PNAS RESEARCH ARTICLE | PSYCHOLOGICAL AND COGNITIVE SCIENCES NEUROSCIENCE OPEN ACCESS

Perceptual and neural constraints on photometric measures of heterochromatic brightness

Shuchen Guan[a,1], Jing Chen[b,1], Robert Ennis[a], Matteo Toscani[c], Matteo Valsecchi[d], Andrea van Doorn[a,e], Jan Koenderink[a,e,f], and Karl R. Gegenfurtner[a,2]

Affiliations are included on p. 7.

Edited by Martin S. Banks, University of California, Berkeley, CA; received March 24, 2026; accepted July 23, 2026

Luminance and the candela provide the foundation of photometry and define standard measures of the visual impact of light. They are defined by classical photometric methods that fail to capture key aspects of brightness judgments across different colors in spatially extended, real-world scenes: Highly saturated colors appear brighter, mixtures are subadditive relative to luminance, and short wavelengths contribute more strongly than expected. Despite these long-standing discrepancies, progress has been limited by the lack of scalable methods to measure heterochromatic brightness under steady viewing. Here, we introduce a ranking-based paradigm that enables reliable measurement of brightness across large stimulus sets. Observers ranked 144 colored stimuli spanning hue and intensity, yielding highly consistent results across sessions, observers, and display environments, including large-scale online studies. This approach reveals a stable global structure of heterochromatic brightness preserved across devices and viewing contexts. Using these data, we evaluate a broad range of candidate models. Luminance, radiance, and established color appearance models leave key systematic patterns unexplained. In contrast, simple nonlinear pooling rules, specifically taking the weighted maximum of the red, green, and blue channel values, provide a near-ceiling account of observed rankings. Neural recordings reveal a corresponding temporal dissociation: High-frequency responses track luminance, whereas low-frequency responses align with nonlinear predictions. Experiments with spectrally tunable illumination confirm the same pattern in spatially extended scenes. Together, these results establish a scalable framework for measuring heterochromatic brightness and identify a simple computational rule governing brightness under steady viewing.

luminance | photometric standards | color | heterochromatic brightness | lighting

Significance

Luminance, defined by the candela, is the international standard for quantifying the visual impact of light and underpins lighting, imaging, and display technologies. Yet it is mainly based on flicker measurements that do not reflect how we typically view light in steady scenes. Although deviations from luminance have long been recognized, the absence of reliable, scalable measurements has prevented viable alternatives. Here, we introduce a framework, where observers rank the brightness of hundreds of colored patches, and show that brightness under steady viewing is best predicted by taking the weighted maximum of red, green, and blue channels. This result identifies a fundamental limitation of current photometric standards and provides a principled basis for predicting perceived brightness in complex lighting environments.

Artificial lighting accounts for a substantial fraction of global electricity consumption (1, 2). Its efficiency is evaluated using photometric quantities derived from the candela, the only SI base unit defined through a behavioral percept (3, 4). Luminous intensity is computed by weighting a light source's spectrum with the CIE luminous efficiency function V(λ), derived from a combination of psychophysical methods including heterochromatic flicker photometry and step-by-step brightness matching between two adjacent fields sharing a common border (5–7). In flicker photometry, two differently colored lights are alternated rapidly until temporal modulation disappears, isolating a transient luminance-cancellation signal that yields additive measurements: The luminance of a mixture equals the sum of its components (8, 9). Additivity is therefore embedded in the definition of the candela and underlies modern photometric standards. Under typical viewing conditions, however, light sources and illuminated surfaces are spatially extended and temporally stable rather than restricted to adjacent fields or rapidly alternating stimuli. Under these conditions, perceived heterochromatic brightness departs systematically from additive luminance (*SI Appendix*, Fig. S1): Saturated colors appear brighter than predicted, mixtures are subadditive, and short-wavelength stimuli contribute more strongly than allowed by V(λ) (10, 11). With spectrally tunable LED lighting now widely used (12), these discrepancies reveal a growing mismatch between photometric standards and perceived brightness.

Such deviations have long been recognized. Early studies reported failures of additivity (13–17), and the Helmholtz–Kohlrausch effect showed that chromatic purity enhances perceived brightness beyond luminance predictions (18, 19). Color appearance models introduced nonlinear corrections (20, 21), yet retained luminance as their underlying reference. Overall, progress was hindered by the difficulty of obtaining reliable measurements and the absence of a systematic model comparison. Our own prior work explored brightness judgments across a variety of tasks and identified nonlinear channel contributions as an important determinant of perceived brightness (22). Here, we revisit this problem using a

Author contributions: S.G., J.C., R.E., M.T., M.V., A.v.D, J.K., and K.R.G. designed research; S.G. and J.C. performed research; S.G., J.C., R.E., M.T., and K.R.G. analyzed data; and S.G., J.C., and K.R.G. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]S.G. and J.C. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: gegenfurtner@uni-giessen.de.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2610398123/-/DCSupplemental.

Published August 19, 2026.

scalable ranking paradigm that allows observers to order large sets of colored stimuli by brightness. The resulting measurements are highly reliable across laboratory and large-scale online experiments. We show that steady brightness systematically departs from additive luminance and is instead accurately predicted by nonlinear pooling across color channels. Neural recordings reveal a corresponding dissociation between luminance-driven responses at high temporal frequencies and nonlinear brightness signals under slower stimulation, and experiments with spectrally tunable lighting confirm the same relationship in spatially extended real-world environments.

Results

Behavioral Ranking Task. A prerequisite for reevaluating luminous intensity is a measurement that is both reliable and scalable. Observers ranked 144 colored stimuli according to perceived brightness using a sequential elimination ranking task (Fig. 1*A*). In each trial, participants selected the brightest stimulus from a subset until a complete ordering was obtained. This procedure yields a full ranking while keeping the number of decisions manageable. The resulting mean rank structure across stimuli and participants is shown in Fig. 1*B*. Systematic variations are evident across hue and stimulus intensity, indicating that steady heterochromatic brightness exhibits a well-defined global structure rather than random variability across stimuli. The stability of this structure is illustrated in Fig. 1*C*. In a subset of observers retested more than 6 mo later, mean rankings were nearly identical, with a Spearman correlation of 0.99 between sessions. This high test–retest reliability indicates that the ranking procedure captures a stable perceptual ordering rather than transient decision strategies or contextual effects.

To evaluate whether such measurements can be obtained at large scale outside the laboratory, we next asked whether the ranking task itself produces stable results across different displays. Unlike classical brightness matching procedures, the elimination ranking task depends primarily on ordinal comparisons among stimuli and is therefore less sensitive to absolute calibration differences. After measuring the spectral characteristics of the same observers' personal displays, we evaluated how the observed deviations from the laboratory display affected the stimulus set and found that the resulting rank-order information was largely preserved (*SI Appendix*, Fig. S2 *A*, *C*, and *E*). When the laboratory observers repeated the ranking task at home using their own monitors (Fig. 1*D*), the resulting brightness orderings closely matched those obtained in the laboratory, despite differences in presentation devices. This robustness enabled large-scale online testing. In the main black-background online sample, an additional 182 participants completed the task remotely without explicit control of displays and viewing conditions (*SI Appendix*, Fig. S3*A*). Despite substantial device variability and the absence of calibration, the average rankings closely reproduced the laboratory results, with correlations above 0.98. Together, these results show that steady brightness judgments obtained with the ranking paradigm are reproducible across observers, devices, and environments, allowing heterochromatic brightness to be measured at a scale that was previously difficult to achieve with traditional methods.

This large and reliable dataset allows a direct empirical test of a central assumption of photometry: that luminous intensity should predict perceived heterochromatic brightness under steady viewing. If luminous intensity were additive under steady viewing, luminance should accurately predict these rankings. It does not. When luminance values were compared with the empirical brightness ranks, more than one third of the variance remained unexplained (Fig. 1*E*). A purely physical measure of spectral energy, radiance, performed somewhat better than luminance but still left substantial variance unexplained (Fig. 1*F*). The widely used color appearance model CAM16 also performed worse than radiance and also failed to capture the global rank structure of our dataset (Fig. 1*G*). These results indicate that steady brightness perception cannot be captured by linear spectral weighting and is not adequately described by standard color appearance models. In contrast, a nonlinear channel-based prediction provided a near-ceiling account of the observed rankings (Fig. 1*H*). This

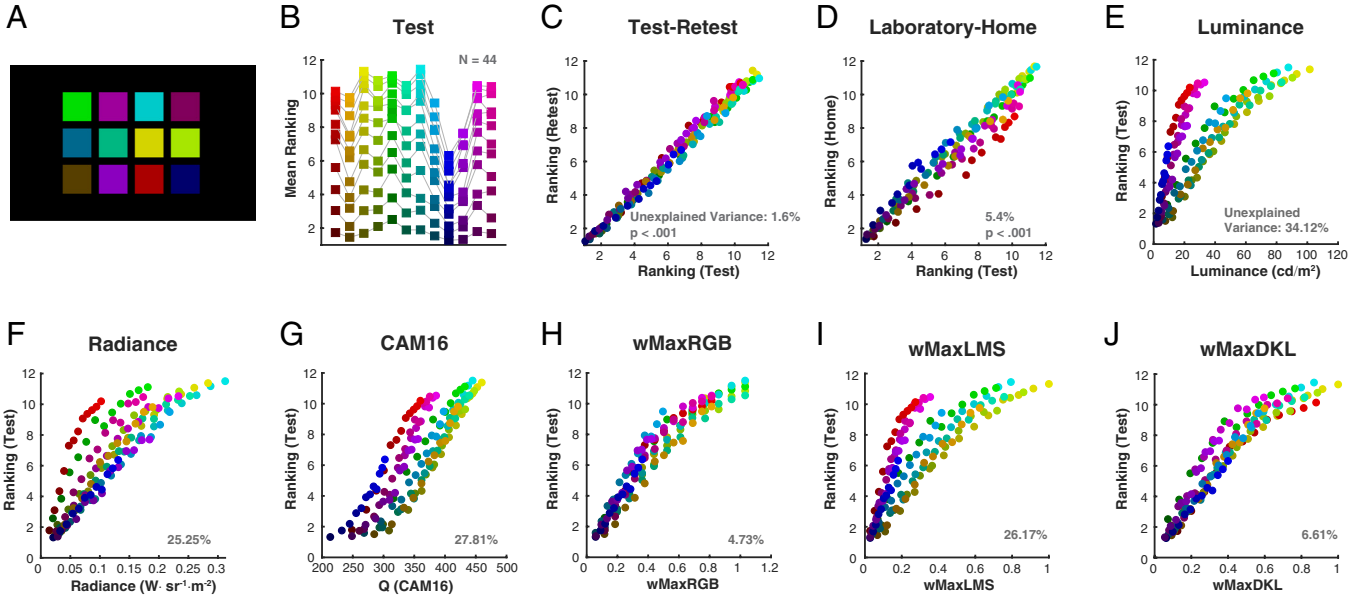

Fig. 1. Reliable large-scale measurement of heterochromatic brightness reveals failure of additive photometry under steady viewing. (*A*) Schematic of the ranking paradigm. On each trial, observers ranked 12 patches from the pool of 144 patches spanning 12 hues and 12 intensity levels defined in RGB space. (*B*) Mean brightness rankings in the laboratory sample. (*C*) The test–retest correlation for a subset of observers measured more than 6 mo apart. Each point represents one stimulus. (*D*) Correlations between laboratory vs. home testing. (*E*) Prediction of empirical brightness rankings by luminance, (*F*) radiance, (*G*) the CIE CAM16 model and (*H*) nonlinear weighted maximum models in RGB, (*I*) LMS cone, and (*J*) DKL spaces.

computation reduced unexplained variance to below five percent and reproduced the global rank structure observed across laboratory and online samples.

These results show that steady heterochromatic brightness cannot in general be predicted from additive luminance. We therefore asked how signals from different color channels should be combined to account for the observed rank structure. To address this question, we evaluated a generalized Minkowski pooling rule in which weighted responses from the three display primaries are combined with exponent n. The weights allow the contributions of the red, green, and blue channels to differ, while the exponent controls the degree of nonlinearity in the pooling operation. When $n = 1$, the model reduces to linear summation, equivalent to additive photometry. As n increases, the pooling rule becomes increasingly nonlinear and approaches a maximum operation.

Model performance improved markedly as the pooling exponent increased and approached an asymptote near $n = 2$ (Fig. 2*A*). Within this nonlinear regime, the fitted channel weights converged on a well-defined optimum (Fig. 2*B*). This optimal weighting differed strongly from the weighting implied by luminance, indicating that steady brightness cannot be explained by additive photometry even after allowing arbitrary channel weights. Across the broader model comparison (Fig. 2*C* and *SI Appendix*, Fig. S3), the nonlinear pooling model provided a near-ceiling account of the data, approaching the test–retest reliability limit. In contrast, luminance and most traditional color appearance models performed poorly, leaving substantial unexplained variance and, in many cases, performing worse than radiance. More recent models that incorporate explicit hue weighting and the Helmholtz–Kohlrausch effect showed improved performance, but still failed to capture the global rank structure. A linear model with optimized channel weights (wSumRGB, *SI Appendix*, Fig. S3*B*) performed substantially better than these approaches, indicating that part of the failure of luminance arises from inappropriate channel weighting. However, even this optimized linear model fell far short of the nonlinear pooling rule, demonstrating that the primary limitation of existing approaches is the assumption of additivity itself.

To examine whether such computations might arise from known visual representations, we next evaluated models based on nonlinear pooling in physiologically motivated color channels. Nonlinear pooling over cone signals (maxLMS) provided only modest improvements over luminance (Fig. 1*I*). This is expected: L and M cones overlap heavily in their spectral sensitivities due to their recent evolutionary divergence (23, 24), making them poorly suited as inputs to a max operation. Pooling in a rectified opponent color space (wMaxDKL) performed substantially better (Figs. 1*J* and 2*C*), consistent with the idea that brightness signals may arise after chromatic recombination into more independent channels. That RGB outperforms wMaxDKL reflects the fact that display primaries are near-optimal for spanning the gamut of natural object colors (25) and are therefore well matched to the

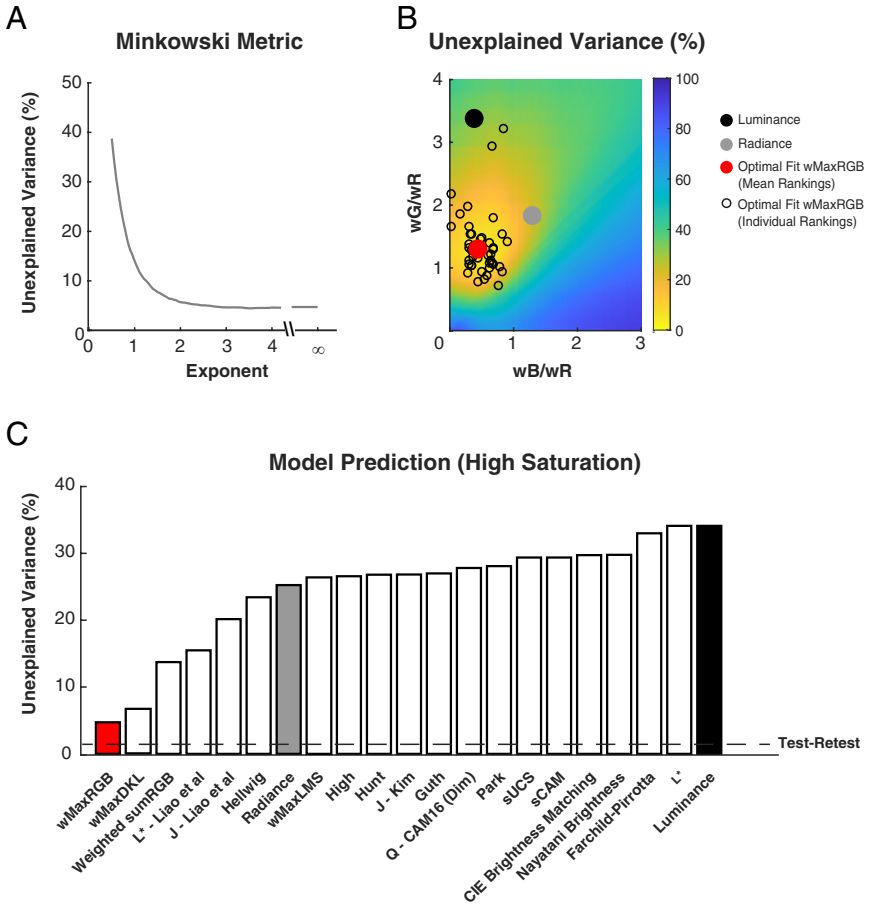

Fig. 2. Nonlinear pooling over color channels accounts for steady heterochromatic brightness. (*A*) Model performance as a function of Minkowski exponent n for the high saturation stimulus set. (*B*) Weight optimization landscape for nonlinear pooling in the high saturation dataset. The heatmap shows unexplained variance across relative weight combinations for the three primary channels. The red marker indicates the optimal weights for fitting the mean ranking data, while the black and gray markers represent the luminance and radiance weights of the display primaries, respectively. The open circles indicate the optimal weights obtained from fits to the individual observers' ranking data, illustrating stable individual variation around the population optimum. (*C*) Comparison of representative model classes for the high saturation dataset. The dashed line shows the test–retest reliability (same in Fig. 1*C*).

environmental statistics that shape the representation of color in the visual cortex.

These results converge on a simple computational rule for steady heterochromatic brightness. Responses in the three primary channels are first scaled by channel-specific weights and then combined through a strongly nonlinear, near-maximum pooling operation. The optimal wMaxRGB weights [1, 1.3, 0.44] assigned comparable contributions to the red and green channels, while giving the blue channel substantially greater influence than implied by luminance weighting. In the wMaxDKL model, the weights were [1, 0.88, 0.40] for L + M, L − M, and S − (L + M) and similarly indicate a substantial contribution of S-cones. This pattern closely mirrors the empirical rank structure, in which short-wavelength stimuli appear systematically brighter than predicted by additive photometry. For simplicity, we refer to this compact approximation as wMaxRGB. It provides a compact description of steady brightness judgments across the stimulus set.

To test whether these findings generalize beyond highly saturated stimuli, we expanded the stimulus set to include multiple chroma levels (*SI Appendix*, Fig. S4*A*). Although the optimal channel weights shifted somewhat across chroma (*SI Appendix*, Fig. S4*B*), the qualitative ranking of model classes remained unchanged. Nonlinear pooling continued to outperform additive photometric models across stimulus sets (*SI Appendix*, Figs. S5–S7). We also examined whether contextual factors influenced the inferred model parameters. Separate online cohorts performed the ranking task on displays with different backgrounds (black, gray, and white). This had little effect on the estimated channel weights (*SI Appendix*, Fig. S10). This indicates that the failure of additivity is robust across stimulus conditions and viewing contexts and that steady brightness is consistently best described by nonlinear pooling over color channels.

These behavioral results therefore indicate that steady heterochromatic brightness is best described by nonlinear pooling across color channels rather than by additive luminance. An important question is whether this distinction is also reflected in neural signals. Classical photometry relies on flicker-based measurements that are known to engage luminance-sensitive pathways, whereas the present ranking task probes sustained brightness judgments under steady viewing. We therefore asked whether neural responses measured at different temporal frequencies reflect these two computational regimes.

Neural Basis. The behavioral results revealed a clear dissociation between flicker-based additivity and steady perceptual brightness, suggesting that distinct neural mechanisms may underlie these signals. We therefore recorded steady-state visual evoked potentials (SSVEP, *SI Appendix*, Fig. S8 *B* and *C*) while observers viewed color discs flickering either at 15 Hz or at 3 Hz (*SI Appendix*, Fig. S8 *A* and *D*). The higher frequency approximates the temporal regime of heterochromatic flicker photometry, whereas the lower frequency more closely resembles steady viewing conditions.

At 15 Hz, SSVEP amplitudes closely tracked stimulus luminance (Fig. 3*A*). Across stimulus conditions, responses were strongly correlated with luminance values, consistent with classical flicker-based measurements and their physiological basis in luminance-sensitive pathways (8, 26, 27). The MaxRGB model failed to predict the 15 Hz SSVEP (*SI Appendix*, Fig. S8*F*). In contrast, at 3 Hz the relationship between SSVEP amplitude and luminance was markedly weaker (Fig. 3*B*). Luminance accounted for only a limited fraction of response variance at this lower temporal frequency.

When 3 Hz SSVEP amplitudes were compared with predictions from nonlinear pooling models, the correspondence was substantially stronger (Fig. 3*C*). Thus, high-frequency neural responses are consistent with additive luminance mechanisms, whereas low-frequency responses align more closely with the nonlinear MaxRGB pooling rule that predicts steady heterochromatic brightness. This frequency-dependent dissociation mirrors the distinction between flicker photometry and sustained perceptual judgments observed in the behavioral experiments.

Real-World Lighting. To determine whether the dissociation between luminance and steady brightness extends beyond display-based stimuli, we tested brightness judgments under spectrally tunable LED illumination in a controlled room environment (*SI Appendix*, Fig. S9 *A* and *B*). Two spatially separated light sources were adjusted to create conflicts between competing brightness predictors while maintaining comparable spatial layout and overall scene structure.

When luminance and radiance were placed in opposition (Fig. 4*A*), a majority of observers selected the higher-radiance condition as brighter (Fig. 4*D*). When luminance conflicted with the nonlinear pooling model derived from the psychophysical data (wMaxRGB, Fig. 4*B*), observers predominantly chose the condition favored by the nonlinear model (Fig. 4*E*). Across participants, the proportion of choices consistent with the nonlinear model

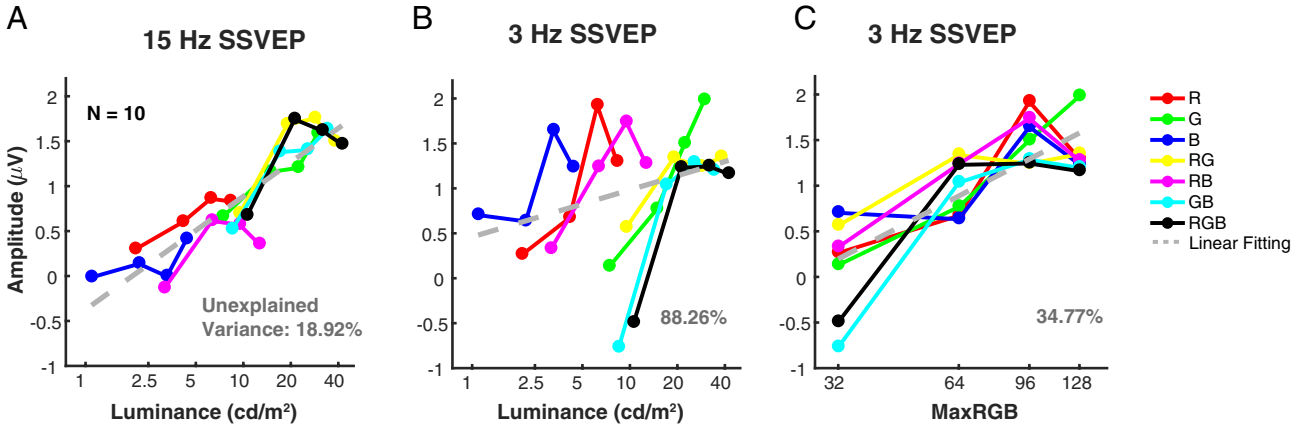

Fig. 3. Frequency-dependent neural responses dissociate luminance from steady brightness. (*A*) SSVEP amplitude at 15 Hz plotted as a function of stimulus luminance. Neural responses closely follow luminance predictions, consistent with classical flicker based measurements. (*B*) SSVEP amplitude at 3 Hz plotted as a function of stimulus luminance. (*C*) SSVEP amplitude at 3 Hz plotted as a function of nonlinear pooling model predictions. Each point represents one stimulus condition averaged across observers. Details of the SSVEP amplitude calculations can be found in *Materials and Methods*.

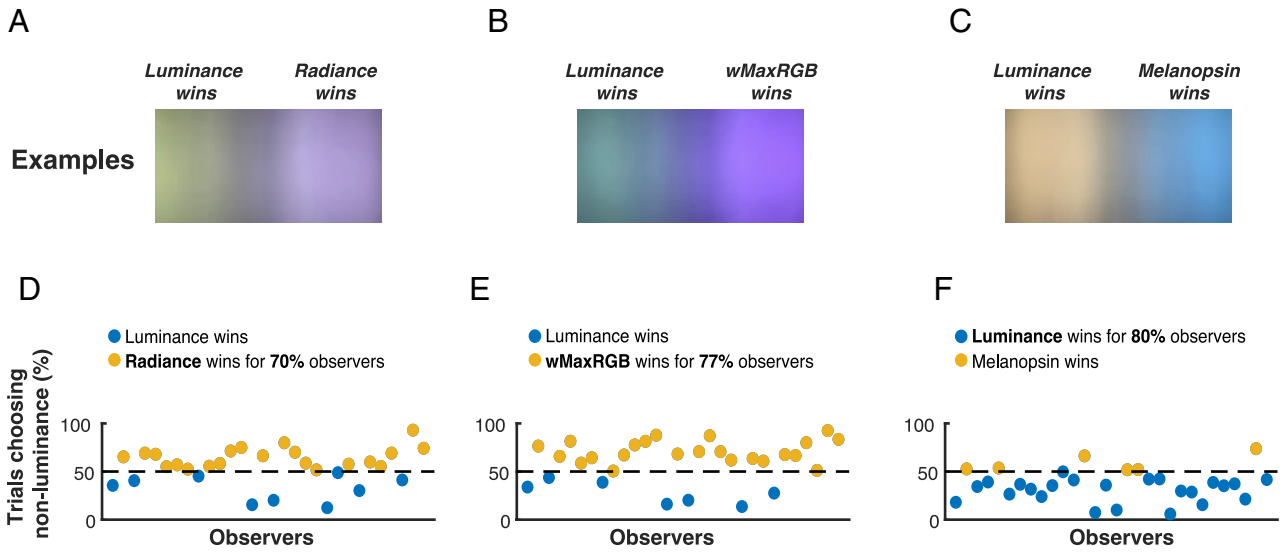

Fig. 4. Nonlinear pooling predicts brightness under real-world LED illumination. The *Top* panels show example lighting pairs with conflicts between Luminance and other predictors. (*A*) The left light has higher Luminance, while the right light has higher Radiance. (*B*) Higher Luminance (left light) vs. higher wMaxRGB (right light). (*C*) Higher Luminance (left light) vs. higher Melanopsin Excitation (right light). The *Bottom* panels (*D*–*F*) show percentage of trials where observers chose non-Luminance predictors over Luminance in conflicting scenarios. Each circle represents an observer (blue: Luminance >50%, yellow: other predictors >50%). See *SI Appendix*, Fig. S9 (*C*–*F*) for more conflicting scenarios.

significantly exceeded chance and also exceeded choices consistent with luminance. In contrast, when luminance was placed in opposition to melanopsin-based excitation (Fig. 4*C*), observers largely followed luminance rather than melanopsin predictions (Fig. 4*F*). Thus, not all alternative predictors outperform luminance. Instead, the advantage appears specific to nonlinear pooling rules that capture steady heterochromatic brightness.

These real-world lighting experiments demonstrate that the failure of additivity is not restricted to isolated display patches but extends to spatially extended illuminated environments. When spectral content is manipulated under steady viewing, perceived brightness aligns more closely with nonlinear pooling than with additive luminance.

Across psychophysical measurements, cortical responses, and real-world illumination, steady heterochromatic brightness consistently departs from additive luminance and is better explained by nonlinear pooling across color channels. The optimal channel weights underlying this computation are shown in *SI Appendix*, Fig. S10.

Discussion

The present results show that the perceived brightness of colored lights under steady viewing systematically departs from additive luminance and is better predicted by nonlinear pooling across color channels. Modern photometry, however, is largely grounded in classical photometric methods, particularly heterochromatic flicker photometry together with adjacent-field brightness matching, both of which are based on simplified viewing conditions (5, 6, 8, 9, 28, 29). Flicker photometry directly isolates a luminance-cancellation signal, whereas adjacent-field brightness matching constrains judgments to two uniform fields sharing a common border, a geometry closely related to minimum-distinct-border photometry. These methods yield additivity, which the candela and all derived photometric quantities inherit (3, 4, 10, 30). Our findings indicate that this structure does not extend to steady perceptual brightness. Under sustained viewing conditions, which characterize most natural scenes and everyday visual experience, brightness judgments are highly reliable yet follow a distinct computational rule.

The frequency-dependent neural dissociation provides a mechanistic link between these regimes. At higher temporal frequencies, cortical responses closely follow luminance, consistent with the physiological basis of flicker photometry (8, 26, 27). At lower temporal frequencies, responses align more closely with the nonlinear pooling rules that also explain perceptual rankings. This pattern reflects a distinction between transient luminance mechanisms and sustained cortical representations that integrate chromatic and intensity information. Luminance-selective responses are relatively sparse in cortex and are most prominent in areas such as MT (31–33), whereas neurons in early and intermediate visual cortex commonly combine chromatic and luminance contrast (34–40). Nonlinear pooling over opponent channels therefore provides a plausible account of steady brightness as a cortical computation rather than a direct consequence of the linear cone-weighted summation that underlies photometric luminance.

Color appearance models were developed to address known discrepancies between luminance and perceived brightness (20, 21). These models introduce nonlinear corrections that improve perceptual correspondence while retaining luminance as a foundational axis. The present findings suggest that under sustained viewing conditions, additivity itself is not preserved. The superior performance of nonlinear pooling rules over luminance and representative appearance models indicates that steady brightness is not merely a corrected form of luminance but reflects a distinct computational regime. A maximum-based pooling rule is not foreign to visual computation. Variants of MaxRGB underlie widely used color representations in computer graphics and design, including the value component of HSV color space (41). Their success reflects the ability of simple maximum operations to capture salient perceptual structure. The present results indicate that a related nonlinear pooling rule also provides a compact description of steady heterochromatic brightness in human observers.

The strong performance of the RGB-based model deserves comment. RGB primaries may seem like an arbitrary display choice, but they are not. An RGB basis is close to optimal for representing the range of colors found in natural objects using only three positive channels (25). This is why display manufacturers independently converge on similar primaries and presumably why

color representations in the visual system, shaped by the statistics of natural scenes, are efficiently described in this space. Consistent with this, our results are robust across displays with different primaries: Observers tested in the laboratory and on their own home displays produced essentially identical brightness rankings and channel weights (Fig. 1*D* and *SI Appendix*, Fig. S10). RGB is also not the unique space in which the max model succeeds. Rectified DKL channels perform nearly as well as RGB, suggesting that the critical computation is nonlinear pooling over relatively independent color channels at later cortical stages (Figs. 1*J* and 2*C*).

These findings do not challenge luminance within the temporal regime for which it was defined. Flicker photometry and rapidly modulated stimuli are well described by additive spectral weighting. Rather, the results delineate a boundary between transient luminance mechanisms and the perceived brightness of colored lights under the steady viewing conditions that characterize most natural scenes. Under these conditions, brightness follows nonlinear pooling across color channels rather than additive luminance. This distinction may become increasingly relevant in modern lighting environments, where spectrally tunable LEDs allow large changes in spectral composition while maintaining identical luminance.

Materials and Methods

Here we give a summary description of our methods. More details about display calibration, color space transformations, model implementations, fitting procedures, EEG preprocessing, and additional control analyses are provided in *SI Appendix, Supplementary Materials and Methods*.

Behavioral Brightness Ranking.

Laboratory experiments. Forty-five observers with normal color vision [Ishihara-tested (42)] participated; one was excluded due to low catch-trial accuracy (<75%). 21 observers completed a retest after ≥6 mo. All participants provided written (or electronic, for online studies) informed consent prior to participation. All procedures were approved by the local ethics committee (Giessen LEK 2020-0015) and complied with the Declaration of Helsinki.

Stimuli were presented on a calibrated OLED display running in sRGB mode (linearized RGB; details in *SI Appendix, Supplementary Materials and Methods*). Whenever we speak of RGB, we refer to linearized sRGB. Observers viewed 12 colored patches (4.6° squares) arranged in a 4 × 3 grid on a black background and ranked them from brightest to darkest by sequential clicking. Each patch set sampled 12 hues (primary, secondary, and intermediate directions in RGB space) and 12 intensity levels (10 to 80%). Each observer completed 66 trials (including six catch trials with monotonic intensity variation). Patch positions and trial order were randomized. Catch trials were used to verify task compliance. For each observer, pairwise comparisons within catch trials (396 pairs) were evaluated against model-consistent predictions; observers below 75% accuracy were excluded. For remaining observers, rankings were averaged across repetitions to obtain a mean brightness rank for each of 144 stimuli. Reliability was assessed using Spearman correlations within and across sessions.

Model performance was evaluated by correlating predicted brightness with behavioral rankings. Tested models included standard photometric and color appearance measures [e.g., luminance (6), CIELAB L* (21), CIECAM02/16 brightness (43, 44)], as well as classical and modern extensions incorporating chromatic contributions (see *SI Appendix, Supplementary Materials and Methods* for full list and parameterizations). In addition, we evaluated a generalized Minkowski formulation over RGB:

$$I^* = \left((w_R R)^n + (w_G G)^n + (w_B B)^n \right)^{1/n}.$$

Weights w_R, w_G, w_B were optimized to minimize unexplained variance. The limiting cases correspond to weighted sum (n = 1) and weighted max (n→∞). Fits were performed at both individual and group levels. Equivalent analyses were conducted in LMS and DKL color spaces after normalization [details in *SI Appendix, Supplementary Materials and Methods*; (45, 46)].

Online experiments. To test generalizability, the same task was performed outside the laboratory. First, the laboratory observers repeated the experiment at home on their own displays. Second, a larger sample was collected via Prolific (three groups with black, gray, or white backgrounds; total N = 486 after exclusions). Stimuli and procedures were identical to the laboratory experiment, except that display calibration was not controlled. For a subset of home displays (N = 43), spectral measurements were obtained and compared to the laboratory display (*SI Appendix, Supplementary Materials and Methods*). Despite variability in display characteristics, behavioral rankings showed high consistency across environments. Analyses followed the same pipeline as in the laboratory. We additionally quantified cross-display robustness by comparing model predictions across measured display gamuts and evaluating consistency of pairwise rankings. Model fits (including optimized weights for wSumRGB and wMaxRGB variants) were computed separately for each dataset.

Chroma variation experiment. To test the role of chroma, we extended the stimulus set to the full RGB cube. Observers performed the same ranking task in laboratory, home, and Prolific settings (total N = ~190 after exclusions). For each hue–intensity combination, multiple chroma levels were generated by interpolating toward the corresponding gray. This yielded 732 stimuli spanning hue, intensity, and chroma. Trials contained 12 patches sampling all dimensions. Procedures and exclusion criteria were identical to the main experiment. Model fits were computed both across all chroma levels and separately within each chroma slice. This allowed us to assess i) how chroma modulates perceived brightness, ii) how model weights change with chroma, and iii) consistency with the fixed-chroma experiment. Rankings were highly consistent across experimental settings, supporting robustness of the effects.

Statistical analysis. All analyses were conducted in MATLAB. Model performance was quantified using Spearman correlation and unexplained variance ($1 - \rho^2$). Weight parameters were estimated by minimizing unexplained variance. Statistical comparisons and robustness analyses across observers, environments, and stimulus sets are detailed in *SI Appendix, Supplementary Materials and Methods*.

SSVEP. To link perceptual judgments to neural responses, we measured SSVEPs in 10 observers with normal vision. Stimuli were centrally presented color discs (1° radius) flickering at 3 Hz or 15 Hz with sinusoidal contrast modulation. Seven colors (primary and secondary combinations) were tested at multiple intensity levels. EEG was recorded from occipital electrodes (O1, Oz, O2) at 1,000 Hz. Each trial (4 s) consisted of four successive intensity epochs. Frequency-domain amplitudes were computed via Fourier transform, and SSVEP responses were quantified by subtracting adjacent-frequency noise and averaging across harmonics [<30 Hz, (47)]. Analyses were conducted separately for 3 Hz and 15 Hz conditions. Neural responses were correlated with model predictions (including luminance and MaxRGB-type metrics). Details of preprocessing and signal extraction are provided in *SI Appendix, Supplementary Materials and Methods*.

Real-World Lighting Experiments. To test ecological validity, we measured brightness judgments under real illumination using spectrally tunable LED sources. Thirty observers viewed a neutral room illuminated by two independently controlled light fields. On each trial, observers adapted to a reference illuminant (D65), followed by a brief presentation of two competing illuminants (left vs. right), and indicated which side appeared brighter. Illuminants were constructed to create conflicts between different predictors (e.g., higher luminance vs. higher radiance or chromatic contributions). A total of 21 predictor pairs were tested, each instantiated by multiple spectral combinations. Catch trials ensured task compliance. Spectral power distributions were measured with calibrated spectroradiometers and converted into predictor values (luminance, radiance, melanopsin excitation (48), sumRGB, maxRGB, and optimized weighted variants). RGB representations were derived from spectral partitions approximating cone fundamentals [*SI Appendix, Supplementary Materials and Methods*; (25)]. For each predictor pair, we computed the proportion of trials in which observers selected the illuminant favored by that predictor. This provided a direct behavioral comparison between competing models under realistic viewing conditions.

Data, Materials, and Software Availability. Anonymized human responses data have been deposited in https://osf.io/65bwf (49).

ACKNOWLEDGMENTS. We are grateful to Doris Braun for valuable discussions of an earlier version of this manuscript. This project was funded by European Research Council ERC AdG Color 3.0 (884116), Deutsche Forschungsgemeinschaft DFG Collaborative Research Center SFB/TRR 135 (Project No. 222641018), and by DFG Excellence Cluster EXC 3066/1 "The Adaptive Mind" (533717223).

Author affiliations: [a]Department of Psychology and Center for Mind, Brain and Behavior, Justus-Liebig-Universität Giessen, Giessen 35394, Germany; [b]School of Psychology, Shanghai University of Sport, Shanghai 200438, China; [c]Department of Psychology, Faculty of Science and Technology, Bournemouth University, Poole BH12 5BB, United Kingdom; [d]Department of Psychology, University of Bologna, Bologna 40127, Italy; [e]Department of Experimental Psychology, Utrecht University, Utrecht 3584 CS, the Netherlands; and [f]Laboratory for Experimental Psychology, Catholic University of Leuven, Leuven 3000, Belgium

1. EIA, "How much electricity is used for lighting in the United States?" (US Energy Information Administration, 2024). https://www.eia.gov/tools/faqs/faq.php?id=99&t=3.
2. C. Jensen, "What is energy efficient light?: A socio-technical analysis of lighting in transition," PhD dissertation, Aalborg Universitet, Aalborg, Denmark (2014). https://doi.org/10.13140/RG.2.1.1679.6567.
3. C. H. Page, P. Vigoureux, "The International System of Units (SI)" (US Department of Commerce, National Bureau of Standards, 1977). National Bureau of Standards Special Publication Number 330, https://www.nist.gov/pml/special-publication-330.
4. D. Newell, E. Tiesinga, *The International System of Units (SI) NIST Special Publication 330* (National Institute of Standards and Technology, ed. 2019, 2019).
5. CIE, *History of the CIE* (International Commission on Illumination, 1988).
6. CIE, *Commission Internationale de l'eclairage Proceedings, 1924* (Cambridge University Press, Cambridge, 1926).
7. K. S. Gibson, E. P. T. Tyndall, *Visibility of Radiant Energy* (US Government Printing Office, 1923), vol. 19.
8. B. Lee, P. Martin, A. Valberg, The physiological basis of heterochromatic flicker photometry demonstrated in the ganglion cells of the macaque retina. *J. Physiol.* **404**, 323–347 (1988).
9. R. A. Bone, J. T. Landrum, Heterochromatic flicker photometry. *Arch. Biochem. Biophys.* **430**, 137–142 (2004).
10. G. Wyszecki, W. S. Stiles, *Color Science: Concepts and Methods, Quantitative Data and Formulae* (John Wiley and Sons, 2000).
11. R. M. Boynton, "Human color perception" in *Science of Vision*, K. N. Leibovic, Ed. (Springer, 1990), pp. 211–253.
12. H. L. Mahoney, T. M. Schmidt, The cognitive impact of light: Illuminating ipRGC circuit mechanisms. *Nat. Rev. Neurosci.* **25**, 159–175 (2024).
13. D. B. Judd, "Report of US secretariat committee on colorimetry and artificial daylight" in *Proceedings of the 12th Session of the CIE*, CIE, Ed. (Bureau Central de la CIE, Paris, France, 1951). p. 11.
14. A. Dresler, The non-additivity of heterochromatic brightnesses. *Trans. Illum. Eng. Soc.* **18**, 141–165 (1953).
15. S. L. Guth, N. J. Donley, R. Marrocco, On luminance additivity and related topics. *Vision Res.* **9**, 537–575 (1969).
16. G. S. Wasserman, C. B. Gillman, Subadditivity and superadditivity of heterochromatic lights. *Psychol. Rev.* **77**, 338 (1970).
17. P. K. Kaiser, G. Wyszecki, Additivity failures in heterochromatic brightness matching. *Color Res. Appl.* **3**, 177–182 (1978).
18. V. A. Kohlrausch, Zur photometrie farbiger lichtern. *Das Licht* **6**, 259–279 (1935).
19. W. Cowan, C. Ware, "On the brightness of colours that differ in hue or saturation" in *Proceedings of the 11th Session of the CIE*, CIE, Ed. (Bureau Central de la CIE, Paris, France, 1987), pp. 116–119.
20. M. D. Fairchild, *Color Appearance Models* (John Wiley and Sons, 2013).
21. CIE, "Recommendations on uniform color spaces, color difference equations, psychometric color terms"" (Bureau Central de la CIE, Paris, France, 1978).
22. J. Koenderink, A. van Doorn, K. Gegenfurtner, Color weight photometry. *Vision Res.* **151**, 88–98 (2018).
23. G. Jacobs, *Comparative Color Vision* (Elsevier, 2013).
24. T. Baden *et al.*, A standardized nomenclature for the rods and cones of the vertebrate retina. *PLoS Biol.* **23**, e3003157 (2025).
25. J. Koenderink, A. v. Doorn, K. R. Gegenfurtner, Colors and things. *i-Perception* **11**, 2041669520958431 (2020).
26. D. Regan, Objective method of measuring the relative spectral-luminosity curve in man. *J. Opt. Soc. Am.* **60**, 856–859 (1970).
27. J. B. Siegfried, D. I. Tepas, H. G. Sperling, R. H. Hiss, Evoked brain potential correlates of psychophysical responses: Heterochromatic flicker photometry. *Science* **149**, 321–323 (1965).
28. W. de W. Abney, *Researches in Colour Vision* (Sampson Low, Marston and Co, 1913).
29. W. de W. Abney, E. R. Festing, XII. The Bakerian Lecture—Colour photometry. *Philos. Trans. R. Soc.* **177**, 423–456 (1886).
30. C. I. E., *International Lighting Vocabulary* (CIE, Paris, 1970).
31. K. R. Gegenfurtner *et al.*, Chromatic properties of neurons in macaque MT. *Vis. Neurosci.* **11**, 455–466 (1994).
32. W. H. Merigan, J. Maunsell, How parallel are the primate visual pathways? *Annu. Rev. Neurosci.* **16**, 369–402 (1993).
33. R. B. Tootell *et al.*, Functional analysis of human MT and related visual cortical areas using magnetic resonance imaging. *J. Neurosci.* **15**, 3215–3230 (1995).
34. R. Shapley, M. J. Hawken, Color in the cortex: Single- and double-opponent cells. *Vision Res.* **51**, 701–717 (2011).
35. D. C. Kiper, S. B. Fenstemaker, K. R. Gegenfurtner, Chromatic properties of neurons in macaque area V2. *Vis. Neurosci.* **14**, 1061–1072 (1997).
36. A. K. Garg, P. Li, M. S. Rashid, E. M. Callaway, Color and orientation are jointly coded and spatially organized in primate primary visual cortex. *Science* **364**, 1275–1279 (2019).
37. P. Lennie, J. Krauskopf, G. Sclar, Chromatic mechanisms in striate cortex of macaque. *J. Neurosci.* **10**, 649–669 (1990).
38. L. G. Thorell, R. L. de Valois, D. G. Albrecht, Spatial mapping of monkey VI cells with pure color and luminance stimuli. *Vision Res.* **24**, 751–769 (1984).
39. E. N. Johnson, M. J. Hawken, R. Shapley, The spatial transformation of color in the primary visual cortex of the macaque monkey. *Nat. Neurosci.* **4**, 409–416 (2001).
40. K. R. Gegenfurtner, D. C. Kiper, S. B. Fenstemaker, Processing of color, form, and motion in macaque area V2. *Vis. Neurosci.* **13**, 161–172 (1996).
41. A. R. Smith, Color gamut transform pairs. *ACM Siggraph Comput. Graph.* **12**, 12–19 (1978).
42. S. Ishihara, *Ishihara's Tests for Colour Deficiency* (Kanehara and Company, ed. 24 plates, 1997).
43. C. Li *et al.*, Comprehensive color solutions: CAM16, CAT16, and CAM16-UCS. *Color Res. Appl.* **42**, 703–718 (2017).
44. N. Moroney, M. Fairchild, R. Hunt, C. Li, "The CIECAM02 color appearance model" in *Proceedings of the IS&T/SID Tenth Color Imaging Conference*, The Society for Imaging Science and Technology, Ed. (Society for Imaging Science and Technology, Springfield, VA, 2002).
45. A. M. Derrington, J. Krauskopf, P. Lennie, Chromatic mechanisms in lateral geniculate nucleus of macaque. *J. Physiol.* **357**, 241–265 (1984).
46. J. Krauskopf, D. R. Williams, D. W. Heeley, Cardinal directions of color space. *Vision Res.* **22**, 1123–1131 (1982).
47. T. L. Retter, B. Rossion, C. Schiltz, Harmonic amplitude summation for frequency-tagging analysis. *J. Cogn. Neurosci.* **33**, 2372–2393 (2021).
48. M. Spitschan, Melanopsin contributions to non-visual and visual function. *Curr. Opin. Behav. Sci.* **30**, 67–72 (2019).
49. S. Guan *et al.*, Data for "Perceptual and neural constraints on photometric measures of heterochromatic brightness." OSF. https://osf.io/65bwf. Deposited 2 August 2026.