**maryelizabeth**.**sutherland**@us.nature.com

Telephone. +1 212 451 8575

Dear Dr. Sutherland,

We would like to submit for consideration at *Nature* an article reporting discoveries that bear directly on the long-standing question: are color categories in humans innate? Our work uses data from the macaque monkey, a model of the human that we use to unlock what has been a major barrier—the confound of language. Macaque monkeys have the same cone types as humans and very similar central circuits for vision, from retina through high-level regions of the ventral visual pathway. But since monkeys do not have language, data obtained with them are not confounded by the likely impact of language on categorization behavior. The idea that monkeys could be a key has been recognized for decades; the challenge has been to come up with a testing paradigm that itself does not use language or introduce other confounds such as inadvertently teaching or reinforcing idiosyncratic categories. Bae et al. {Bae, 2015 #18149} developed a non-verbal color-matching paradigm in humans that recovers color categories; we adapted this paradigm for use in monkeys. To ensure that the animals were not reinforced for idiosyncratic matches, every trial in our paradigm has four discrete possible matches of which only one is veridical (and rewarded). The errors are meaningful. To obtain enough of them for cross validation required hundreds of thousands of trials across four monkeys, which took almost five years of data collection. Using this paradigm, we discovered that monkeys have two consensus color categories, corresponding to “warm” and “cool” colors. We then developed a new analytic approach, an extension of the Target Confusability Competition (TCC) model of Schurgin et al {Schurgin, 2020 #17999}, to investigate the source of these categories: do they correspond to a nonlinear mapping of the stimulus space to the “true” perceptual space, or to an attractor within the space? We find that the consensus categories in monkeys are explained by a nonlinear mapping, one which likely impacts the interpretation of any study of human perception and cognition that uses color. Finally, we use these discoveries to empirically determine for the first time a perceptually uniform color space, one that is not confounded by language. The work is important not only because it provides important knowledge about how color works and could have practical application, but also because it answers fundamental questions about the origin of concepts.

Theories about color are among the oldest theories in psychology, dating to the ancient Greeks, and they have played a pivotal role in theories of the mind throughout history, including work by Descartes, Locke, Newton, Hume, and Wittgenstein; color continues to be a fertile testbed today (e.g., {Cohen, 2010 #18145}). Why has color been so important? Because color is not only a fundamental aspect of human experience (one many people would like to understand) but also a ready inroad into understanding the concepts more broadly. Human beings have a rich conceptual understanding of the world, a faculty that is thought to be a distinguishing feature of the human mind. Where do concepts come from? The question is at the heart of philosophy of mind and is increasingly a preoccupation of cognitive science {Wynn, 1992 #18661;Hespos, 2004 #18660;Tenenbaum, 2011 #17954;Jara-Ettinger, 2020 #18669}. The history of the question has taken its own fascinating turns. Key among them was the rise of the British Empiricists including John Locke whose writings had substantial influence on the US founding fathers. The distinctly American ideal of the “self-made man” has its origins in Locke’s idea that humans are innately endowed only with an ability to form sensory representations, what Locke called sensory “ideas”. In a view that resonates in contemporary scholarship, Locke asserted that the human mind at birth is a blank slate, a “tabula rasa”. Abstract (complex) concepts, he argued, are built up as people grow up, by stitching together disparate sensory (simple) ideas—hence the notion that man is “self-made”. This view is juxtaposed by a view traced to Descartes that humans have been gifted by evolution with innate mechanisms that compute from sensory data veridical representations of the world that embody information that is not in the sensory data. These innate mechanisms may, it has been argued, include representations of certain kinds of concepts. For example, Susan Carey has argued persuasively that the concept of “object” has arisen in evolutionary time so that newborns come equipped with representations of three-dimensional things that are independent of themselves {Carey, 2009 #18670}.

But as Carey points out, “whether there are innate conceptual representations is an empirical question” (Carey, pg. 33). So, what data are required? Again, Carey is instructive: “In order to study the origin of concepts, one must characterize their developmental and evolutionary trajectories, and to do that one must discover what kinds of mental representations and which specific concepts *nonhuman animals* and human infants and children have” (my emphasis)(Carey, pg. 5)…”we have no alternative but to do the hard empirical work” (Carey, pg. 29). Our paper takes up the challenge to collect the relevant data in nonhuman animals—and as Carey guessed, the work has not been trivial.

We exploit color as a paradigm, following in the footsteps of many others who have taken advantage of color to understand hallmarks of intelligent systems such as discrimination, perceptual constancy, memory, and communication (e.g., {Hurlbert, 1991 #4698;Webster, 1995 #8619;Bloj, 1999 #2801;Golz, 2002 #851; Fougnie, 2012 #18676; Lara, 2014 #9993;Zaslavsky, 2018 #16161;Schurgin, 2020 #17999;Panichello, 2021 #18671;Twomey, 2021 #18514}). The utility of color derives from the fact that it is a continuous variable with a well-specified physical basis (wavelength) that humans naturally categorize. In human culture, color categories become meaning-laden abstract concepts. The ripeness of a strawberry, the health of an infant, the passion of a valentine’s heart, all follow from concepts of redness. Indeed, it is the link between the color-state of objects and their behavioral relevance that earns color its vitality. Colors have other recognized features of concepts besides falling into discrete categories. For example, they involve social interactions among people (evident in color language); children learn colors relatively late (after shape), suggesting that color is a sophisticated representational system with expressive power; and many people develop complex intuitive theories about color. That color is conceptual is evident in the way people categorize each other by skin color, which often leads to (racist) judgements lacking any sensory evidence. The building block for color concepts is color categorization. Which raises the question: where do color categories come from?

Within color research, it is a long-standing open question whether humans come into the world with innate color categories. Some authorities contend we do (e.g. see {Lindsey, 2006 #8246;Lindsey, 2015 #8219;Skelton, 2017 #14994}; others argue we do not (e.g. {Davidoff, 1999 #6901;Block, 2023 #18672}). The tradition with the largest number of proponents is the view that color categories are innate. This idea was propelled by the influential work of Berlin and Kay who argued that all humans are innately endowed with the same set of “basic” color categories, and that the acquisition by languages of color terms for these categories is determined by a stereotyped sequence {Berlin, 1969 #1165}. Berlin and Kay were inspired by Ewald Hering, whose Opponent Colors Theory has remained for 150 years the accepted view of color appearance {Fairchild, 2015 #18520;Wolfe, 2022 #18569}. Opponent Colors Theory (OCT) states that all colors can be necessarily and sufficiently described by the extent to which they are reddish-versus-greenish, blueish-versus-yellowish, and blackish-versus-whiteish. The universality of this behavior, according to OCT, is attributed to hardwired innate neural mechanisms. The idea that color categories are innate is encouraged by research showing that 5-month-olds’ “categorical distinctions [align] with common distinctions in color lexicons and are organized around hues that are commonly central to lexical categories across languages” {Skelton, 2017 #14994}. Cross-cultural work on young children has been interpreted as showing support for “universalistic models of color categorization” because “color term knowledge does not modify categorical perception, at least during the early stages of childhood” {Franklin, 2005 #7003}. “These findings suggest that color categorization is partly organized and constrained by the biological mechanisms of color vision and not arbitrarily constructed by language” {Skelton, 2017 #14994}. Work on young humans involves heroic effort and has insurmountable limitations: a lot of data must be necessarily discarded by subjective criteria (“fussiness” or “failure to participate”), and it is typically only possible to assess behavior after several months of life, which brings confounds related to some cultural exposure, language, and learning. There remains no conclusive evidence from behavior or neurophysiology that color categories are innate, which leaves room for the less prominent position that color categories are learned {Davidoff, 1999 #6901;Roberson, 2000 #6899} (see review, Chapter {Block, 2023 #18672}. This argument adopts the view that color perception is shaped by language, e.g. through a Sapir-Whorfian mechanism {Goldstein, 2009 #6998}, and is supported by cross-cultural variability in color naming as well as perceptual learning studies {Ozgen, 2002 #18674}. But again, there is no conclusive evidence that categories are not innate. Ned Block aligns with Susan Carey in the view that the issue can only be addressed “by appeal to empirical evidence”. He also underscores the substantial obstacle to progress being a “lack of evidence that animals do not have the relevant concepts” (Block, pg. 266).

The paper not only uncovers color categories in monkeys, providing evidence of the language-independent color concepts of the trichromatic primate brain, but also shows how these color categories come about: through a nonlinear mapping of the stimulus space onto a truly uniform perceptual space represented in the observer’s mind. This discovery allowed us to reconstruct a perceptually uniform color space, unconfounded by language. These results will have direct relevance for the interpretation of all studies that have used color to study perceptual and cognitive operations, including the papers referenced above, for they show that the standard color space assumed to be uniform is, in fact, not uniform. The significance of this last point may not be immediately obvious, yet it is it crucial. As Newton showed, the spectrum is not a satisfactory representation of an objective color space because the spacing of the colors is not uniform and the rainbow does not include purples (hence the origin of the circular color wheel). Color societies such as the Commission Internationale de l’Eclairage (C.I.E.) have sponsored thousands of hours of research attempting to come up with a uniform color space, using similarity judgements made by human observers. As recently pointed out {Witzel, 2019 #18675}, this introduces a pernicious circularity: the similarity judgements (made by human observers) could likely be influenced by language since all human cultures have language. Our paper breaks the circularity.

We think the manuscript will be of broad interest, not just for cognitive scientists, but for neuroscientists, evolutionary biologists, anthropologists, linguists, computer scientists, philosophers, artists, and engineers. It takes up a question that has deep roots in scholarship about the mind within a domain—color—that is of profound intrinsic interest. The primary discovery of color categories in monkeys is clear and important, but the paper does not stop there. It shows how the categories come about and uses that explanation to generate a new color space that will have practical applicability in many fields, from industry to research on memory and attention.

Best wishes,