shaped by experience, is referred to as memory color. A memory color consists of a typical color and a tolerance range (the range over which particular colors are considered acceptable examples of the memory color). Memory color is closely related to the concept of diagnostic color. Objects that have a typical color are considered color-diagnostic. For example, grass is usually green and bananas are usually yellow (color-diagnostic), but cars vary considerably in their color (non-diagnostic). While both grass and bananas can vary from green to yellow to brown, the most canonical color of grass is green and a banana is yellow (these colors also correlate with the objects' most behaviorally valued state, lush and ripe). Color-diagnostic objects automatically elicit awareness of and overt attention to their memory colors and vice versa (Naor-Raz et al., 2003; Nijboer et al., 2006), which makes memory color useful for object recognition and object quality assessment. When color is diagnostic, memory colors enhance the speed and accuracy of object and scene recognition, especially when shape is ambiguous or occluded (for a review see Bramao et al., 2011). When color-diagnostic objects vary in color due to changes in state (fruit ripening, food rotting, faces palling during illness, metal rusting, surfaces accumulating dust, wetting, drying, burning, ageing, etc), memory color may be useful for recognizing these state changes, informing assessments of an object's behavioral value and guiding decisions about how to interact with it.

When Hering first proposed the notion of memory color in 1878 he suggested that knowledge about the typical color of objects may influence perception of actual object colors (cognitive penetrability), and in particular that such knowledge might support color constancy by serving as a reference to recover information about the illuminant in a scene (Hering, 1878). Many studies since have found evidence that memory color can modulate color appearance ('memory color effects'): A color-diagnostic object can induce the impression of its memory color, even in the absence of a color signal (i.e. a grayscale banana appears tinged in yellow) (e.g. Adams, 1923; Hansen et al., 2006). Similarly, when a categorically ambiguous color, one that falls near the boundary of a color category (say yellowish-orange) is displayed on a color-diagnostic object (like a banana or a carrot) object identity can shift how the color is perceived (perceived/categorized as orange on a carrot and yellow on a banana) (Mitterer and Ruiter, 2008). And finally, consistent with Hering's hypothesis, memory colors can support color constancy, though their contribution may be relatively limited compared to low-level mechanisms like adaptation and simultaneous contrast (Granzier and Gegenfurtner, 2012).

However, memory color effects remain controversial, due to variable results across studies and the nagging possibility that positive results could be attributable to demand characteristics (if a study participant knows that a banana is yellow, they may provide a behavioral response consistent with this knowledge without actually experiencing a change in perception) (for a review see Adeyefa-Olasupo and Flombaum 2018; Valenti and Firestone, 2018). The psychophysical experiments presented in Chapter 3 (testing the role of memory color for skin in color constancy) and Chapter 4 (probing memory colors under real-world three dimensional viewing conditions in which retinal mechanisms of color are crippled) uncover novel memory color findings and speak to these broader concerns.

Memory color for skin as a cue for color constancy

Several studies have found evidence that memory color contributes to color constancy: when illumination cues are limited, color-diagnostic objects are more tolerant to variations in illumination (more color constant) than non-diagnostic objects (e.g Hurlbert and Ling, 2005); further, the presence of color-diagnostic objects in a scene can enhance the color constancy of other non-diagnostic objects in the scene, though this work suggests color-diagnostic objects' contribution to color constancy is modest compared to low-level mechanisms like adaptation (Granzier and Gegenfurtner, 2012). Among color-diagnostic objects, skin is one of the most diagnostic. In fact, the earliest use of the word 'color' seems to have been in the 13th century as 'skin color' (Zaidi, 2011). Skin is a polychromatic surface whose color statistics are remarkably stable and form a distinctive profile in cone-contrast space that varies predictably under changing illumination (Bianco and Schettini, 2012; Crichton et al., 2012). Moreover, people have precise memory colors and preferences for skin and show little tolerance for color deviations in reproductions of skin (Bartleson and Bray, 1962; Chauhan, et al., 2015; Hunt et al.,1974; Sanders, 1959; Smet, et al., 2011; Yendrikhovskij, et al., 1999). Given these observations, and the prevalence of skin in visual experience, it has long been theorized that skin might serve as an especially useful reference surface toward stabilizing color percepts, but this has never been behaviorally tested.

In Chapter 3 I manipulated #theDress image to provide the first behavioral evidence that the visual system can use the spectral signals from skin combined with stored knowledge about skin color (memory color) to infer the spectral bias of the illumination in a scene and bring about illuminant-invariant percepts of other objects in the scene. Contrary to prior reports showing that memory color contributions to color constancy are modest compared to low level cues like adaptation and simultaneous contrast (Granzier and Gegenfurtner, 2012), the present results found skin was far more effective at stabilizing #theDress' colors than a uniform chromatic surround (designed to engage low-level mechanisms). While skin is not the only cue the visual system uses to stabilize color percepts, this work demonstrates that it is a sufficient and especially powerful one. The finding also highlights the value of #theDress as a stimulus for probing high-level cues to color constancy, as its colors could not be stabilized by low-level cues (the uniform surround condition failed to bring about color constant behavior).

Memory color effects under real world, three-dimensional viewing conditions

In chapter 4 I presented an assessment of the impact of memory on the color appearance of real three dimensional color-diagnostic objects (fruit), faces, and non-diagnostic objects (toys) using a novel paradigm that exploited narrow-band low pressure sodium (LPS) light to cripple the retinal mechanisms of color vision while preserving all other aspects of visual experience. Contrary to prior reports (e.g. Hansen et al., 2006; Witzel et al., 2011), the study did not detect canonical memory color effects for familiar fruit objects: fruits were not perceptually tinged in their typical color under LPS light. The study did uncover a surprising paradoxical impact of memory on the color appearance of faces: rather than appearing tinged in their normal (reddish) color, faces appeared greenish under LPS light. How do we reconcile these findings with the prior literature?

Why were memory colors for fruit not detected under LPS light?

The literature on the canonical memory color effect (MCE)—where an objectively achromatic object appears tinged with its typical color—dates back nearly a century and is exceptionally inconsistent. Having reviewed the checkered literature and spent considerable time experimenting in the lab, I have concluded that the effect is contingent on the ambiguity of the sensory signal, consistent with a standard Bayesian framework: heightened sensory ambiguity increases the influence of memory color priors in mediating color percepts. I posit that the mixed results reported in the literature largely stem from 1) variable ambiguity inherent to variable task designs, and 2) commonly failing to account for achromatic biases related to illumination priors (Winkler et al., 2015).

Several lines of evidence suggest the importance of ambiguity in eliciting the MCE: First, the earliest study documenting an MCE was carried out in 1923 and demonstrated that when a shape that had a trained color-association was cut from gray cloth and viewed under *dim lighting conditions*, subjects reported that it appeared to have its typical color (Adams, 1923). I have replicated this finding in an informal classroom setting using a black and white American flag: when the flag is viewed under bright light it appears achromatic, but under very dim light (though not scotopic) it appears colorful (canonical red, white, and blue) (unpublished). The effect arises because dim light ostensibly creates sufficient ambiguity in the sensory signal to elicit the memory color effect. Second, when a color that falls near the boundary of a color category (say yellowish-orange) is displayed on a color-diagnostic object (like a banana or a carrot) object identity can shift how the color is perceived (perceived as orange on a carrot and yellow on a banana) (Mitterer and Ruiter, 2008). In this case, the ambiguity stems from the color's proximity to a category boundary (it is categorically ambiguous). And third, in experiments that find an impact of memory color on color constancy, acute illumination ambiguity is an explicit feature of the experimental design (intended to enable the isolation of memory color contributions to color constancy).

The most recent sweep of memory color studies are most similar in design to the 1923 study from Adams—they probe MCEs for color-diagnostic objects rendered in grayscale—except they are carried out on *bright* computer displays, which lack the ambiguity of the Adams-style dim light setting. Curiously, these studies still report significant MCEs (e.g. a grayscale banana reportedly appears yellowish; Hansen et al., 2006; Witzel et al., 2011). Why? An often-overlooked characteristic of these findings is that they usually only recover memory color effects for yellowish-orangeish objects (not red, blue, or green). I speculate that the MCEs recovered in these studies arise due to achromatic biases related to illumination priors, specifically the known blue-yellow asymmetry that biases people to infer blueish illumination in near-neutral (achromatic) settings (Winkler et al., 2015; Adeyefa-Olasupo, 2018).

Taken together, these observations are consistent with the conclusion that canonical memory color effects are contingent on sensory ambiguity, which likely accounts for our failure to recover typical MCEs for fruit and faces under LPS light. While LPS light generated weird sensory signals, they were not particularly ambiguous—the illumination was bright and its chromatic bias was discernable. This ambiguity-contingency also comports with a sensible processing strategy: Many color-diagnostic objects undergo intrinsic color changes that provide valuable information about object state (e.g. a piece of fruit ripening). In general, one would not want the visual system to shift the perceived color of objects toward their memory colors under unambiguous viewing conditions, because doing so would obscure meaningful intrinsic color variations.

Why did faces appear greenish under LPS light?

If canonical memory color effects are contingent on sensory ambiguity and LPS light doesn't introduce much sensory ambiguity, why did we observe an impact of memory on the apparent color of faces? The green face effect ('paradoxical memory color') is different from the canonical memory color effect. The perceptual color shift is not towards the typical color of face skin (reddish), but away from it (greenish). While the sensory signal was not ambiguous, it did violate a strong prior for healthy face color. Specifically, it did so in a way that mimics how unhealthy face color signals violate priors for healthy face color. Both illness and LPS lighting produce an objective decrement in face redness, i.e. a shift toward green. Though the signals are not objectively (spectrally) green, both sick faces under normal illumination and healthy faces viewed under LPS light appear greenish. In effect the visual system exaggerates the color signal's deviation from expected healthy face color, thereby inducing an illusory greening effect that makes salient a behaviorally meaningful violation of healthy face color appearance.

The finding is a novel and striking example of the ability of cognition to influence perception, and is especially important because it provides an example of cognitive penetrability that can't be explained by demand characteristics—after all, if subjects' color reports reflected their knowledge about the typical color of faces, they would report that faces appeared reddish, the canonical color of faces.

5.2.5. Color constancy: what are the cues and how are they weighted?

How the brain transforms the inherently ambiguous color signals that enter the eye to create surface color representations that are stable across variations in illumination has been a fundamental question in the field of color vision for decades, yet it remains poorly understood. While all retinal images are inherently ambiguous with respect to illumination, the reliability of certain properties of the environment (e.g. the spectral properties of typical illuminants) and the physics of 3D image formation produce statistical regularities that are usually sufficient to constrain the ill-posed problem of recovering relatively stable color percepts despite substantial changes in illumination. It is rare to encounter a natural image or scene for which the available illumination cues are sufficiently ambiguous to break color constancy, and it is vanishingly rare for such an image to result in more than one stable color percept. #TheDress image was a significant discovery to the field of color vision, as it constituted the first striking example of an ambiguous color image, one that could be resolved as more than one stable categorical color percept. While there are many examples of color illusions that demonstrate the subjective nature of color perception—images for which identical patches of pixels are made to appear differently colored by manipulating the context in which they appear (e.g. Chapter 1, Figure 2)—any given exemplar gives rise to a single color percept, within and across individuals. One exception might be the mach color card (Bloj et al., 1999), whose perceived color varies depending on how its 3D structure is resolved, but eliciting such variability does require manipulating the available depth cues using stereoscopic filters. The dress is an image for which two individuals can look at the same image, on the same screen, and report different color percepts.

In Chapter 3 I provided a rigorous characterization of the phenomenon, establishing that: 1) the image is in fact perceived as more than one categorical color percept; 2) it is susceptible to reversals (though it is 'sticky'); 3) the variable percepts arise because the illumination cues in the image are ambiguous and people make different implicit inferences about the lighting conditions in the image; and 4) low-level and high-level factors mediate this process (e.g. scale, spatial frequency, attention, prior knowledge and experience).

One implication of these findings is that people have different priors and/or weight illumination cues differently, but that most of the time there is sufficient information in a scene to arrive at the same interpretation. Understanding how color constancy is achieved will require determining the rules that dictate how different cues are weighted and how dynamic this process is with respect to cue availability, cue reliability, cue covariance, and prior expectations. Maloney (2002) suggests the method of perturbation analysis, developed in studies of depth and shape perception. In this method, rather than selectively removing cues from a scene, cues are instead selectively perturbed to signal conflicting illuminants, and observers are queried about their percepts to determine which cues they rely on and to what extent. But synthesizing realistic complex scenes with these properties has been a challenge. #TheDress is an example of one such image. It is composed of pixels that correlate with both ends of the daylight axis (blueish skylight and yellowish sunlight), effectively cueing two conflicting illuminants with similar prior probabilities. The cues to the lighting geometry are similarly under constrained; they are consistent with two plausible, but conflicting interpretations: 1. a single light source is backlighting the dress, casting it's surface in shadow (shadows have a high prior probability of being blueish, as they are usually illuminated by diffuse blue skylight); 2. a global light in plane with the dress source is providing direct illumination (direct illumination has a high prior probability of being yellowish, the color of sunlight and most artificial light sources) (Figure 2).

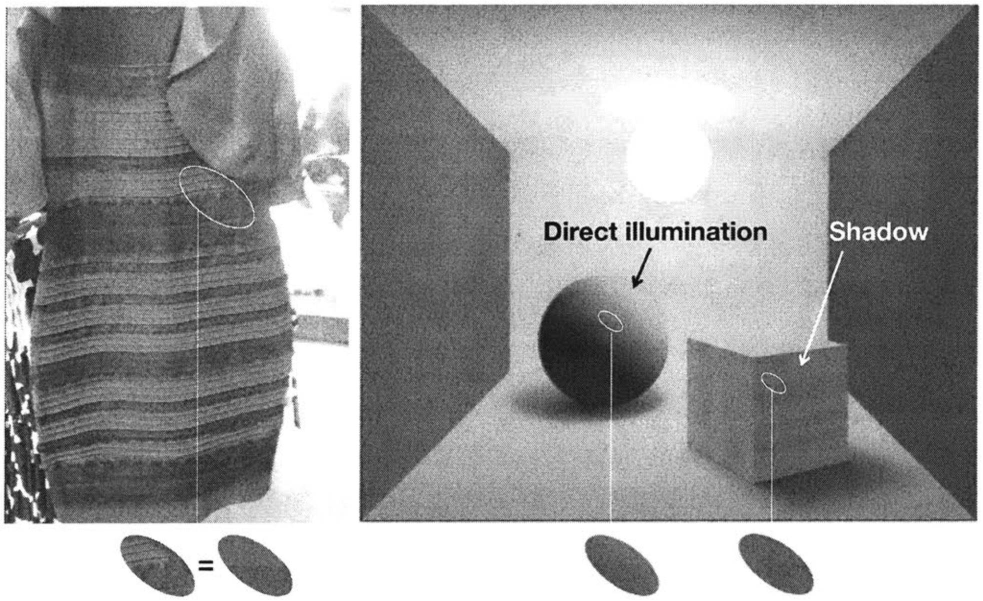


Figure 2. What are the lighting conditions? Is the dress (left panel) backlit and so cast in *cool shadow*? Or is it directly illuminated by bright *warm light*? The wavelengths that reach the eye from a blue and black surface directly illuminated by bright warm light are the same as the wavelengths that reach the eye from a white and gold surface illuminated by cool shadow (right panel); which appear as a low-saturation blue and brown when viewed in isolation (bottom ovals).

However, this is just one image. A complete understanding of what gives rise to #theDress' unique balance of realism and ambiguity should enable the creation of more stimuli, which has not yet been achieved. Since the emergence of the dress stimulus, several other images have been identified via social media that also elicit multiple perceptual color interpretations (Figure 3B-E). An analysis of the full image set could provide a common set of characteristics that predispose an image to this striking form of color ambiguity[2], offering promise for the development of controlled stimuli with these properties. If a sufficient number of exemplars naturally emerge, it may become possible to train a generative adversarial network (GAN) to generate novel exemplars. The development of such stimuli would enable complex cue perturbation studies toward uncovering the cues and weighting rules individuals use to achieve

---

[2] At first blush, there are some obvious similarities: 1) all the images in the set are composed of two low saturation, opponent colors; 2) one color component is usually perceived as achromatic, the other as chromatic (consistent with observers attributing the chromatic bias of one component to the illuminant); 3) all appear to have non-uniform light fields and poor exposure quality; and 4) all are tightly cropped, making judgments about scene geometry and lighting configuration ill-posed.

color constancy, and would enable novel physiological and neuroimaging studies of its brain basis.

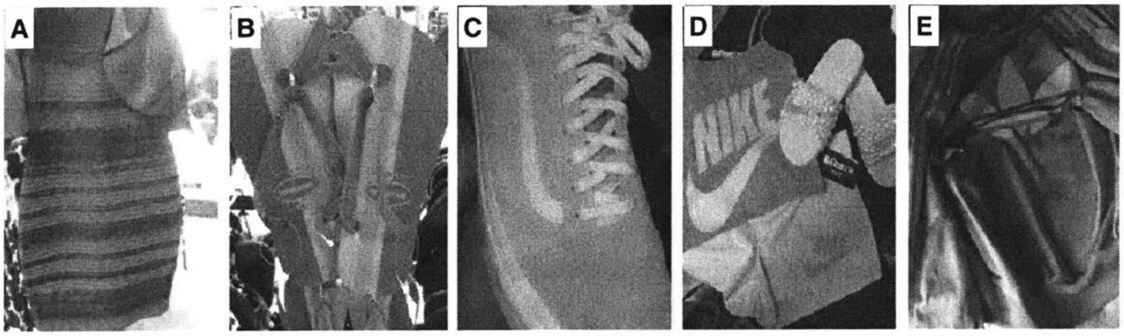


Figure 3. Ambiguous color images that give rise to multiple color percepts. A) "#heDress": Blue & Black; White & Gold; Blue and Brown. B) Blue & Black; White & Yellow; Blue and Gold (Arthur, 2017). C) Pink & White; Gray & Teal (Alisha, 2017). D) Mint & Gray; White & Pink; Blue & Pink (Rafaella, 2017). E) Blue & White; Black & Brown; Green & Gold (Mariam, 2016).

On the other hand, a list of cues and their weightings will never be sufficient to capture human experience. Ultimately, we need a more complete model of how people invert the causal processes underlying image formation. Ongoing efforts to understand vision as a process of inverting a forward graphics model to uncover the causal structure of the world that created the image look especially promising (Janner et al., 2017; Yildirim et al., 2018).

5.2.6. Where is perceptual surface color represented in the brain?

Conflicting data in the literature have raised questions about where perceptual surface color representations (invariant to illumination and subject to memory color effects) are achieved in the brain. While physiological results in monkeys and fMRI results in humans have established that color representations in V4 reflect perceptual dimensions of color (Conway et al., 2007; Brouwer and Heeger, 2009, 2013; Bohon et al., 2016), selective color constancy impairments acquired after lesion damage have not provided conclusive results—some cases implicate V4 (Clarke et al., 1998, Kennard et al., 1995), while others implicate V1 (Kentridge et al., 2007). Further, recent fMRI studies in humans have found that illuminant-invariant surface color can be decoded as early as V1 (Bannert and Bartles, 2017), as can memory color effects (e.g. the yellow of a gray banana) (Bannert and Bartles, 2013). Given the abundance of cortical feedback to V1, and the low temporal resolution of fMRI, it is conceivable (in fact likely) that the color information decoded from V1 in these studies actually reflects representations that have been revised by top-down feedback from high-order areas. In order to resolve the question of where color constant surface color is first achieved in the brain, I propose the use of MEG to test, with high temporal resolution, *when* color constant percepts can first be decoded in the human brain. Given that we have temporal benchmarks for feed-forward V1 and

V4 representations, the timing of MEG decoding should be helpful in determining where perceptual surface color representations are first achieved (in spite of the low spatial resolution of the tool). Such efforts will also be informative about the role of feedback in revising color representations at earlier stages of the visual pathway. I am working towards this goal in a current line of work not presented in this thesis.

5.2.7. A general framework for the ventral visual pathway?

The results presented in this thesis align with a general framework of ventral visual processing characterized by a transition from feature to real-world entity representations, organized by specialized subsystems for color, shape, faces, and scenes. Early visual cortex is retinotopically organized and strictly concerned with low-level feature extraction. When these signals reach the ventral stream, the transition begins from retinotopic feature-based representations to viewer-centered real world object/entity representations. Early in the transition, cortex is organized by segregated feature and partial object/entity representations: the posterior component of the face processing system is largely concerned with face features (Friewald et al., 2009; Issa and DiCarlo, 2012); the posterior component of the color-biased system (region Pc, corresponding to parts of V4 and VO-1) is concerned with representing perceptual hue (Conway et al., 2007; Brouwer, and Heeger, 2009, 2013); and the posterior component of LO is concerned with global shape, independent of semantic category (Grill-Spector et al., 1999). As processing proceeds anteriorly, shape features are transformed to higher-order representations of face and object forms (increasingly view-invariant and globally complete) (Meyers et al., 2015), while surface features (color, texture, and gloss), more speculatively, are transformed to higher-order representations of material quality. Finally, at the most anterior extent of the ventral pathway, these signals are unified to achieve complete real world object/entity representations. Meanwhile, representations of scene layout are established in a parallel subsystem (Epstein and Kanwisher, 1998), sensitive to shape and surface features.

Consistent with this framework, anatomical studies of macaque IT suggest the region is comprised of multiple parallel specialized subsystems (Kravitz et al., 2013). Importantly, these studies also establish that the system is not a strictly feed-forward serial hierarchy: extensive interconnectivity within IT and across the entire visual pathway (including feedback projections directly to V1) are observed, as well as elaborate connectivity patterns linking subcortical and prefrontal regions to different parts of IT. This complex connectivity is likely critical for resolving sensory and perceptual ambiguities, enforcing task demands, bringing memory to bear on perception, ascertaining behavioral affordances, forming concrete and abstract associations, and achieving learning. Understanding these rich complex processes will form the basis of decades of future vision research. This framework will serve as a useful road map, and color specifically, a useful model system to constrain the undertaking.

5.2.8. Closing remarks

This thesis presented an assessment of the functional organization of color, shape, and category information in the ventral visual pathway of humans and a comparison to non-human primates that suggests a broad homology across primate species and provides insights into the computational goals of the system and how it evolved. The results invite future work into the connectivity and developmental origins of the tripartite face/color/place system. The psychophysical experiments presented in this thesis yielded tantalizing theories about the high level neural and cognitive processes that mediate color perception. Testing these theories and probing the fine-grained computations supported by these regions will require novel experimental and data-analytical frameworks in human imaging (including MVPA, RSA, and SVM)—as well as more invasive paradigms in macaques, a research program made more tractable by the system's apparent homology to macaques.

5.2.9. References

Adams, G. K. (1923). An Experimental Study of Memory Color and Related Phenomena. *The American Journal of Psychology, 34*(3), 359. doi:10.2307/1413955

Adeyefa-Olasupo, I., & Flombaum, J. I. (2018). Tutorial Review: Evidence for the Memory Color Effect, 1923–2016. doi:10.31234/osf.io/ahqsm

Adeyefa-Olasupo, I E. (2018) Perception is shaped along the (L+M)–S axis (and possibly confused with a memory color effect) bioArXiv Preprints.

Alisha [zjmah]. (2017, October 11). The real shoe is pink & white okay!? [Twitter moment]. Retrieved March 20, 2019, from https://twitter.com/zjmah/status/918234911290613760

Arthur [positivedemi]. (2017, November 17). Que cor voces tao vendo? [Twitter moment]. Retrieved March 20, 2019, from https://twitter.com/positivedemi/status/799346903112630272

Bannert, M. M., & Bartels, A. (2017). Invariance of surface color representations across illuminant changes in the human cortex. *NeuroImage, 158*, 356-370.

Bannert, M., & Bartels, A. (2013). Decoding the Yellow of a Gray Banana. *Current Biology,23*(22).

Bartsch, M. V., Boehler, C. N., Stoppel, C. M., Merkel, C., Heinze, H., Schoenfeld, M. A., & Hopf, J. (2014). Determinants of Global Color-Based Selection in Human Visual Cortex. *Cerebral Cortex*, (9).

Bartsch, M. V., Loewe, K., Merkel, C., Heinze, H., Schoenfeld, M. A., Tsotsos, J. K., & Hopf, J. (2017). Attention to Color Sharpens Neural Population Tuning via Feedback Processing in the Human Visual Cortex Hierarchy. *The Journal of Neuroscience,* (43).

Bianco, S., & Schettini, R. (2012). Color constancy using faces. *2012 IEEE Conference on Computer Vision and Pattern Recognition.*

Bloj, M. G., Kersten, D., & Hurlbert, A. C. (1999). Perception of three-dimensional shape influences colour perception through mutual illumination. *Nature,* (6764).

Bohon, K. S., Hermann, K. L., Hansen, T., & Conway, B. R. (2016). Representation of Perceptual Color Space in Macaque Posterior Inferior Temporal Cortex (the V4 Complex). *ENeuro, 3*(4).

Bouvier, S. E., & Engel, S. A. (2005). Behavioral Deficits and Cortical Damage Loci in Cerebral Achromatopsia. *Cerebral Cortex,* (2).

Bramão, I., Reis, A., Petersson, K. M., & Faísca, L. (2011). The role of color information on object recognition: A review and meta-analysis. *Acta Psychologica,* (1).

Brouwer, G. J., & Heeger, D. J. (2009). Decoding and Reconstructing Color from Responses in Human Visual Cortex. *Journal of Neuroscience,* (44).

Brouwer, G. J., & Heeger, D. J. (2013). Categorical Clustering of the Neural Representation of Color. *Journal of Neuroscience,* (39).

Cant, J. S., & Goodale, M. A. (2006). Attention to Form or Surface Properties Modulates Different Regions of Human Occipitotemporal Cortex. *Cerebral Cortex,* (3).

Cant, J. S., Arnott, S. R., & Goodale, M. A. (2008). FMR-adaptation reveals separate processing regions for the perception of form and texture in the human ventral stream. *Experimental Brain Research,* (3).

Cavina-Pratesi, C., Kentridge, R., Heywood, C., & Milner, A. (2010). Separate Channels for Processing Form, Texture, and Color: Evidence from fMRI Adaptation and Visual Object Agnosia. *Cerebral Cortex,* (10).

Cavina-Pratesi, C., Kentridge, R. W., Heywood, C. A., & Milner, A. D. (2009). Separate Processing of Texture and Form in the Ventral Stream: Evidence from fMRI and Visual Agnosia. *Cerebral Cortex,* (2).

Chauhan, T., Xiao, K., Yates, J., & Wuerger, S. (2015). Estimating discrimination ellipsoids for skin images. *Journal of Vision,* (12).

Clarke, S., Walsh, V., Schoppig, A., Assal, G., & Cowey, A. (1998). Colour constancy impairments in patients with lesions of the prestriate cortex. *Experimental Brain Research, 123*(1-2).

Conway, B. R., & Tsao, D. Y. (2009). Color-tuned neurons are spatially clustered according to color preference within alert macaque posterior inferior temporal cortex. *Proceedings of the National Academy of Sciences,* (42).

Conway, B. R., Moeller, S., & Tsao, D. Y. (2007). Specialized Color Modules in Macaque Extrastriate Cortex. *Neuron,* (3).

Epstein, R., & Kanwisher, N. (1998). The Parahippocampal Place Area: A Cortical Representation of the Local Visual Environment. *NeuroImage,* (4).

Fleming, R. W. (2014). Visual perception of materials and their properties. *Vision Research.*

Freiwald, W. A., & Tsao, D. Y. (2010). Functional Compartmentalization and Viewpoint Generalization Within the Macaque Face-Processing System. *Science,* (6005).

Freiwald, W. A., Tsao, D. Y., & Livingstone, M. S. (2009). A face feature space in the macaque temporal lobe. *Nature Neuroscience,* (9).

Goddard, E., Mannion, D. J., Mcdonald, J. S., Solomon, S. G., & Clifford, C. W. (2011). Color responsiveness argues against a dorsal component of human V4. *Journal of Vision,* (4).

Granzier, J. J., & Gegenfurtner, K. R. (2012). Effects of Memory Colour on Colour Constancy for Unknown Coloured Objects. *I-Perception,* (3).

Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzchak, Y., & Malach, R. (1999). Differential Processing of Objects under Various Viewing Conditions in the Human Lateral Occipital Complex. *Neuron,* (1).

Hansen, T., Olkkonen, M., Walter, S., & Gegenfurtner, K. R. (2006). Memory modulates color appearance. *Nature Neuroscience,* (11).

Heywood, C. A., & Kentridge, R. W. (2003). Achromatopsia, color vision, and cortex. *Neurologic Clinics,* (2).

Hiramatsu, C., Goda, N., & Komatsu, H. (2011). Transformation from image-based to perceptual representation of materials along the human ventral visual pathway. *NeuroImage,* (2).

Humphrey, G. K., Goodale, M. A., Jakobson, L. S., & Servos, P. (1994). The Role of Surface Information in Object Recognition: Studies of a Visual Form Agnosic and Normal Subjects. *Perception,* (12).

Hung, C., Yen, C. C., Ciuchta, J. L., Papoti, D., Bock, N. A., Leopold, D. A., & Silva, A. C. (2015). Functional Mapping of Face-Selective Regions in the Extrastriate Visual Cortex of the Marmoset. *Journal of Neuroscience,* (3).

Hunt, R., Pitt, I., & Winter, L. (1974). The Preferred Reproduction of Blue Sky, Green Grass and Caucasian Skin in Colour Photography. *The Journal of Photographic Science,* (3).

Hurlbert, A. C., & Ling, Y. (2005). If it's a banana, it must be yellow: The role of memory colors in color constancy. *Journal of Vision, 5*(8), 787-787

Hurlbert, A. C. (2007) Colour constancy. *Current Biology*, 17, R906-907.

Issa, E. B., & Dicarlo, J. J. (2012). Precedence of the Eye Region in Neural Processing of Faces. *Journal of Neuroscience,* (47).

Komatsu, H., & Goda, N. (2018). Neural Mechanisms of Material Perception: Quest on Shitsukan. *Neuroscience.*

Kravitz, D. J., Saleem, K. S., Baker, C. I., Ungerleider, L. G., & Mishkin, M. (2013). The ventral visual pathway: An expanded neural framework for the processing of object quality. *Trends in Cognitive Sciences,* (1).

Kennard, C., Lawden, M., Morland, A.B., Ruddock, K.H. (1995). Colour identification and colour constancy are impaired in a patient with incomplete achromatopsia associated with prestriate cortical lesions. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 260*(1358).

Kentridge, R. W., Heywood, C. A., & Weiskrantz, L. (2007). Color contrast processing in human striate cortex. *Proceedings of the National Academy of Sciences, 104*(38).

Kuehni, R. G. (2015). How many object colors can we distinguish? *Color Research & Application,41*(5), 439-444. doi:10.1002/col.21980

Lafer-Sousa, R., & Conway, B. R. (2017). #TheDress: Categorical perception of an ambiguous color image. *Journal of Vision,* (12).

Lafer-Sousa, R., Hermann, K. L., & Conway, B. R. (2015). Striking individual differences in color perception uncovered by 'the dress' photograph. *Current Biology,* (13).

Maloney, L. T. (2002). Illuminant estimation as cue combination. *Journal of Vision,* (6).

Mapelli, D., & Behrmann, M. (1997). The role of color in object recognition: Evidence from visual agnosia. *Neurocase,* (4).

Mariam [MARIAM1A]. (2016, February 24). My jacket I bought is so cute. [Twitter moment]. Retrieved March 20, 2019, from https://twitter.com/MARIAM1A/status/702638409022504960

Mars, R. B., Jbabdi, S., Sallet, J., O'reilly, J. X., Croxson, P. L., Olivier, E., . . . Rushworth, M. F. (2011). Diffusion-Weighted Imaging Tractography-Based Parcellation of the Human Parietal Cortex and Comparison with Human and Macaque Resting-State Functional Connectivity. *Journal of Neuroscience,* (11).

Meyers, E. M., Borzello, M., Freiwald, W. A., & Tsao, D. (2015). Intelligent Information Loss: The Coding of Facial Identity, Head Pose, and Non-Face Information in the Macaque Face Patch System. *Journal of Neuroscience,* (18).

Miceli, G., Fouch, E., Capasso, R., Shelton, J. R., Tomaiuolo, F., & Caramazza, A. (2001). The dissociation of color from form and function knowledge. *Nature Neuroscience,* (6).

Mitterer, H., & Ruiter, J. P. (2008). Recalibrating Color Categories Using World Knowledge. *Psychological Science,* (7).

Naor-Raz, G., Tarr, M. J., & Kersten, D. (2003). Is Color an Intrinsic Property of Object Representation? *Perception,* (6).

Nijboer, T. C., Zandvoort, M. J., & Haan, E. H. (2006). Seeing red primes tomato: Evidence for comparable priming from colour and colour name primes to semantically related word targets. *Cognitive Processing,* (4).

Orban, G. A., Essen, D. V., & Vanduffel, W. (2004). Comparative mapping of higher visual areas in monkeys and humans. *Trends in Cognitive Sciences,* (7).

Orban, G. A., Zhu, Q., & Vanduffel, W. (2014). The transition in the ventral stream from feature to real-world entity representations. *Frontiers in Psychology.*

Passingham, R. E., Stephan, K. E., & Kötter, R. (2002). The anatomical basis of functional localization in the cortex. *Nature Reviews Neuroscience,* (8).

Pelphrey, K. A., & Wyk, B. C. (2011). Functional and Neural Mechanisms for Eye Gaze Processing. *Oxford Handbooks Online.*

Pelphrey, K. A., Singerman, J. D., Allison, T., & Mccarthy, G. (2003). Brain activation evoked by perception of gaze shifts: The influence of context. *Neuropsychologia,* (2).

Price, C. J., & Humphreys, G. W. (1989). The Effects of Surface Detail on Object Categorization and Naming. *The Quarterly Journal of Experimental Psychology Section A,* (4).

Rafaella [RafaellaWaldorf]. (2017, June 27). E rosa com branco, mas a luz da foto deixa o short meio verde/azul, sei la. [Twitter moment]. Retrieved March 20, 2019, from https://twitter.com/RafaellaWaldorf/status/879842369687998464

Rosenthal, I., Ratnasingam, S., Haile, T., Eastman, S., Fuller-Deets, J., & Conway, B. R. (2018). Color statistics of objects, and color tuning of object cortex in macaque monkey. *Journal of Vision,* (11).

Sawayama, M., Adelson, E. H., & Nishida, S. (2017). Visual wetness perception based on image color statistics. *Journal of Vision,* (5).

Schalk, G., Kapeller, C., Guger, C., Ogawa, H., Hiroshima, S., Lafer-Sousa, R., Saygin, Z.,M., Kamada, K., Kanwisher, N. (2017). Facephenes and rainbows: Causal evidence for functional and anatomical specificity of face and color processing in the human brain. *Proceedings of the National Academy of Sciences,* (46).

Sharan, L., Liu, C., Rosenholtz, R., & Adelson, E. H. (2013). Recognizing Materials Using Perceptually Inspired Features. *International Journal of Computer Vision,* (3).

Sharan, L., Rosenholtz, R., & Adelson, E. (2010). Material perception: What can you see in a brief glance? *Journal of Vision,* (8).

Simmons, W. K., Ramjee, V., Beauchamp, M. S., Mcrae, K., Martin, A., & Barsalou, L. W. (2007). A common neural substrate for perceiving and knowing about color. *Neuropsychologia,* (12).

Smet, K., Ryckaert, W. R., Pointer, M. R., Deconinck, G., & Hanselaer, P. (2011). Colour appearance rating of familiar real objects. *Color Research & Application,* (3).

Stasenko, A., Garcea, F. E., Dombovy, M., & Mahon, B. Z. (2014). When concepts lose their color: A case of object-color knowledge impairment. *Cortex.*

Steeves, J. K., Humphrey, G. K., Culham, J. C., Menon, R. S., Milner, A. D., & Goodale, M. A. (2004). Behavioral and Neuroimaging Evidence for a Contribution of Color and Texture Information to

Scene Classification in a Patient with Visual Form Agnosia. *Journal of Cognitive Neuroscience,* (6).

Tanaka, J. W., & Presnell, L. M. (1999). Color diagnosticity in object recognition. *Perception & Psychophysics,* (6).

Valenti, J., & Firestone, C. (2018). Finding the "odd one out": Memory color effects and the logic of appearance.

Vurro, M., Ling, Y., & Hurlbert, A. (2010). Memory colours of polychromatic objects. *Journal of Vision, 9*(8), 333-333.

Wada, A., Sakano, Y., & Ando, H. (2014). Human cortical areas involved in perception of surface glossiness. *NeuroImage.*

Wagner, H. G. (1965). Outlines of a Theory of the Light Sense. Ewald Hering , Leo M. Hurvich. *The Quarterly Review of Biology,* (4).

Winawer, J., & Witthoft, N. (2015). Human V4 and ventral occipital retinotopic maps. *Visual Neuroscience.*

Winkler, A. D., Spillmann, L., Werner, J. S., & Webster, M. A. (2015). Asymmetries in blue–yellow color perception and in the color of 'the dress'. *Current Biology, 25*(13).

Yang, J. N., & Maloney, L. T. (2001). Illuminant cues in surface color perception: Tests of three candidate cues. *Vision Research,* (20).

Yendrikhovskij, S. N., Blommaert, F. J., & Ridder, H. D. (1999). Color reproduction and the naturalness constraint. *Color Research & Application,* (1).

Yildirim, I., Belledonne, M., Freiwald, W., & Tenenbaum, J. (2018). Efficient inverse graphics in biological face processing.

Zaidi, Q. (2011). Visual inferences of material changes: Color as clue and distraction. *Wiley Interdisciplinary Reviews: Cognitive Science,* (6).