Color categories in macaque monkey

Daniel J. Garside, Audrey L.Y. Chang, Hannah M. Selwyn, Bevil R. Conway

Abstract

A distinctive ability of the mind is the use of conceptual representations. Are concepts innate or do they form through culture? Addressing the origin of concepts requires data in subjects who are not influenced by language, and an experimental approach. Here we exploit color as a tool of cognitive science—color is continuous, yet readily categorized, and color categories become meaning-laden concepts. For example, a ripe strawberry, an angry spouse, a valentine’s heart, all derive meaning from redness. We test whether color categories are innate by measuring behavior in macaque monkeys, a model of the human that lacks language but has essentially the same cone types and central visual circuits as humans. We adopted a match-to-sample paradigm used previously to uncover color categories in human subjects from subtle errors in matches. Using this method, macaques showed two consensus color categories. These do not correspond to basic color terms but rather to “warm” and “cool”. Using a novel extension of a recently developed analytic method, we determined that the consensus categories derive from a nonlinear mapping between the purportedly uniform color space of the stimuli and a true perceptually uniform color space. From the empirical results, we reverse-engineered a uniform perceptual space unconfounded by language. The results show that human color categories are likely not innate and provide evidence that purportedly uniform color spaces are inexorably shaped by the most elementary, universal, human categories of warm and cool.

# Introduction

Color categories are typically identified by color terms, of which the so-called Basic Color Terms are prominent {Berlin, 1969 #1165}. In English, these are black, white, red, green, yellow, blue, brown, orange, pink, purple, and grey. One view is that basic color categories are universal {Regier, 2005 #8218}, the brain is innately programed with concepts of Basic Colors, or a subset of them {Lindsey, 2006 #8246}. This idea is supported by modeling the commonalities in color naming across cultures {Jameson, 2009 #7488; Baronchelli, 2010 #8500; Lindsey, 2015 #8219; Abbott, 2016 #8474} and work in 5-month-old {Skelton, 2017 #14994} and 8-month-old infants {Ozturk, 2013 #6777}. Another view is that color categories emerge through the influence of language, culture, and behavior {Regier, 2009 #8250; Cibelli, 2016 #18503}, an idea promoted by variability in color naming patterns across languages {Paramei, 2018 #16202} and neurophysiological studies that fail to find clear evidence of neural correlates of basic color categories. One recognized approach to test these alternatives involves studies of trichromatic non-human primates. The gene sequence and spectral tuning of the pigments for the three classes of cones in these animals is very similar to that in humans. {Lindsey, 2009 #8224}

{Roberson, 2000 #6899;Davidoff, 1999 #6901}

{Goldstein, 2009 #6998;Roberson, 2008 #7588;Regier, 2007 #8489;Regier, 2005 #8218;Franklin, 2005 #7003;Kay, 2003 #16203;Levinson, 2000 #8498;Davies, 1998 #8247;Saunders, 1997 #9544;Zollinger, 1988 #8259}

{Zhou, 2010 #16276}

In pigeons {Wright, 1971 #18677} and bees {Kühn, 1927 #18678}

Color categorization appears to be a universal human phenomenon. Widespread variability in culture-specific color terminology suggests that color categories are linguistic in nature; however, fundamental similarities in color naming and color categorization across languages suggests that there may be some underlying structure which is universally inherent to human cognition and neurophysiology (Berlin and Kay 1991; Gibson et al. 2017). Studying categories in non-linguistic animals allows us to pick apart the relative contributions of language and innate factors. It has been previously shown that non-linguistic animals can be trained to perform tasks that involve the categorization of color (Sandell, Gross, and Bornstein 1979; Fagot et al. 2006). What has been unclear is whether these categories are used without explicit externally-motivated training.

We developed a method to test for evidence of categorical behavior without explicit categorical training, based on a task which has been extensively used in the investigation of working memory: a delayed forced-choice task. In this task, the participant is shown a colored circle on screen, which they remember the color of, and then after a delay, they select a circle of matching color from a set of differently colored circles.

This task has traditionally been used in working memory experiments because color was seen as a simple continuous scale with well-defined perceptual uniformity across the scale. Unfortunately, for working memory researchers, these assumptions have been shown to be ill-founded - (Bae et al. 2015) found that certain colors were remembered more accurately than others and that responses for certain colors were biased towards other colors. The pattern in responses could be accounted for by a model that encoded a memorized color in two distinct ways simultaneously - as a point on a continuous scale and *also* as a member of one of a number of categories that carved up colorspace. For humans they found that a four-category model fit their data well.

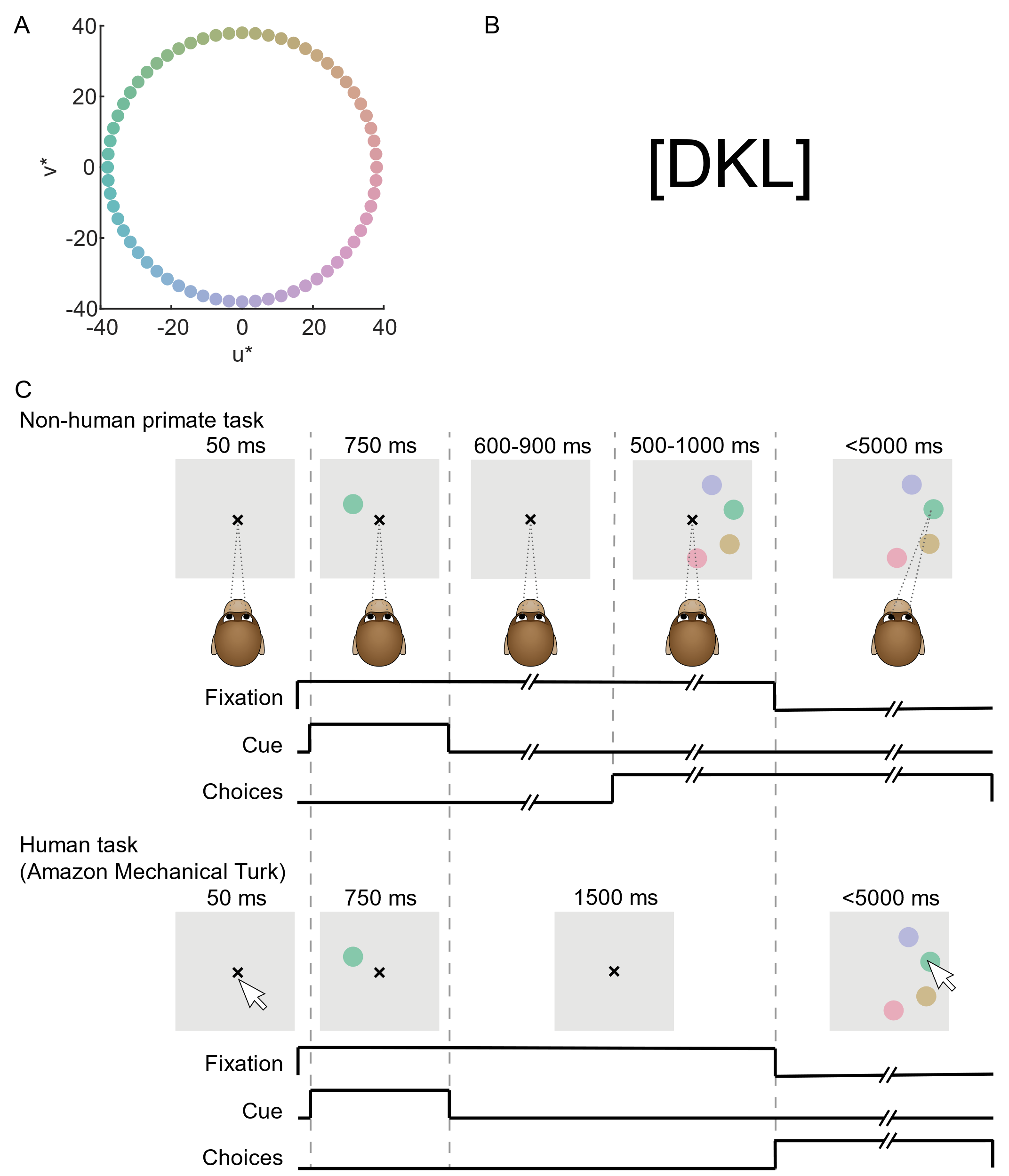
In seeking to explain this result, they showed that these categories correspond relatively well with categories identified in a separate experiment (with separate observers) to the categories recovered from people’s linguistic categories for color. This shows us that, at least at the population level, there is an apparent correspondence between the location of color category centers and the biases in response. It is possible to train macaques on a delayed forced-choice task such as this, thus we are able to use such a task, alongside the theoretical constructs of (Bae et al. 2015), to investigate innate use of color categories in macaques.

# Methods

## Subject details

Four male rhesus macaques (Macaca mulatta), weighing 8–10 kg , were implanted with headposts. Surgical implantation protocol has been described previously (Lafer-Sousa and Conway 2013). Designation of the subjects are PO, CA, BU, and MO. All procedures were approved by the Animal Care and Use Committee of the National Eye Institute and complied with the regulations of the National Institutes of Health.

## Behavioral tasks



**4-Alternative Forced Choice (4-AFC), Delayed Match to Sample Paradigm.** A. Cue and choice colors in the plane of CIELUV. Colors were defined to be equi-luminant and equi-saturated in CIELUV. B. Cue and choice colors in DKL colorspace. C. The timing and visual organization of the delayed match-to-sample task that the animals performed.

#### Stimuli

Stimuli were discs presented on a CRS Display++ screen. The color of stimuli varied only in hue, and were sampled from 64 equally spaced points on a circle in CIELUV space ([[fig:StimuliAndParadigm]](#fig:StimuliAndParadigm) ), with a white point of XXX (xy), a radius of 37, and a luminance of XXX (L\* = 76.0693). These values were chosen to maximize gamut while maintaining a fixed saturation and luminance. The background was XXX. CIELUV was used, in contrast to previous work which has used CIELAB, because CIELUV has the benefit of an associated chromaticity diagram. We also noted that nominally equi-saturated stimuli defined in CIELAB tended to have significant variation in apparent saturation, whereas the same in CIELUV were much closer to visually equi-saturated. Luminance noise was added by XXX to the extent of YYY.

The experiment was controlled by multiple computers running ‘Kofiko’ (a MATLAB/Psychtoolbox based software for working with monkeys ).

#### 4-Alternative Forced Choice (4-AFC): Non-human primates

Non-human primates were trained on a color-matching task. Trials begin with fixation (50 ms) on a white cross in the center of the screen. A cue stimulus (colored disc) is shown to one side of the fixation cross (750 ms). The position of the cue is invariant throughout a daily session. Following cue presentation, the monkey must maintain fixation (600-900 ms) before the choice stimuli appear on the screen alongside the fixation cross (500-1000 ms). The choices are positioned at constant eccentricity and with equal spacing in the hemifield opposite the cue stimulus, with the exact positions of the stimuli varying randomly trial-to-trial. One choice is always a direct match to the cue, and the other three are randomly sampled from the remaining stimuli, without replacement. Upon offset of the fixation cross, the animal makes a selection by saccade, and is rewarded for selecting the choice that is identical to the cue. Animals were head-fixed at a distance of XXX from the screen. Stimuli had a radius of XXX degrees of visual angle, at an eccentricity of XXX/degrees from a central fixation.

#### 4-Alternative Forced Choice: Human participants

Human participants were recruited via Amazon Mechanical Turk to perform an analogous version of the non-human primate 4-AFC task. Participants click on an initial fixation cross to request a trial, after which a cue is shown to one side of the fixation cross (750 ms). After cue offset, a fixation cross is shown and the cursor is hidden to de-incentivize mouse movement (1500 ms). Four choices are then shown, and participants make their selection by clicking.

## Data Analysis

#### Mixture Modeling

To assess the bias in responses for each cue, we computed the distribution of responses on trials where the monkey made an incorrect choice. For each completed trial, we calculated the error as the angular difference between the correct option and the chosen option. For each cue, we computed the number of times the monkey selected each incorrect choice, normalized by the number of times each choice color was available as a choice option for all completed trials of the given cue (though this was approximately uniformly distributed). We then fit a Gaussian with a variable floor ([[eq:GaussianEquation]](#eq:GaussianEquation)) to the error distribution for each cue, using the MATLAB fit function with the equation defined as a\*exp(-(((x-b)^2)/(2\*c^2)))+d.

$$\label{eq:GaussianEquation} \eqnmarkbox[purple]{p1}{a} \cdot \exp \frac{-(x- \eqnmarkbox[violet]{mu}{\mu} )^2}{2 \eqnmarkbox[blue]{sigma}{\sigma}^2} + \eqnmarkbox[gray]{d}{d}$$

This fit was weighted by the number of times each choice color was an option for the given cue across all completed trials (though as before, this was approximately uniformly distributed). Bias was taken as the difference between the cue and the peak of the corresponding Gaussian, for each cue color ( in [[eq:GaussianEquation]](#eq:GaussianEquation)). These values, for each stimulus, are plotted as the black lines in [[fig:BiasCurvesCombined]](#fig:BiasCurvesCombined). Where this line falls closer to the center of the figure than the 0 line, there is negative bias (which in this representation is counter-clockwise), and vice versa for values above the 0 line. Confidence intervals were extracted from

#### Target Confusability Competition (TCC)

One disadvantage of the mixture model for our analysis is that we can only use it to analyze the subset of trials where the animal made an incorrect response[[1]](#footnote-1). In order to use the full dataset (both incorrect and correct trials), we developed a generative model, based on the *Target Confusability Competition (TCC)* model of (Schurgin, Wixted, and Brady 2020). The key elements of the TCC model are a similarity function, which determines the similarity between stimulus and stimulus through a non-linear mapping of distance to similarity, and a value of , which can be thought of as describing the amount of noise acting on the system. Taken together, these two elements can be used to predict the probability that a choice of colour will be picked, from the set of , on a trial where the cue is . Our implementation of the model differs in some key ways to that described in (Schurgin, Wixted, and Brady 2020):

1. We do not assume that the underlying function is the same for each stimulus. (Schurgin, Wixted, and Brady 2020) collapse across stimuli for the majority of their analyses (though, see their Figures 1D and Extended Data Figure 5). Since we are most interested in the differences between the functions for the different stimuli, it does not make sense for us to collapse our data. We therefore deal with a “similarity matrix”, whereas (Schurgin, Wixted, and Brady 2020) could refer to their collapsed version as a “similarity function”.
2. We make no assumptions about the underlying function that determines similarity. (Schurgin, Wixted, and Brady 2020) use an exponential function with additional perceptual noise (see their Figure 1F), based on observations gained from collecting data on various simultaneous judgment tasks. We choose not to do this, because we expect that if biases are present, this would modify the shape of the function differently for each stimulus. We refer to fits made this way as “free similarity matrix” fits, since each elements of the matrix is “free” to float as it wishes, independently of those elements around it.
3. We fit our model on a single dataset, whereas (Schurgin, Wixted, and Brady 2020) derive their similarity functions and values for from independent datasets.
4. In fitting a “free similarity matrix” noise in the system can either be represented by the value of or in modifying the “contrast” of the similarity matrix (the relationship between the highest values and the lowest values in the matrix), since we apply no constraints on the floor or peak of the function. We therefore assume a value of for free similarity matrix fits.
5. Since we use an AFC method, as opposed to a pseudo-continuous response space, we are able to take advantage of an alternative computational method for computing the probabilities of a particular choice being made. We use the correction factors of (McGraw and Wong 1992) (their Table 3) to estimate the probability , where are samples from independent normal variables, with means representing the pairwise similarity values between and , and variances determined by . This decreases the runtime of our model by several orders of magnitude compared to the method used by (Schurgin, Wixted, and Brady 2020) (See the function modelPDF in TCC\_Code\_InManuscriptOrder\Model\TCCUncorrelated.m from <https://osf.io/j2h65/> for comparison).

Using this model, we used parameter estimation techniques to construct a similarity matrix where each cell represented the similarity between stimulus and stimulus . Such a similarity matrix is shown in [[fig:SimilarityMatrixPollux]](#fig:SimilarityMatrixPollux) for a single animal (Monkey P). This model, and this visualization method, allows us to assess not only the mean of the bias in the responses, but the shape of the response-bias-curve, which gives us an insight into the source of the bias.

#### Cognitive Bias vs. Stimulus-Space Non-Uniformity

For these experiments, we used a nominally perceptually-uniform colorspace: CIELUV. This space has been derived psychophysically, with the goal of minimizing differences in perceptual non-uniformity across the space, for color differences of small magnitudes (the apparent color difference between two points in one part of the space should be equal to the apparent color difference between two points in another part of the space when that the cartesian distance between the two points in each case be the same).

However, non-uniformities within the space are known to exist (ref?), and uniformity for small color differences does not necessarily assure uniformity for larger color differences (ref? Teunissen?). Likewise, uniformity for the conditions under which the psychometric measurements from which the space was determined (considering: spatial, temporal, spectral etc.) does not necessarily assure uniformity across all possible viewing conditions (ref).

With this in mind, it is reasonable to consider what the effect of residual non-uniformity might be on the results of our experiment. As discussed by (Panichello et al. 2019) (their Figure S5) non-uniformities in colorspace could also potentially lead to systematic biases on tasks such as ours. The logic goes as follows: our points are uniformly distributed in our chosen space ([[fig:StimuliAndParadigm]](#fig:StimuliAndParadigm)A), but if this space is actually non-uniform compared to the colorspace implicitly being used by an observer, then these same points will be *non*-uniformly distributed in a hypothetical ‘perfect colorspace’. It follows that for each cue color, surrounding distractor points might actually be closer or further away than anticipated. If the nearest neighbors on one side of the cue are actually chromatically closer than the neighbors on the other side, one would expect these to be chosen at a higher frequency than the others, creating a systematic bias. (Panichello et al. 2019) lack a robust framework within which to test these ideas, and conclude that there would be no effect on their conclusions. We reach a contrary conclusion.

It is difficult to distinguish biases from different sources using the Mixture Modeling approach...

This task becomes tractable in the TCC framework. (why?) A reasonable definition of cognitive bias might be: an agent is more likely to pick choice as a match to cue than they are to select choice as a match to cue , and that this behavior would not be reciprocal (they would not be more likely to pick choice as a match to cue than they are to select choice as a match to cue ). By this definition, this type of bias would appear as spread or displacement away from the negative diagonal in the similarity matrix, that *was not* symmetric across the negative diagonal (symmetry across the diagonal would represent reciprocity). There are numerous ways that cognitive bias can be envisaged/implemented. A cartoon example is shown in [[fig:JustBias]](#fig:JustBias).

Spread away from the negative diagonal which is mirrored across the diagonal represents non-uniformity of the stimulus space. In areas where the behavioral space is oversampled, one would see spread away from the negative diagonal (adjacent colors are more similar than the average). In areas that are undersampled, one would see a pinch into the negative diagonal (adjacent colors are less similar than the average). A cartoon example is shown in [[fig:JustColSpace]](#fig:JustColSpace).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| |  | | --- | | image | | |  | | --- | | image | | |  | | --- | | image | | |  | | --- | | image | |

**Distinguishing between different sources of bias using TCC models: cognitive bias vs. non-uniformity of stimulus space.** Similarity matrices representing different theoretically driven mechanisms can result in the same average bias value. A mixture model cannot distinguish between these different sources, whereas a TCC model readily can. [2](#fig:JustBias): an example of how cognitive bias might appear - each row of the matrix is shifted leftwards or rightwards. [3](#fig:JustColSpace): an example of how non-uniformity in stimulus space might appear - the similarity between each cue and its neighbors is increased or decreased, resulting in an expansion of the higher similarity region of the matrix symmetrically around the negative diagonal for colors which are more similar to their neighbors than average, and a contraction for colors that are less similar to their neighbors than average. [4](#fig:JustBias_subset): The values representing the similarity function for cue 20 (the row highlighted in red in [2](#fig:JustBias)), with the circular median shown as a vertical dashed line. [5](#fig:JustColSpace_subset): As in [4](#fig:JustBias_subset) but for [3](#fig:JustColSpace). Note how the circular median of both functions is 25.

In this example, it is possible to see how responses could be biased even in the absence of cognitive bias - consider cue 20, for example. The similarity of cue 20 to all the possible choices is represented by the row of elements at location 20. Tracking left to right from the y-axis, see how there is only a small area of similarity to the left of the negative diagonal and a larger area of similarity to the right of the diagonal. Although the most similar choice is still 20, there is a longer tail of the similarity distribution to the right than to the left, and thus responses will be shifted to higher values on average.

To estimate how much of the bias could be attributed to non-uniformity of colorspace, we fit an alternative version of the TCC model. In this version, we use a single similarity function, defined by [[eq:SimilarityFunction]](#eq:SimilarityFunction) (which is controlled by two parameters: and , which together control the slope of the function and the extent of the flat-top of the function at zero)[[2]](#footnote-2) and instead allow the stimuli chromaticities to float. This is akin to asking: what set of relationships between the stimuli in stimulus space can best explain the data we observe?

$$\label{eq:SimilarityFunction} \eqnmarkbox[purple]{explambda}{\exp(x\cdot\lambda)} \eqnmarkbox[cyan]{convolution}{\circledast} \eqnmarkbox[blue]{sigma}{\mathcal{N}(0,\sigma^2)}$$

#### A behaviorally-derived colorspace

Following the fitting of the model described above, the set of modified stimuli chromaticities can be extracted ([[fig:MACBEHspace]](#fig:MACBEHspace). These can be thought of as the chromaticity values for the stimuli that we used, no longer represented in CIELUV, but now in a behaviorally-derived color space instead. It is then possible to define a uniformly distributed set of colors in this new space, reparameterize them by their relationship to the original colors in the new space (e.g. new chromaticity is 60% of the angular distance on the path between stimulus and stimulus ), and then determine where this new set of colors would be in CIELUV ([[fig:UniformStimsInCIELUV]](#fig:UniformStimsInCIELUV)).

# Results

#### Monkeys exhibit biases

The animals examined show a hallmark of color categorization behavior: memory biases towards a set of particular points in a perceptually uniform colorspace. In [[fig:BiasCurvesCombined]](#fig:BiasCurvesCombined) it can be seen that the biases deviate substantially and systematically from zero, with the attractor points being found where the bias line crosses the zero line from positive to negative (going counter-clockwise). These points are highlighted with colored lines, with the filled areas around these lines showing the confidence intervals on these crossing points. Repeller points found where the line crosses the zero line from negative to positive.

![](data:application/pdf;base64,)

**Bias as a function of hue, for data collapsed over 4 animals.**

#### Shared color categories across monkeys

We see that all tested monkeys share two common attractor points ([[fig:BiasCurvesIndividual]](#fig:BiasCurvesIndividual)), which we interpret as evidence of two shared color categories: a warm/orange-ish category (between 0 and 45), and a cool/blue-ish category (between 180 and 225).

#### Individual differences between monkeys

In one animal [[fig:BiasCurvesCastor]](#fig:BiasCurvesCastor) we see evidence of additional categories: strong evidence for a greenish category and weak evidence for a purple category (the “strength" of a category can be gleaned from looking at the local gradient at the zero-crossing point)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| |  | | --- | | image | | |  | | --- | | image | | |  | | --- | | image | | |  | | --- | | image | |

**Bias curves for individual animals**

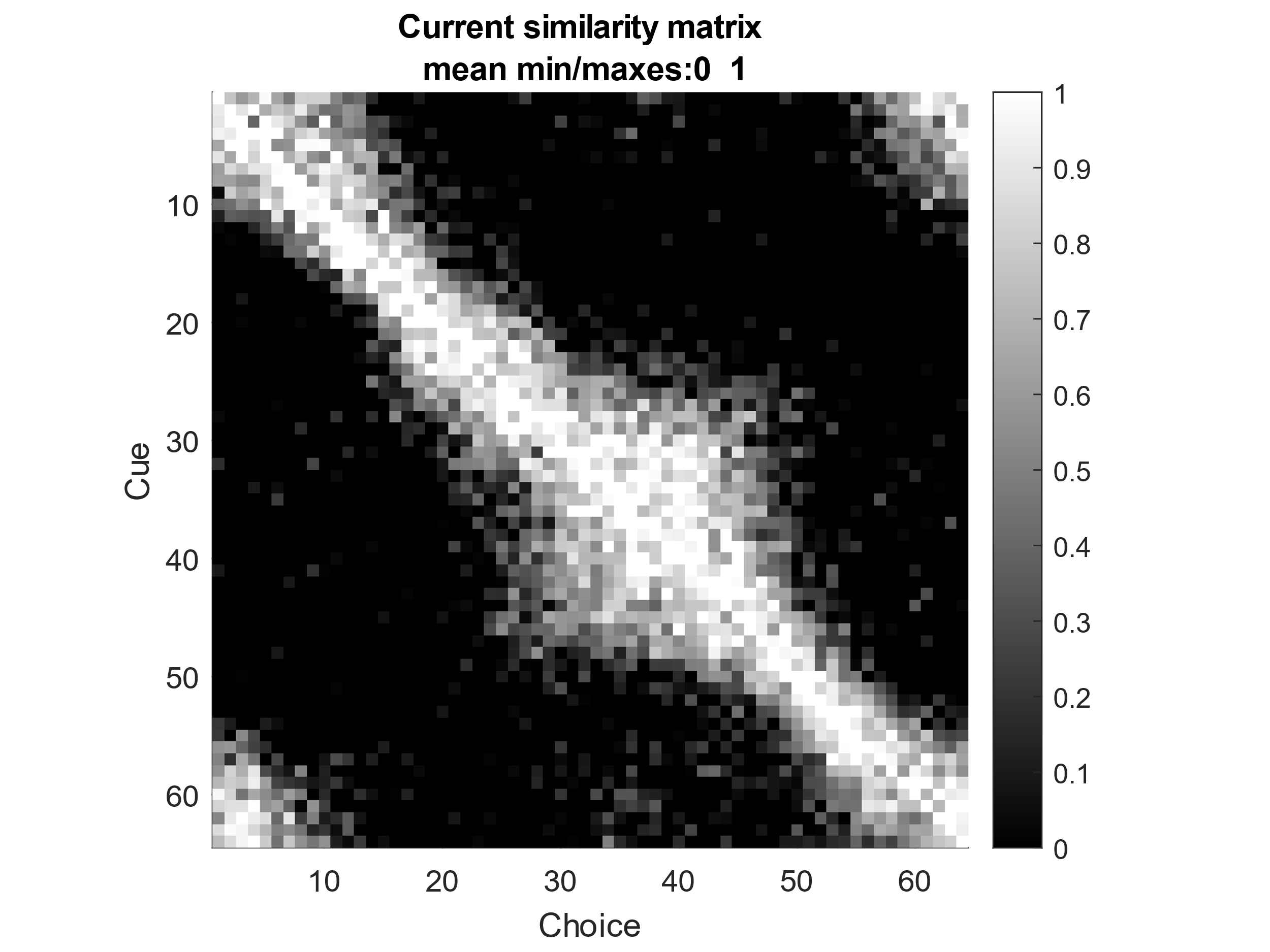
#### Biases Compared to Humans

Comparison to our human data, Mech Turk and also rig (maybe?)

Comparison to (Bae et al. 2015)

Comparison to (Panichello et al. 2019)

#### Cognitive Bias vs. Non-uniformity in Perceptual Space



**Free Similarity model fit for combined data from all animals** Similarity between stimulus and stimulus , where one is the cue and one is the choice. This is a “free” similarity matrix - in that no particular relationship is pre-supposed between any of the stimuli (such as, for example: closer stimuli will be more similar). This figure can be compared to Figure 1D in (Schurgin, Wixted, and Brady 2020), except there the rows are circularly shifted so that the the x-axis becomes the relative distance, rather than the absolute value of the stimulus.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| |  | | --- | | image | | |  | | --- | | image | | |  | | --- | | image | | |  | | --- | | image | |

**Free Similarity model fits for individual animals**

#### A behaviorally-derived colorspace

Extracted colorspace shown in [[fig:MACBEHcolorspace]](#fig:MACBEHcolorspace).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| |  | | --- | | image |   Stimuli in CIELUV | |  | | --- | | image |   Stimuli in behaviorally-derived space | |  | | --- | | image |   Stimuli sampled in behaviorally-derived space, projected back into CIELUV |

**A behaviorally-derived colorspace.** Ipsum!

#### Longitudinal analysis

Segmenting our data into subgroups of 5000 datapoints allowed us to both look at whether the determined categories varied over time, and also allowed us to perform a power analysis. From the monkeys studied it was clear that during our data collection period the categories remained static (within our measurement uncertainty), and also that the categories we saw are reliable enough to be seen with substantially less data. See Figures XXX

# Discussion

#### Where do categories come from?

In the absence of language, we can infer that the shared categories we observe arise either due to innate biological factors, environmental factors such as the distribution of colors in the terrestrial environment, or a combination of the two. The categories that we identify align well with the daylight locus (the line between the blue of the daytime sky, and the yellow of the sun, which itself closely follows the Planckian locus), and also the warm/cool object/background distinction previously identified (Rosenthal et al. 2018). It is plausible that what we observe is the presence of two fundamental categories - ‘likely to be an object of interest’ and ‘likely to *not* be an object of interest’.

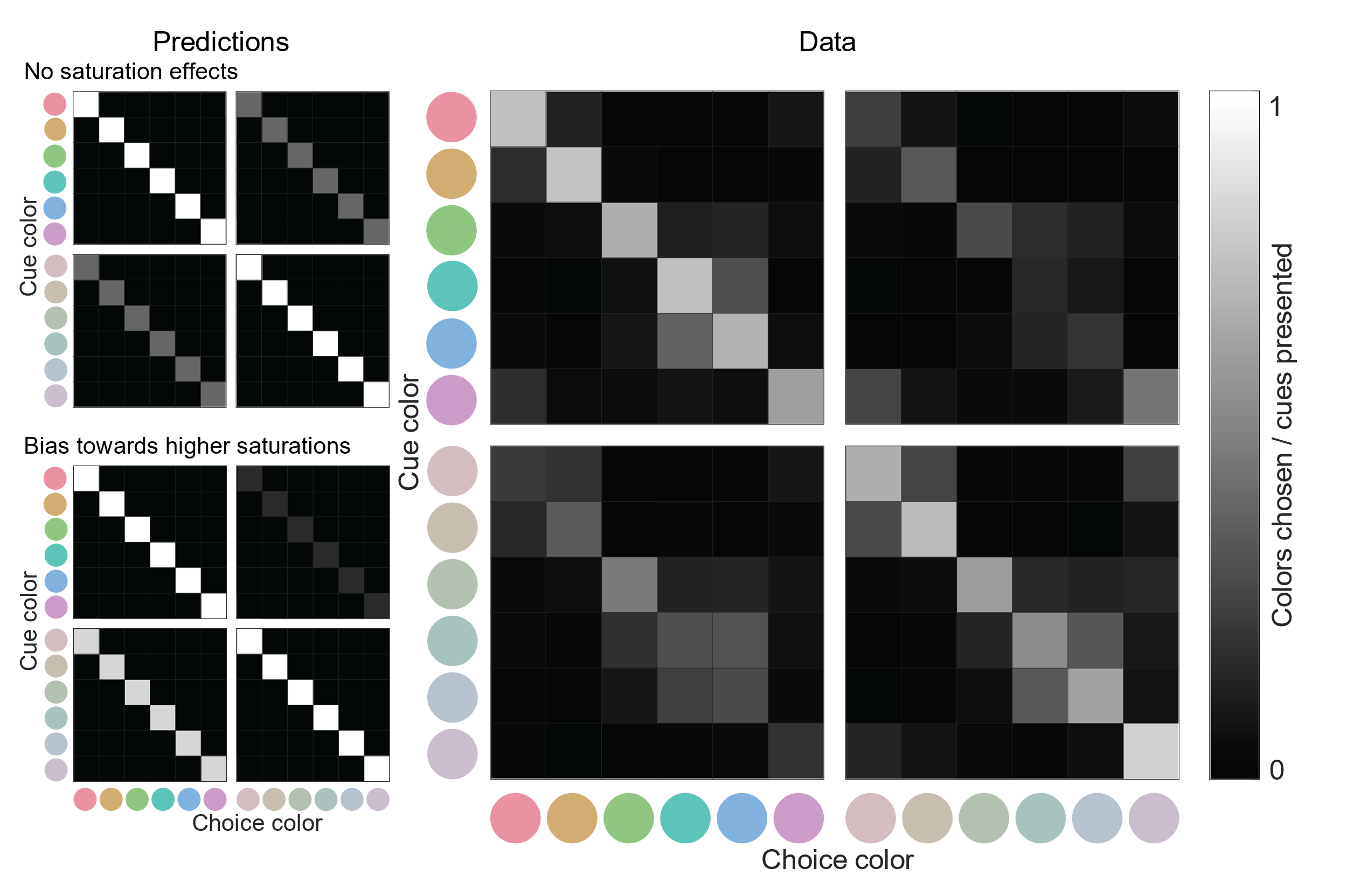
#### Different color spaces for different tasks, at different stages of the visual hierarchy

We have derived a behaviorally-derived colorspace, based on the data collected on our task, and this colorspace seems to differ meaningfully from CIELUV. There are multiple possibilities as to why this difference could occur. One potential source for this difference could be that multiple different colorspaces, optimised for different tasks, could exist at multiple different levels of the visual hierarchy. For example, it seems reasonable to say that both detection and memory are tasks that we would expect a visual system to perform, and yet color vision might require different forms to serve each of these functions. It is possible that this type of memory task recovers the colorspace that is used for short-term memory, where it makes sense to consider the trade off between accuracy and efficiency/stability (Panichello et al. 2019), and that the distinction with existing colorspaces, which are derived from data on simultaneous matches represents this distinction in task demands. Multiple levels of representation of color in the primate visual hierarchy presents an opportunity for those representations to be fruitfully different.

#### Language

#### Saturation bias

Non-uniformities in CIELUV may also plausibly result in our nominally iso-saturated colors actually being variably saturated. This would be a concern, as it would be a reasonable prediction that higher saturation colors would be more salient, and thus more likely to be selected as responses. We see no (or very little) bias towards higher saturation colors in a control experiment. In [[fig:saturationBias]](#fig:saturationBias) it can be seen that there are a reasonable number of errors where an animal picks a higher saturation version of the same hue (lower left quadrant). Still, it is also seen that the number of errors of the inverse type (upper right quadrant) is roughly equal in number.



**Saturation Bias.** Heatmap of cues and corresponding choices. Selections along the negative diagonal correspond to correct choices. Choices along the negative diagonal in the bottom left and top right quadrants show trials on which an incorrect choice was made in such a way that the hue was correct but the higher or lower saturation versions of the cue were chosen (respectively). Note: the main diagonal is expected to be filled in at a greater extent regardless of performance level since the correct choice is shown on every trial, whereas only a subset of the incorrect choices are shown.

#### Comparison with humans

# Limitations

?

# Conclusion

# Data availability statement

Data is available at: zenodo(?)...

# Code availability statement

Code as used in this paper: doi... Latest version of code: github...

# Acknowledgments

# Author contributions

Bae, Gi-Yeul, Maria Olkkonen, Sarah R. Allred, and Jonathan I. Flombaum. 2015. “Why Some Colors Appear More Memorable Than Others: A Model Combining Categories and Particulars in Color Working Memory.” *Journal of Experimental Psychology: General* 144 (4): 744–63. <https://doi.org/10.1037/xge0000076>.

Berlin, Brent, and Paul Kay. 1991. *Basic Color Terms: Their Universality and Evolution*. University of California Press.

Fagot, Joël, Julie Goldstein, Jules Davidoff, and Alan Pickering. 2006. “Cross-Species Differences in Color Categorization.” *Psychonomic Bulletin & Review* 13 (2): 275–80. <https://doi.org/10.3758/BF03193843>.

Gibson, Edward, Richard Futrell, Julian Jara-Ettinger, Kyle Mahowald, Leon Bergen, Sivalogeswaran Ratnasingam, Mitchell Gibson, Steven T. Piantadosi, and Bevil R. Conway. 2017. “Color Naming Across Languages Reflects Color Use.” *Proceedings of the National Academy of Sciences* 114 (40): 10785–90. <https://doi.org/10.1073/pnas.1619666114>.

Lafer-Sousa, Rosa, and Bevil R. Conway. 2013. “Parallel, Multi-Stage Processing of Colors, Faces and Shapes in Macaque Inferior Temporal Cortex.” *Nature Neuroscience* 16 (12): 1870–78. <https://doi.org/10.1038/nn.3555>.

McGraw, Kenneth O., and S. P. Wong. 1992. “A Common Language Effect Size Statistic.” *Psychological Bulletin* 111 (2): 361. <https://doi.org/10.1037/0033-2909.111.2.361>.

Panichello, Matthew F., Brian DePasquale, Jonathan W. Pillow, and Timothy J. Buschman. 2019. “Error-Correcting Dynamics in Visual Working Memory.” *Nature Communications* 10 (1): 3366. <https://doi.org/10.1038/s41467-019-11298-3>.

Rosenthal, Isabelle, Sivalogeswaran Ratnasingam, Theodros Haile, Serena Eastman, Josh Fuller-Deets, and Bevil R. Conway. 2018. “Color Statistics of Objects, and Color Tuning of Object Cortex in Macaque Monkey.” *Journal of Vision* 18 (11): 1–1. <https://doi.org/10.1167/18.11.1>.

Sandell, Julia H., Charles G. Gross, and Marc H. Bornstein. 1979. “Color Categories in Macaques.” *Journal of Comparative and Physiological Psychology* 93 (4): 626–35. <https://doi.org/10.1037/h0077594>.

Schurgin, Mark W., John T. Wixted, and Timothy F. Brady. 2020. “Psychophysical Scaling Reveals a Unified Theory of Visual Memory Strength.” *Nature Human Behaviour* 4 (11): 1156–72. <https://doi.org/10.1038/s41562-020-00938-0>.

1. This is for messy and annoying reasons. Firstly, since in our paradigm the choices consist of “the correct choice” plus three “distractors” (see ), the correct choices are greatly over-represented. Imagine: by guessing at random, the correct choice would be picked far more frequently than any of the other stimuli, since it is presented on *every trial*, whereas the other stimuli are only presented as distractors with a probability of . This could be normalized out, as we do for the other values, but then a more insidious issue becomes apparent: For “high-error” choices, the odds of there being a similar choice to that one is less than the odds of there being a similar choice to the “0-error” choice, and thus the probability of selecting the correct answer, when normalised, is lower than one would expect. You can think of it as: for the high-error choice, there are *2 chances* to pick another choice option that is close to the high-error choice (since one choice is going to be the distant-by-definition 0-error choice), whereas for the 0-error choice there are *3 chances*. An additional note for clarity: this is an issue for us because of the sampling required for the AFC paradigm, and is not an issue of concern for those who use a response mechanism where all possible responses are simultaneously presented. [↑](#footnote-ref-1)
2. These two parameters can be though of theoretically as the *similarity function* (how similar is to ), and the *perceptual function* (at what point do stimuli become indistinguishable from one-another). Unfortunately, in this parameterization of the function, the parameters are highly correlated, which makes recovery of these values via model fitting rather difficult. An alternative parameterization where the parameters were maximally uncorrelated would be preferable. [↑](#footnote-ref-2)