Psychological Bulletin

ISSN: 0033-2909

2023, Vol. 149, Nos. 7–8, 487–505
https://doi.org/10.1037/bul0000399

Pervasive Failure to Report Properties of Visual Stimuli in Experimental Research in Psychology and Neuroscience: Two Metascientific Studies

Zhicheng Lin[1], Qi Ma[2], Xiaolin Huang[1], Xuebing Wu[1], and Yang Zhang[2]
[1] Department of Applied Psychology, The Chinese University of Hong Kong
[2] Department of Psychology, Soochow University

Transparency in research reporting is crucial for evaluating the reproducibility and validity of research, including potential confounding factors (internal validity) and generalizability (external validity). Here, we focus on visual stimuli—stimuli routinely used to elicit mental processes and behaviors—as a case study to systematically assess and evaluate current practices in reporting visual characteristics, including display setup, stimulus size, luminance/color, and contrast. Our first study scrutinized recent publications ($N = 360$) in leading journals in neuroscience and psychology—spanning vision, cognitive, clinical, developmental, and social/personality psychology. The second study examined recent publications ($N = 114$) on visual attentional bias in clinical samples, involving tasks known to be sensitive to visual properties. Analyzing the full text and supplemental materials of these articles, the two studies reveal a systematic lapse in current practices of reporting characteristics of visual stimuli. This reporting failure was not due to authors making visual materials available online, which was rare (<20%) and could not replace the reporting of visual characteristics. Failure to report stimulus properties hinders efforts to build cumulative science: (a) direct replications become challenging if not impossible; (b) internal validity may be compromised; and (c) generalizability across stimulus properties is prematurely assumed, and its evaluation is precluded in the first place. Our findings have immediate implications for journal policies on reporting practices, urging for explicit emphasis on transparent reporting of stimulus properties, particularly when perceptual components are involved. To assist in this effort, we provide templates for reporting study setup, visual displays, and visual stimuli.

Public Significance Statement
Numerous editorials, writing guides, author guidelines, and manifestos for reproducible science have focused on the importance of transparency in data, code, and research materials. We identify a pervasive failure to report properties of visual stimuli in research in psychology and neuroscience—which may undermine research reproducibility and validity—and outline potential solutions.

Keywords: visual properties, transparency, reproducibility, generalizability (external validity), internal validity

Supplemental materials: https://doi.org/10.1037/bul0000399.supp

The past decade has witnessed a growing emphasis on the importance of replication (Frank & Saxe, 2012; Nosek et al., 2022; Zwaan et al., 2017) and generalizability (Durgin et al., 2012; Yarkoni, 2022) in psychology, neuroscience, and allied fields. Indeed, an alarming number of replication failures (Open Science Collaboration, 2015) have instigated methodological reflections and reforms aimed at

Zhicheng Lin https://orcid.org/0000-0002-6864-6559
Data collection and analysis for journals in vision science, cognitive psychology, and neuroscience were preregistered at https://osf.io/9jhrq. The data set and summary results are available online at https://osf.io/gpn2f/.

The authors thank Bowen Chen for assistance in literature coding. The study was supported by the National Key R&D Program of China (Grants 2021ZD0204200, 2022YFB4500601), the National Natural Science Foundation of China (Grants 32071045, 32171049), Guangdong Basic and Applied Basic Research Foundation (Grant 2019A1515110574), and Shenzhen Fundamental Research Program (Grant JCYJ20210324134603010). The authors have no known conflicts of interest to disclose.

Zhicheng Lin designed the study, supervised the research, analyzed the data, drew the figures, and wrote the article; Qi Ma extracted articles for the first study, coded the two studies, and helped make Figure 1; Xiaolin Huang and Xuebing Wu extracted articles for the second study and contributed to the coding; and Yang Zhang contributed to part of the drawing of Figure 2 and provided managerial support during the early stage of the first study.

Correspondence concerning this article should be addressed to Zhicheng Lin, Department of Applied Psychology, The Chinese University of Hong Kong, 2001 Longxiang Boulevard, Longgang District, Shenzhen, Guangdong 518172, China, or Yang Zhang, Department of Psychology, Soochow University, No. 333, Ganjiang East Road, Wuzhong District, Suzhou, Jiangsu 215006, China. Email: zhichenglin@gmail.com or yzhangpsy@suda.edu.cn

improving the reproducibility and robustness of scientific research. Irreproducible/unreplicable research has been partly attributed to questionable research practices (John et al., 2012), such as selective reporting of studies, outcomes, and analyses. A notorious example is *p*-hacking, which involves using biased and invalid methods to obtain statistically significant results (Simmons et al., 2011). As remedies, replication has been suggested to assess reproducibility and robustness in past research. Practices like disclosure (e.g., Aczel et al., 2020), preregistration (before peer review or, in the form of Registered Reports, after peer review; Bosnjak et al., 2022; Chambers & Tzavella, 2022), and other open-science practices have been championed to increase transparency, curtail questionable research practices, and consequently improve reproducibility and robustness in future research (De Boeck & Jeon, 2018; Nelson et al., 2018; Nosek et al., 2022).

Accordingly, prior to data collection or analysis, researchers are encouraged to preregister their study plan, including disclosing methodological and analytical specifics (Bosnjak et al., 2022). When preparing their articles, authors are further urged, and at times mandated,[1] to disclose all dependent variables (measures), conditions (groups, predictors), and data exclusions (participants or observations), and to make data, code, and materials available.

Here, we show that the current guidelines are inadequate, neglecting a critical issue: the presentation and appearance of study materials. Specifically, through two metascientific analyses of recent publications in psychology and neuroscience, we demonstrate a systematic failure in reporting properties of visual stimuli in contemporary practices. We contend that these reporting practices (a) render direct replications cumbersome if not unattainable; (b) potentially jeopardize internal validity due to factors such as confounds; and (c) prematurely assume generalizability across stimulus properties, thereby affecting external validity. The adverse consequences of these practices are likely to vary across different research areas and topics, and hence should be assessed on a case-by-case basis. Researchers are therefore advised to seriously consider the impact of stimulus properties on their findings, to not dismiss them outright without a sound empirical basis, particularly when perceptual processes are involved.

In what follows, we first scrutinize the current approach of translating the goals of transparency and openness into research practices, pinpointing a neglect of material presentation and appearance in current guidelines. We argue that this oversight is problematic, potentially undermining research replicability, validity, and generalizability. As further evidence, we offer concrete examples illustrating how visual stimulus properties can influence a broad range of commonly studied processes, from perceptual and cognitive to affective and social. We then delineate our approach to examining current reporting practices concerning visual stimulus properties, before delving into the Method and Results sections. In the Discussion, we outline possible structural reasons underlying the present practices, elaborate on situations where the reporting of stimulus properties is likely to be important, and examine the implications for both policy and practices—offering reporting templates for study setup, visual displays, and visual stimuli.

The Road to Reproducible and Replicable Research

From the vantage point of transparent science communication, we focus on how research is reported in journal articles to allow direction replication and assessment of internal and external validity. There is a consensus on the importance of transparency and openness in countering irreproducible research and fostering a cumulative research enterprise. On paper, it also seems deceptively simple to implement the principles of transparency and openness into research practices: just fully describe the research. Indeed, current recommendations and guidelines often use terms like "full," "enough," and "any" when dispensing advice. Hence, during preregistration, researchers are advised to "describe any relevant study materials" (Bosnjak et al., 2022, p. 612); in article preparation, they are further urged to describe "in full the study design and data" (Munafo et al., 2017, p. 6), to provide "enough detail so readers will know how each result was obtained" (Gernsbacher, 2018, p. 405), to "fully (describe) the study design, procedures, and materials to allow independent replication" (Aczel et al., 2020, pp. 4–5), and to ensure "full sharing of methodological details" (Pham & Oh, 2021, p. 173).

Unfortunately, such prescriptions, though concise, are not completely actionable. After all, how does one *fully* describe the research? Recent editorials (Lindsay, 2017), writing guides (Gernsbacher, 2018), author guidelines (Nosek et al., 2015), and a manifesto for reproducible science (Munafo et al., 2017) have predominantly emphasized the availability of data, code, and research materials (see Supplemental Box S1). For example, in the widely adopted Transparency and Openness Promotion (TOP) guidelines, research materials are recommended or mandated to be posted to a trusted repository (Nosek et al., 2015). These perspectives are echoed in current journal guidelines, which rightly emphasize transparency in data, code, and research materials.

Yet, the manner in which research materials are presented in a given study—their appearance—has received little emphasis or elaboration. From a participant's viewpoint, the essence of an object lies in its very appearance. To the extent that research materials are used to tap into the targeted psychological/neuronal processes and behaviors, their properties—which define their appearance—are an integral part of a study, much like properties such as size and color are integral to a piece of clothing. Imagine a study where researchers present two types of images to participants on a computer screen and observe a difference in some measures. To replicate this effect, we need not only access to these images but also an understanding of their specific presentations—such as their size and brightness. Without this knowledge of the stimulus properties, it is difficult to ascertain whether potential discrepancies between the replication and the original study reflect the replicability of the original effect or are due to uncounted differences in stimulus properties.

To illustrate with a concrete case, consider a recent study that investigated whether self-face images capture attention subliminally—that is, without the images entering awareness (Wójcik et al., 2019). However, the study did not report the luminance and contrast of the face and mask images. This omission presents a challenge for researchers attempting to replicate the study: if the contrast of the target face image is too high, the face images may become visible; if the contrast is too low, the face images may not trigger any relevant processes, resulting in a floor effect. Consequently, the failure to measure and report luminance and contrast hampers efforts to replicate the study.[2]

[1] For example, https://www.psychologicalscience.org/publications/psychological_science/ps-submissions.

[2] It is notoriously difficult to obtain none-reported information by contacting the original authors. In this specific case, we emailed the first author for clarifications on methodological details, and received no reply with two attempts (on July 28, 2022 and August 4, 2022).

Stimulus Properties and the Evaluation of Validity and Generalizability

Stimulus properties are pertinent not just to the evaluation of replicability but also to the assessment of internal validity and generalizability.[3] Revisiting the comparison between two types of images, researchers typically aim to draw conclusions about an abstract, categorical-level construct of interest, as exemplified by the images—rather than about the images themselves. The psychology and neuroscience literature is replete with such examples, including comparisons between food images and nonfood images to study food information processing, comparisons between emotional and neutral faces to understand emotional processing, comparisons between self-face images and other-face images to explore self-relevant processing, and so on. Yet, to draw causal conclusions about these abstract constructs, it is necessary to rule out alternative accounts, specifically low-level, image-based confounds such as differences in luminance and contrast. Indeed, in widely used picture databases, pictures with different affective ratings have been shown to also differ in low-level image properties, making them a critical confound to measure and tackle (Delplanque et al., 2007). Stimulus properties, therefore, bear directly on the internal validity of our studied construct (for a recent and concise discussion on validities, see Vazire et al., 2022). Undisclosed differences in these properties between images pose a hidden confound that may undermine the internal validity of the studied construct, and thus threaten the validity of the conclusion drawn.

Stimulus properties also directly affect generalizability, otherwise known as external validity in psychometrics. Discourse on generalizability often focuses on subject samples (Henrich et al., 2010) and study contexts (e.g., time periods and geographies; Delios et al., 2022; Lin & Li, 2023). But overlooking the issue of generalization across stimuli may pose a validity challenge in much of psychology research (Judd et al., 2012; Wells & Windschitl, 1999; Yarkoni, 2022). This issue arises when there is a failure to sample stimuli and test the generality over what are presumed to be irrelevant stimulus variables, thereby threatening the very validity of the construct under study (Wells & Windschitl, 1999; Yarkoni, 2022). Suppose we want to test how perceived colorfulness affects emotion. Two things will be critical for proper inferences: we need to sample more than just one set of colors in order to properly generalize from the specific measurement to the abstract construct; we also need to make sure that the effect can generalize across properties unrelated to colorfulness (e.g., size). In other words, generalizability is intrinsically linked to construct validity—the extent to which a measurement taps into the intended construct of interest (Vazire et al., 2022)—with generalizability over stimuli an important domain to evaluate. Indeed, overlooking seemingly trivial variations in stimuli (and other details) may be one cause of the ongoing replication and generalization crisis in psychology and neuroscience (Yarkoni, 2022).

Let us revisit the earlier image comparison example. Suppose we would like to contrast upright faces and inverted faces. To draw any meaningful conclusion about their differences (e.g., a face inversion effect) at the construct level (e.g., holistic processing), we need to ensure that the observed differences can generalize across stimuli—rather than being confined to a specific, limited set of stimuli. In other words, to make proper inferences, we need to ensure stimulus sampling variability (Clark, 1973). This goal can be accomplished by systematically varying the levels of the stimulus factor—this factor can then be modeled as a random effect in data analysis (Judd et al., 2012; Yarkoni, 2022). In addition to variation in stimulus exemplars, variation in stimulus properties—such as in luminance, contrast, and spatial frequency—may also systematically modulate the observed effect. Such modulations provide a rich test bed to evaluate the involved mechanisms (e.g., Goffaux & Rossion, 2006; Vuilleumier et al., 2003).

It is in this spirit of testing generalizability and robustness that recent metastudies attempt to systematically manipulate a multitude of variables in a single experimental design (Baribault et al., 2018; DeKay et al., 2022). For example, to compare two types of images—a subliminal cue (briefly presented and masked) versus a supraliminal one (unmasked)—Baribault et al. (2018) systematically tested the effects of theoretically pertinent variables (variables that were expected to moderate the effect) and nuisance variables (those expected to be generalized over). Hence, the authors varied stimulus timing, color, contrast, symbol, and location, and assigned different implementations to different microexperiments within one large metastudy. Using hierarchical Bayesian models, they found robustness to changes in colors and symbols but observed large heterogeneity in stimulus timing. As revealed by moderator analysis, the so-called "subliminal" effect only manifested itself when the cue was visible.

As the above example illustrates, the question of whether an effect can generalize over a stimulus property is ultimately an empirical one, requiring explicit testing. Metastudies provide an intriguing approach to this, complementing the more traditional factorial method—which involves manipulating a fixed number of levels in the variables of interest and then crossing them. Regardless of the approach we choose, assessing generalizability over a stimulus property necessitates a systematic variation of its levels. But to achieve this, the stimulus property must first be measured. Therefore, the failure to report stimulus properties effectively precludes the critical evaluation of generalizability across these properties.

Modulations by Visual Stimulus Properties Across Perceptual, Cognitive, Affective, and Social Processes

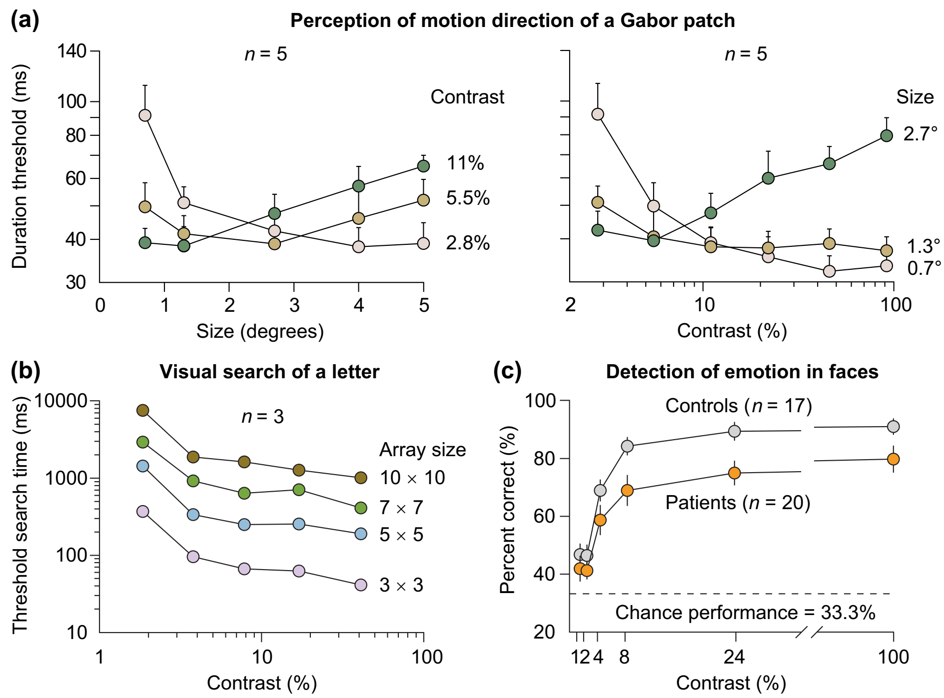
The outcome of a generalizability test is both empirically valuable and theoretically relevant. Predicted or assumed generalizability must undergo empirical testing, and the results subsequently inform our understanding of the phenomenon, either strengthening or challenging our theory. While confirmed generalizability bolsters a theory, unexpected generalizability exposes potential mechanisms that may lead to a new understanding of the effect. Exactly when and how stimulus properties modulate a given effect, currently no theory can specify (if there will ever be one). Yet that precisely speaks to the importance of empirical testing. Examples of stimulus properties modulating a wide range of processes are in order, so as to illustrate the key point—that researchers are well served to seriously consider the impact of stimulus properties rather than dismiss them outright without a sound empirical basis.

At the perceptual end, many phenomena are highly dependent on stimulus properties. For example, as Figure 1a shows, in motion

[3] To the extent that a study cannot be repeated without some changes (e.g., time, participants, environment), replication can be considered a test of generalizability across different contexts.

Figure 1

Effects of Stimulus Contrast on Behavioral Performance in Perceptual, Cognitive, and Affective Research



Note. (a) The task was to indicate the motion direction of a Gabor patch. Performance—measured by the duration of stimulus presentation required for a given probability of successful judgment—improved (i.e., the duration threshold decreased) when the size increased, but only for low-contrast stimuli; the effect was reversed for high-contrast stimuli. Data adapted from "Perceptual consequences of centre-surround antagonism in visual motion processing," by D. Tadin, J. S. Lappin, L. A. Gilroy, and R. Blake, 2003, *Nature*, *424*(6946), pp. 312–315 (https://doi.org/10.1038/nature01800). (b) The task was to search for an uppercase letter among various numbers of distractors (i.e., numerals). Performance—measured by the duration of stimulus presentation required for a given probability of successful search—was strongly affected not only by the number of items but also by the contrast of the items themselves. Each data point represents the arithmetic mean of three participants. Data adapted from "Effect of stimulus contrast on performance and eye movements in visual search," by R. Näsänen, H. Ojanpää, and I. Kojo, 2001, *Vision Research*, *41*(14), pp. 1817–1824 (https://doi.org/10.1016/S0042-6989(01)00056-6). (c) The task was to indicate the expressed emotion of each face image (happy, sad, or neutral), which was presented for 500 ms. For both control and schizophrenia patients, their accuracy was affected by stimulus contrast, particularly at the lower ends. Contrast was defined by Michelson contrast (the six levels on the right corresponded to RMS contrast of <1%, 1%, 2%, 3%, 8%, and 57%). Data adapted from "Sensory contributions to impaired emotion processing in schizophrenia," by P. D. Butler, I. Y. Abeles, N. G. Weiskopf, A. Tambini, M. Jalbrzikowski, M. E. Legatt, V. Zemon, J. Loughead, R. C. Gur, and D. C. Javitt, 2009, *Schizophrenia Bulletin*, *35*(6), pp. 1095–1107 (https://doi.org/10.1093/schbul/sbp109). RMS = root-mean-square. See the online article for the color version of this figure.

perception, increasing stimulus size *improves* the perception of low-contrast motion direction, but *impairs* the perception of high-contrast motion direction (Tadin et al., 2003). The latter effect—spatial suppression—is reduced in schizophrenia (Tadin et al., 2006), major depression (Golomb et al., 2009), and aging (Betts et al., 2005). Thus, studies examining the impact of size on perceptual performance may observe contradictory effects—enhancement or suppression—if the contrast is not known, highlighting the importance of measuring and reporting contrast and size.

The importance of stimulus properties extends beyond perception, affecting cognitive, affective, and social processes. For example, stimulus conditions are a central part of models that attempt to explain visual attention (Reynolds & Heeger, 2009). Specifically, visual properties have been shown to affect observed effects in common attention tasks (e.g., Figure 1b): stimulus luminance/contrast (Näsänen et al., 2001) and size (Proulx & Egeth, 2008) on visual search performance, stimulus luminance/contrast on the magnitude of Stroop interference (Dyer, 1973), and stimulus luminance (Hawkins et al., 1988; Zhao & Heinke, 2014) and size (Herrmann et al., 2010; Yeshurun & Carrasco, 2008) on attention cueing effects. Similarly, working memory is affected by stimulus properties such as contrast (Ikkai et al., 2010) and color (Bae et al., 2014).

Beyond perceptual and cognitive processes, stimulus properties also modulate affective processes across various domains (e.g., Figure 1c; Bayer et al., 2012; Butler et al., 2009; Fotios et al., 2015; Kumar & Srinivasan, 2011; Vuilleumier et al., 2003). For example, font size has been shown to modulate the emotional effects of written words (Bayer et al., 2012), luminance and size on judgments of emotion from facial expression (Fotios et al., 2015), and contrast on emotion identification deficits in schizophrenia (Butler et al., 2009).

The importance of stimulus properties is also evident in social processes. For example, the perception of animacy—the detection of life presence or absence, as in distinguishing between animate agents (such as humans) and inanimate agents (such as robots)—is crucial for social interaction. Judgments of animacy are influenced by the agent's visual appearance, which is dictated by stimulus properties such as form and motion (Cross et al., 2016; Scholl & Tremoulet, 2000). Another route of stimulus impacts is through confounding: an effect that is attributed to social processes might be confounded by differences in stimulus properties. For example, while the comparison of self-face images and other-face images is intended to probe self-relevant processing (Wójcik et al., 2019), their differences could also be due to self-irrelevant factors, such as luminance, contrast, and spatial frequency. Thus, whether stimulus properties directly influence social processes or confound their measurement, overlooking these properties introduces many opportunities for them to undermine our inferences.

Regardless of the process involved, when evaluating generalizability, it is important to consider the specific question, measurement, and a wide range of stimulus space. For example, in visual search (Figure 1b), the set size effect—longer searcher time for larger array size—appears to generalize over contrast, but search time is critically dependent on both stimulus contrast and set size. Likewise, in Figure 1c, the difference between the control and patient groups persists when the contrast level varies from 8% to 100%, but narrows and disappears when the contrast is from 1% to 8%, a range where the stimuli remain visible (Butler et al., 2009).

Current Reporting Practices on Visual Stimulus Properties

So far, we have illustrated the importance of considering stimulus properties in evaluating replicability, internal validity, and generalizability. To take stock of current reporting practices on visual stimulus properties, in the first study, we analyze reporting practices in leading journals in neuroscience and major subfields of psychology (vision, cognitive psychology, clinical psychology, developmental psychology, and social/personality psychology). Presumably, the latest articles in leading journals represent current best—or at least sanctioned—practices in the respective fields. Considering the ubiquity and centrality of vision as a modality in both daily life and research studies, we focus on visual stimuli displayed electronically as a case study. Indeed, compared to other modalities such as auditory, cutaneous, and chemical senses, vision is by far the most commonly used medium for stimulus presentation. Visual stimuli range from geometric shapes, moving dots, letters, characters, faces, objects, and scenes to video clips. They have been routinely used in psychological and neuroscience research to elicit relevant processes and behaviors.

A primary focus in extant open-science guidelines concerns the availability of materials such as stimuli. However, our analysis revealed that 81% of the articles surveyed in the first study did not share stimuli online. But even when stimuli are available, merely having access to them would not be sufficient—be it for direct replication or for the evaluation of generalizability and validity—because the *appearance* of visual stimuli critically depends on the actual properties of the stimuli as they are on the display. Indeed, even the same image on the same device can have very different appearances when just one display characteristic (e.g., γ value) changes. For example, notice in Figure 2d the drastic changes to the same image in brightness, color, contrast, and detail when the display γ changes. This issue of display-contingent visual appearance is a general problem plaguing many domains.[4]

Here, we focus on the reporting of visual characteristics (Figure 2; see a primer on various terms used to describe these properties in the Appendix). Properties to specify stimulus display and appearance can be categorized into two broad kinds: display setup and visual appearance. Information regarding display setup—display model, screen resolution (Figure 2a), and refresh rate (Figure 2b)—is relevant for evaluating the characteristics of the display (e.g., spatiotemporal properties), which constrain the characteristics of the stimuli that can be displayed. Information that defines the visual appearance of the stimuli includes size (Figure 2c), luminance/color (Figures 2e and 2f), and contrast (Figure 2g).

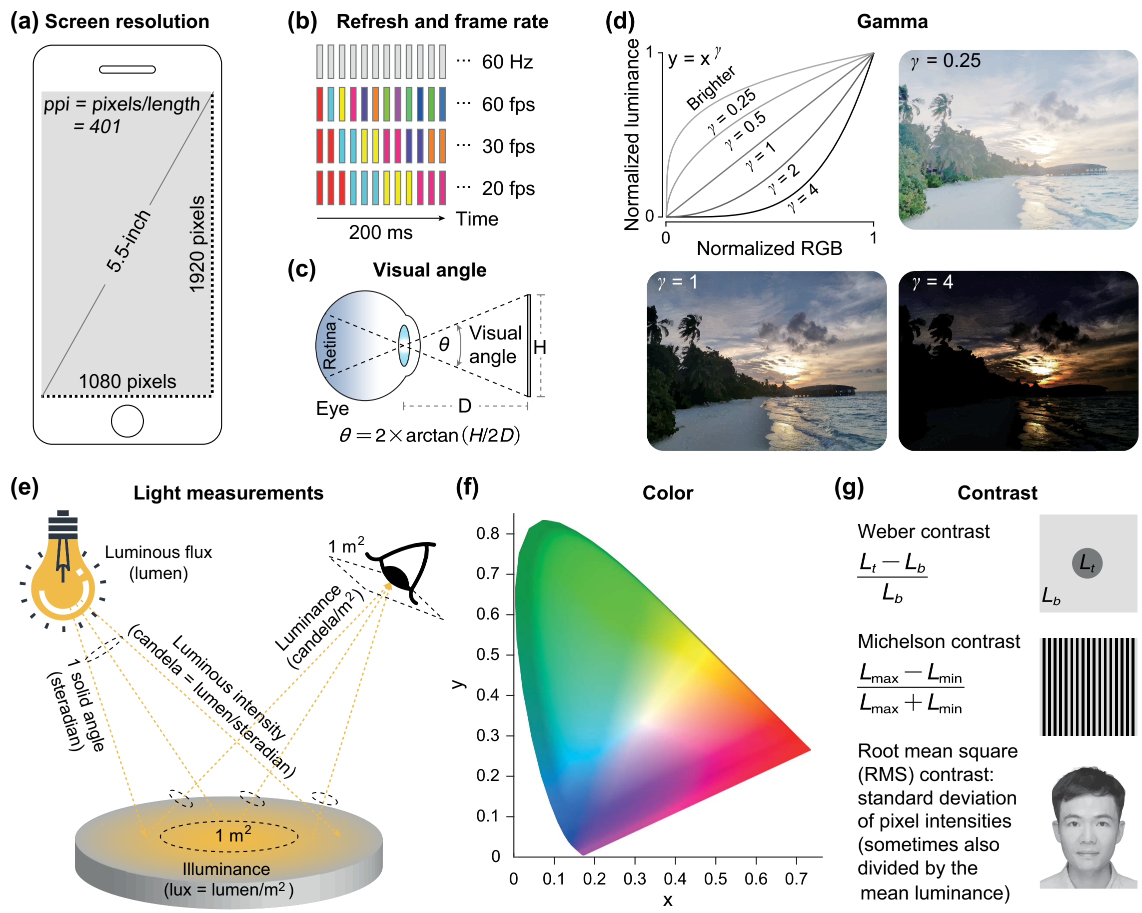
To preview our findings, we found that, while displaying visual stimuli electronically, most studies failed to report visual properties—overall, 8.5% of the articles reported full information and 29.5% reported partial information. These reporting practices—or lack thereof—were widespread, common among the leading journals in psychology and neuroscience. The omission to report stimulus properties may reflect oversight in reporting practices or a tacit assumption on the part of the researchers. In both scenarios, the implicit assumption from the reader's point of view is that the conclusions in the study are robust to variations in stimulus properties. But whether such an assumption is warranted is ultimately an empirical question that requires testing.

While it is impractical—and beyond our scope—to evaluate the validity of the myriad findings covered in the analysis, generally speaking, phenomena in basic perceptual and cognitive domains might be more stimuli driven than those in domains emphasizing deliberate judgment, thinking, and reasoning. In the second study, we examined studies of a basic cognitive phenomenon, namely, visual attentional bias in clinical samples. Attentional bias refers to the preferential allocation of attention to one type of stimuli over another. Attentional bias was chosen as a case study because, among basic cognitive phenomena studied in clinical psychology and neuroscience, it is to our knowledge perhaps the most extensively used. A recent meta-analysis suggests that "attentional bias for positive stimuli occurs rapidly and involuntarily" (Pool et al., 2016, p. 79), implicating *stimulus*-driven processes in attentional bias. Reinforcing the importance of stimulus properties, tasks commonly used in attentional bias have been documented—as discussed previously—to be sensitive to visual properties.

[4] For example, in online apparel retailing, a key challenge is to accurately depict the appearance of apparel despite differing viewing conditions and differing viewing devices.

Figure 2

Key Visual Properties of Displays and Stimuli



Note. (a) Screen resolution describes the spatial resolution of the display, as specified by the number of pixels on the screen (e.g., 1,080 × 1,920); the size of each pixel is measured by pixels per inch (ppi). (b) Refresh rate describes the temporal resolution of a display, as specified by the frequency at which the display refreshes the on-screen image (e.g., 60 Hz); related but different, frame rate describes the rate at which consecutive *static* images (denoted by different colors) appear on a display (e.g., 30 frames per second, fps). (c) Visual angle describes the angular size of a visual stimulus subtended at the eye, as determined by the physical size of the stimulus and its distance from the eye. (d) The γ value of a display specifies the relation between the RGB value or input voltage (x) and the output luminance level (y); lower γ leads to overall brighter images. (e) Luminance describes the amount of visible light reaching the eye from a given direction of space after taking into consideration the eye's sensitivity to light; luminance is closely related to but different from luminous flux, luminous intensity, and illuminance. (f) Color is our perception that distinguishes between different kinds of light based on their spectrums; each color can be specified using three numbers under a given scheme of coordinates called color space; shown on the graph is the *xyY* scheme, where x and y determine the chromaticity (i.e., hue and saturation) of the color and Y, not shown, denotes the luminance of the color. (g) Contrast describes the difference between a stimulus and its background in luminance or color; it is commonly calculated as Weber contrast, Michelson contrast, or root-mean-square (RMS) contrast depending on the type of stimuli. A primer on these concepts is provided in the Appendix; a conceptual and measurement tutorial is also available in Supplemental Materials (https://osf.io/gpn2f; Lin, 2023, February 26). RGB = red, green, and blue. See the online article for the color version of this figure.

The second study shows that publications of attentional bias in clinical samples also failed to report visual properties at a similar rate. As in the first study, 88% of the articles did not share stimuli online, and those articles that did share had a higher ratio of complete reporting (12.7% vs. 6.6%). The failure to report stimulus properties, therefore, represents a hitherto underappreciated and widespread phenomenon in science communication. In the discussion, we outline potential structural reasons for the current practices, explain when reporting stimulus properties is likely to matter, and consider implications for policy and practices.

Method

Transparency and Openness

Data collection and analysis were preregistered for journals in vision science, cognitive psychology, and neuroscience (https://osf.io/9jhrq), but not for journals in clinical, developmental, and social/personality psychology, nor for the attentional bias study. The data set and summary results are available online (https://osf.io/gpn2f; Lin, 2023, February 26). Also included in the online materials are a conceptual and measurement tutorial regarding the technical terms for visual displays and stimuli, a PowerPoint presentation slides, and two MATLAB scripts (one script converts visual angle and the number of pixels from each other; the other script calculates the mean luminance and root-mean-square [RMS] contrast of natural images and equalizes the mean luminance and RMS contrast between images).

Article Selection

In the first study, we sampled leading empirical journals in the respective field. Based on informal discussions with and nominations by psychology colleagues in our departments, 12 journals were identified as preeminent outlets for each field: *Journal of Vision* for vision; *Cognitive Psychology* for cognitive psychology; *Journal of Personality and Social Psychology* for social and personality psychology; *Child Development* and *Developmental Science* for developmental psychology; *Clinical Psychological Science, Journal of Abnormal Psychology, Schizophrenia Bulletin, Depression and Anxiety,* and *International Journal of Eating Disorders* for clinical psychology; and *Nature Neuroscience* and *Neuron* for neuroscience. Note that more journals were included for clinical psychology both because of the diverse, large field of investigation and because of a lack of a clear consensus regarding the top one or two journals. Originally focusing on social, developmental, and clinical psychology, the study was later expanded to include vision, cognitive psychology, and neuroscience. The expansion served two purposes: (a) to assess the scope of the reporting problem identified; and (b) to preregister the study in order to maximize research transparency.

For content analysis, to capture current practices in the respective fields, we selected the latest 30 empirical articles in each journal, with the criteria that the study therein must (a) use visual stimuli/materials on an electronic display; and (b) employ a nonquestionnaire, experimental design. Questionnaires were excluded because they measured self-reported outcomes that generally do not depend on the visual properties of the questions.

Accordingly, under the supervision of the first author and with the assistance of another coder (BC), the second author (A2) searched the full text as well as corresponding supplemental materials (if any) of the 12 journals identified above (eight on and around December 4, 2020; four on and around June 7, 2022). The search started with the latest article in each journal and proceeded in reverse chronological order. The basis for fulfilling the selection criteria—text indicating the involvement of visual stimuli/materials on an electronic display with a nonquestionnaire method—was extracted from each article.

To go beyond the selected journals and to examine phenomena that closely involve stimulus-driven processes, the second study was topic- rather than journal-based, focusing on visual attentional bias in clinical samples. Under the supervision of the first author, in April 2021, two coders (the third and fourth authors, referred to as A3 and A4, respectively) manually searched five databases using a standard systematic search procedure (see Figure 3). As before, the focus was on the latest publications (i.e., those published in 2020 and up to April 9, 2021, when the data were collected), with a total of 114 articles selected.

Article Content Analysis

Basic article-level information was first extracted for each article, including (a) publication year, (b) journal name, (c) article title, and (d) email for correspondence. Next extracted was information concerning the nature of the study, including (a) an excerpt from the article indicating why it was eligible for inclusion; and (b) the specific type of stimuli used (such as words, letters, faces, texts, shapes, images/pictures/photographs, and videos). Each article was then analyzed in two aspects: the reporting of visual stimuli and the sharing of visual stimuli.

Reporting of Visual Stimuli

Reporting of stimuli was evaluated based on the reporting of two types of visual properties. The first was display setup, including the display model number, screen resolution (Figure 2a), and refresh rate (Figure 2b)—information that is relevant but may not be strictly necessary for repeating a given study. The second type was appearance-defining properties, including stimulus size, luminance/color, and contrast—information that is necessary to recreate the stimuli.

As Figure 2c shows, to describe stimulus size—that is, how large the stimulus impinges on the retina of the eye—we need to measure both the physical size of the stimulus and its distance from the viewer. To specify stimulus luminance/color/contrast, the type of stimuli also matters. For simple stimuli with limited variation in luminance (Figure 2e) and color (Figure 2f)—letters, shapes (e.g., dots, lines, triangles, circles), and patterns (e.g., gratings)—the specification is just a few numbers away. Consider the achromatic (i.e., gray scale) stimuli in Figure 2g: the gray circle on a uniform gray background can be specified using two numbers, namely, the luminance of the circle and the background; similarly, a grating—a repeated pattern of dark and bright lines—can be specified by the luminance of the two types of lines. In both cases, luminance values determine the contrast.

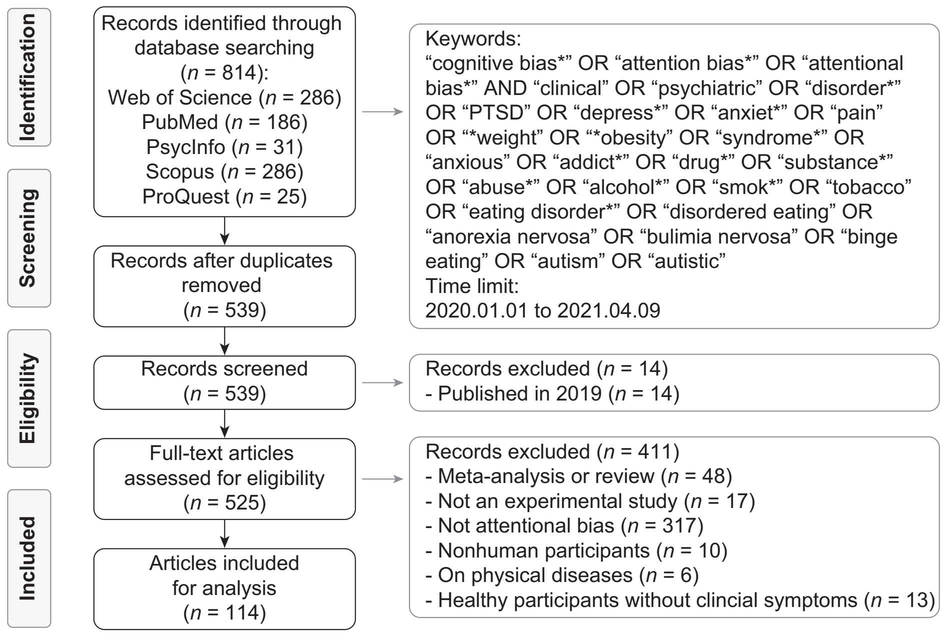
On the other hand, for complex stimuli where the pixels vary greatly in their luminance (and color)—for example, face images (Figure 2g)—an exact specification of luminance for each pixel is not practical. To reproduce or closely recreate the appearance of the stimulus, we need not only access to the actual stimulus (e.g., a picture or video) but also knowledge of the display characteristics, including the γ value (Figure 2d) and minimum and maximum luminance values of the display—or, at a minimum, the display model and image setting (including contrast level and brightness level).

Sharing of Visual Stimuli

As explained above, having access to the original visual stimuli will not allow one to recreate the same visual appearance on a digital display—properties of a digital stimulus (e.g., size, luminance, and contrast) are contingent on display characteristics and specifications. For example, the size of a viewed image, as it subtends at the eye, is determined by both its physical size on the display and its distance from the viewer (Figure 2c; for luminance and contrast, see Figure 2d). Nevertheless, sharing of visual stimuli helps others to replicate the study and convey important (albeit partial) information about the visual appearance. We therefore also examined the sharing of visual stimuli. We evaluated the extent to which the visual materials used in a study were accessible online, focusing on the stimuli of interest in the study while excluding noncritical elements such as fixation, blank screen, and feedback screen.

Figure 3

Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Flowchart of Identifying Recent Experimental Studies on Visual Attentional Bias in Clinical Samples



Note. Five databases were searched based on select keywords and publication time. After screening and eligibility assessment, a total of 114 articles remained. The term "OR" is a logical operator to expand the search by including studies that contain any of the specified terms.

Coding Procedure and Reliability

Coding Procedure

To evaluate the disclosure of visual characteristics and sharing of visual stimuli/materials, four trained coders—A2, A3, A4, and BC—independently combed and encoded the full texts and supplemental materials (if applicable) of all the selected articles (360 in the first study; 114 for the second study).

The completeness of the disclosure in each of the properties—namely, display setup, stimulus luminance/color/contrast, and stimulus size—was independently rated by A2 and BC using a scale of 0–2. When *no* information was mentioned at all regarding the property, it was coded "0"; when *some* information was mentioned but it would not allow readers to fully recover the parameter, it was coded "1," and the relevant information was extracted; when *complete* information was provided or could be calculated, it was coded "2," and the relevant information was also extracted. Examples of ratings of 0, 1, and 2 for each property are shown in Table 1.

The sharing of visual stimuli/materials of interest was independently rated by A2, A3, and A4 using a scale of 0–2. When *no* visual stimuli/materials were shared, it was coded "0"; when *some* but not all visual stimuli/materials were shared, it was coded "1," with the relevant information extracted; when *all* relevant visual stimuli/materials were shared, it was coded "2," with the relevant information also extracted. Discrepancies between raters were resolved through discussions until a consensus was reached.

Note that the rating criteria were overall generous. First, the lack of information regarding the use of generally noncritical visual stimuli, such as fixations and question marks, did not affect the rating. For example, if a study otherwise included full information regarding the stimuli but not the fixation, it would be coded "2" instead of "1." Second, in the same spirit, when information was not provided but can be reasonably inferred, such information would count. For example, when a study indicated that a 10.1-inch Samsung Galaxy Note tablet was used, it was coded "2" instead of "1" because the description points to a specific model (N0020582) that has a specific, fixed screen resolution (1,280 × 800) and refresh rate (60 Hz). Finally, for studies conducted over the internet, the lack of disclosure of information such as screen resolutions and stimuli size did not affect the rating, even though such information can be collected over the internet, such as by matching real object sizes or by exploiting the position of the blind spot of the eye (Li et al., 2020). Instead, it would be coded "NA" (not applicable) in all the properties considered and thus was not included in the statistical analysis. Only when the study involved experiments conducted not over the internet would it be subject to coding of 0–2.

Coder Reliability

To quantify interrater reliability, we compared coding outcomes on the same articles by different raters. For the coding of property reporting, 30 articles were randomly selected and coded by both A2 and BC; for the coding of stimuli/materials sharing, 30 articles were randomly selected and coded by A2, A3, and A4. Interrater reliability was quantified using linearly weighted prevalence-adjusted, bias-adjusted kappa (PABAK, known also as the Brennan–Prediger

Table 1
Rating Scale (0–2) for Basic Properties That Characterize Visual Display and Stimuli

Property	Rating scale with examples		
	0 (no)	1 (partial)	2 (complete)
Display setup (model, resolution, refresh rate)	None	A 22-inch computer screen	A 22-inch LED computer screen Dell P2213; 1,680 × 1,050 pixels; 60 Hz refresh rate
Stimulus luminance, color, contrast (for *simple stimuli*); γ value and minimum and maximum luminance values or image setting of contrast and brightness (for *complex stimuli*)	None	*Simple stimuli*: Gray screen; in black capital letters; maximum contrast (without providing both background and stimulus luminance values) *Complex stimuli*: In grayscale; black background	*Simple stimuli*: Gray screen of 30 cd/m^2; in black capital letters of 0.5 cd/m^2; Michelson contrast ratio of 1.00 *Complex stimuli*: γ value of 1, minimum luminance 0.5 cd/m^2, maximum luminance 220 cd/m^2 (or display contrast adjusted to 80% maximum contrast, brightness to 50% maximum contrast)
Stimulus size	None	Image of 425 × 425 pixels (without providing both the size of pixels and the viewing distance)	Image subtending a visual angle of 4° × 4°

Note. Simple stimuli are stimuli where the luminance and color of the pixels are of limited variation and thus can be characterized using just a few numbers; typical examples include letters, shapes (e.g., dots, lines, triangles, circles), and patterns (e.g., gratings). Complex stimuli are the opposite; typical examples include videos and pictures of rich information (e.g., scenes, real-world photos). LED = light emitting diode.

coefficient). This weighted index accounted for (a) statistical prevalence (i.e., the ratings of 0 and 1 were much more prevalent than ratings of 2), and (b) degree of disagreement (i.e., ratings of 0 and 2 were more different than ratings of 0 and 1 or 1 and 2).

The agreement between the coders was very high across the board. For the coding of property reporting: percentage of agreement = 98.3%, weighted PABAK = 0.96 (95% CI [0.89, 1.0], $p < .001$) for display setup; 93.3% agreement, 0.84 weighted PABAK ([0.61–1], $p < .001$) for luminance and contrast in simple stimuli; 98.9% agreement, 0.97 weighted PABAK ([0.92–1], $p < .001$) for luminance and contrast in complex stimuli; and 98.3% agreement, 0.96 weighted PABAK ([0.89–1], $p < .001$) for size. For the coding of stimuli/materials sharing: percentage of agreement = 95.6%, weighted PABAK = 0.90 (95% CI [0.78, 1.0], $p < .001$).

Results

To evaluate current reporting practices in psychology and neuroscience, in the first study, we sampled leading empirical journals in multiple subfields. We focused on leading journals because publications therein presumably represented best practices in the respective fields. In the second study, we zeroed in on attentional bias in clinical research since it is one of the most extensively studied basic cognitive phenomena, one thought to "[occur] rapidly and involuntarily" (Pool et al., 2016, p. 79). Most research in the analyzed publications was conducted in person (96.7% for the first study, 98.3% for the second study; online studies ranged from 0 to 3 for each journal, as detailed in Table 2). The type of stimuli used was rather diverse, including words, letters, texts, shapes, images (e.g., faces), and videos.

Table 2
Count and Percentage of Online Studies in Individual Journals and on the Topic of Attention Bias

Journal or topic	Total studies	Online studies	
		Count	Percentage (%)
Journal total	360	12	3.3
Journal of Vison	30	1	3.3
Cognitive Psychology	30	0	0
Journal of Personality and Social Psychology	30	0	0
Child Development	30	1	3.3
Developmental Science	30	3	10
Clinical Psychological Science	30	0	0
Journal of Abnormal Psychology	30	0	0
Schizophrenia Bulletin	30	3	10
Depression and Anxiety	30	2	6.7
International Journal of Eating Disorders	30	0	0
Nature Neuroscience	30	2	6.7
Neuron	30	0	0
Attention bias total	120	2	1.7

In addition to evaluating reporting practices, we also examined the extent to which critical visual materials were made available online. Although sharing of visual stimuli only conveyed partial information about the visual appearance of the stimuli—appearance is contingent on display characteristics and user specifications—it nevertheless helps other researchers to replicate and better evaluate the study. Accordingly, the Results section is organized into two parts: reporting practices and sharing of visual materials.

Reporting Practices

Subfields and Journals

The reporting practices of in-person studies were analyzed. The results of the 12 leading journals are shown in Figure 4. The vast majority of the studies failed to report visual properties of the stimuli: Overall, 8.5% of articles reported full information and 29.5% reported partial information. Among the four properties (Figure 4), the complete reporting ratio was highest for stimulus size (Fisher's exact test, $ps < .001$; see Figure 2c and the Appendix), but even then only 18.1% of articles did so, and 17.0% reported partial information—the rest, 64.9%, did not report any information. For display setup (display model, resolution, refresh rate; see Figure 2a and 2b and the Appendix), 6.0% of articles reported full information and 39.4% reported partial information—the rest, 54.6%, did not report any information. The reporting practices were similar for the stimulus luminance/color and contrast (see Figures 2e, 2f, and 2g and the Appendix). For studies using simple stimuli (where the luminance/color of the pixels was of limited variation and could be characterized using a few numbers), 3.4% of articles reported full information and 39.1% reported partial information—the rest, 57.5%, did not report any information. Likewise, for studies using complex stimuli (where the pixels of the stimuli varied greatly in their luminance/color), 2.6% of articles reported full information (either display γ value [Figure 2d; the Appendix] and minimum and maximum luminance values of the display; or, at a minimum, the display model and display image setting of contrast and brightness levels), and 21.2% reported partial information—the rest, 76.2%, did not report any information.

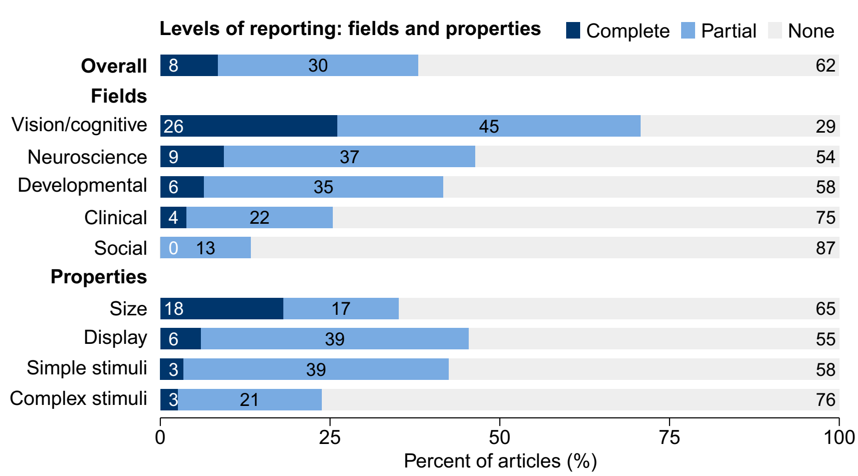
These practices are surprisingly widespread and shared across diverse fields of psychology and neuroscience. As Figure 4 shows, across the four visual properties, complete reporting was highest in vision/cognitive psychology (Fisher's exact test, $ps < .001$), but even then it was only 26.1%, compared with 9.4% in neuroscience, 6.4% in developmental psychology, 3.9% in clinical psychology, and 0.0% in social psychology (for the detailed breakdown to each property, see Supplemental Table S1). Similarly, at the journal level, as Figure 5 shows, complete reporting was highest for *Journal of Vision*, at 39.8%, and lowest for *Clinical Psychological Science*, *Depression and Anxiety*, and *Journal of Personality and Social Psychology*, at 0.0% (for the detailed breakdown to each property, see Supplemental Table S2).

Studies of Attention Bias

It is conceivable that different phenomena may vary in their dependence on stimulus properties. For example, phenomena in basic perceptual and cognitive domains might be driven more by stimuli than phenomena in domains that emphasize deliberate judgment, thinking, and reasoning. Therefore, in the second study,

Figure 4

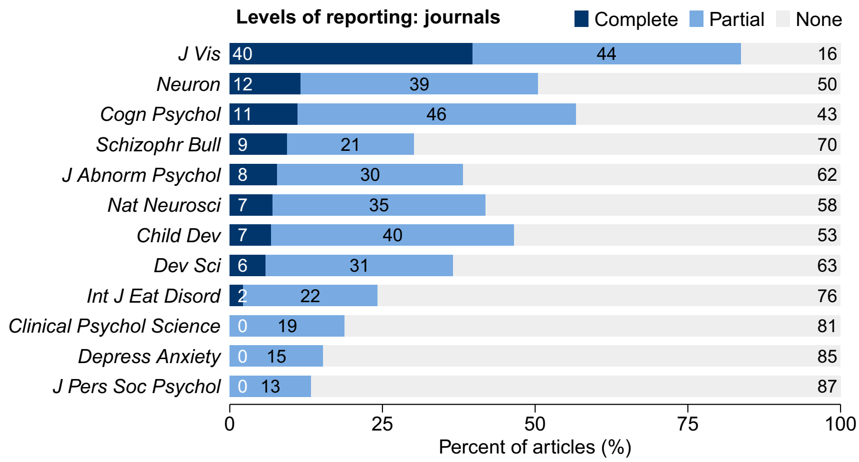
Overall Reporting Practices of Visual Display and Stimulus Properties in 12 Leading Journals in Psychology and Neuroscience and Specific Practices at the Levels of Subfield and Property



Note. The reporting practice in each subfield is tallied across the four properties (display, luminance/color/contrast for simple stimuli [referred to as "simple stimuli" in the figure and hereafter], luminance/color/contrast for complex stimuli [referred to as "complex stimuli" in the figure and hereafter], and size); the practice in each property is tallied across all journals. See the online article for the color version of this figure.

Figure 5

Reporting Practices of Visual Display and Stimulus Properties Across the 12 Leading Journals in Psychology and Neuroscience



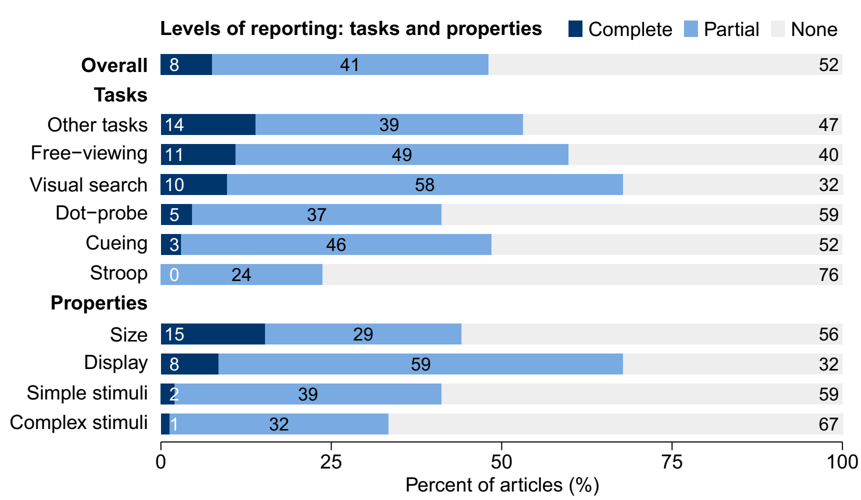
Note. The reporting practice in each journal is tallied across four properties (display, simple stimuli, complex stimuli, and size). See the online article for the color version of this figure.

we asked whether—and to what extent—the omission of communicating stimulus properties identified above transpired in studies of a basic cognitive phenomenon, specifically, visual attentional bias in clinical samples.

As Figure 6 shows, the majority of the studies failed to report visual properties of the stimuli: Overall, 7.5% of articles reported full information (similar to the overall journal results; Fisher's exact test, $p = .82$) and 40.6% reported partial information (higher than the overall journal results; Fisher's exact test, $p < .001$). Among the four properties, complete reporting of stimulus size and display setup was higher than that of luminance/color and contrast for simple and complex stimuli (Fisher's exact test, $ps < .009$). Specifically, for stimulus size, 15.3% of articles reported full information and 28.8% reported partial information—the rest, 55.9%, did not report any information. For display setup, 8.5% of articles reported full information and 59.3% reported partial information—the rest, 32.2%, did not report any information. For studies using simple stimuli, 1.3% of articles reported full information and 32.1% reported

Figure 6

Overall Reporting Practices of Visual Display and Stimulus Properties in Attentional Bias and Specific Practices at the Levels of Task and Property



Note. The reporting practice in each subfield is tallied across the four properties (display, simple stimuli, complex stimuli, and size); the practice in each property is aggregated across all articles. See the online article for the color version of this figure.

partial information—the rest, 66.7%, did not report any information. Likewise, for studies using complex stimuli, 2% of articles reported full information and 39.2% reported partial information—the rest, 58.8%, did not report any information.

Various tasks were used in studies of attentional bias in clinical samples, which were categorized into six types: dot-probe task, free-viewing task, Stroop task, cueing task, visual search task, and others (i.e., the rest of the tasks; for a breakdown of these tasks, see Supplemental Table S3). As Figure 6 shows, reporting practices—or lack thereof—were shared among the six categories of tasks, with similar levels of complete reporting (Fisher's exact test, $ps > .07$). Across the four visual properties, complete reporting was 13.9% in the category of other tasks, 11.0% in free-viewing, 9.7% in visual search, 4.6% in dot-probe, 3.0% in cueing, and 0.0% in Stroop task (for the detailed breakdown to each property, see Supplemental Table B4).

Sharing of Visual Materials

So far, we have documented pervasive failure to report visual properties. Next, we turn to the issue of sharing critical visual materials online. As Figure 7a shows, sharing of visual materials was rare: 81.4% of the articles in the 12 journals and 87.5% of the articles on attention bias did not make visual materials available online (for the detailed breakdown of each property, see Supplemental Table S4). Failing to share visual materials was not linked to a higher degree of stimulus property reporting. If anything, as Figure 7b shows, for the topic of attention bias, the level of complete reporting was higher for those articles sharing critical stimuli online (either partial or complete; 12.7%) than those that did not (6.6%; Fisher's exact test, $p = .04$). The level of complete reporting was not affected by the status of sharing visual stimuli in the 12 journals surveyed in the first study (Figure 7b).

Discussion

In two studies, we identify a prevalent failure to report visual stimulus properties in current research practices, as evidenced by (a) publications in premier journals across psychology and neuroscience and (b) studies investigating a basic cognitive phenomenon in clinical samples. This failure to report stimulus properties is not due to authors sharing stimuli online—in fact, more than 80% of the surveyed articles did not make stimuli available online. Moreover, simply sharing stimuli cannot replace reporting visual properties, as appearance depends on display characteristics and user specifications, which are independent of the stimuli.

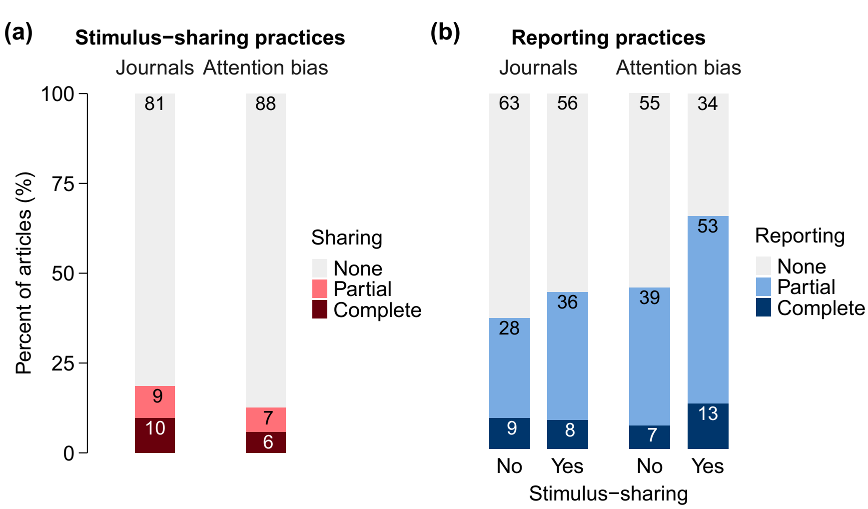
As alluded to in the Introduction, the potential impacts of these practices likely depend on the research topics, phenomena, questions, and measures at hand. It is therefore prudent for researchers to consider the impact of stimulus properties on a case-by-case basis, and not to dismiss such possibilities without a solid empirical foundation. Below, we first briefly consider potential causes for the widespread failure to report visual properties, and then delve into the implication of this practice by outlining when failure to report stimulus properties is likely to matter. Last, to enhance reporting transparency and consequently improve research replicability and validity, we propose practical implications—including reporting templates—for authors, editors, and journals.

Explaining Widespread Failure to Report Stimulus Properties

The culprit for the widespread failure to disclose stimulus properties in current reporting practices may lie in a lack of awareness and the force of inertia. Much like the issue of transparency in data, code, and research materials, which was largely overlooked in

Figure 7

Sharing of Visual Stimuli in the 12 Journals (360 Articles) and on the Topic of Attention Bias (114 Articles) and Reporting Practices as a Function of Whether or Not Making Visual Stimuli Available Online



Note. Each number within the bar plots represents the exact ratio of the associated sharing or reporting level (in %). The "yes" condition includes both partial and complete sharing of critical visual stimuli. See the online article for the color version of this figure.

research practices prior to the recent open science movement (e.g., Gernsbacher, 2018; Lindsay, 2017; Munafo et al., 2017; Nosek et al., 2015), the importance of reporting stimulus properties has remained unnoticed in the current zeitgeist (see Supplemental Box S1).

For example, the acclaimed TOP guidelines (Nosek et al., 2015) for journal policy focus on eight aspects of a article: citation, data, analytic methods (code), research materials, design and analysis, study preregistration, analysis plan preregistration, and replication. Echoing the TOP guidelines, Munafo et al. (2017) emphasized the importance of making "data, materials, and code supporting research outcomes" accessible (pp. 5–6). Similarly, in an editorial for the journal *Psychological Science*, Lindsay (2017) focused on the importance of "sharing of data and materials" (pp. 700–701). Likewise, in a writing guideline, Gernsbacher (2018) highlighted the value of making "materials and data publicly available" (p. 406). Unfortunately, these influential guidelines largely overlook the issue of how materials such as visual stimuli have actually been used or presented in the study. This oversight could stem from neglecting the fact that the appearance of visual stimuli depends not only on the stimuli themselves, but also on their presentation on the digital display—as quantified by visual properties such as contrast.

When Does Reporting Stimulus Properties Matter?

Perceptual Argument

Stimulus properties such as luminance, contrast, and size dictate the visual appearance of a stimulus.[5] Thus, any phenomenon involving a perceptual component or an interaction with perception naturally requires the reporting of these stimulus properties. Although perceptual domains are an obvious area of consideration, as Figure 1 illustrates, perception also plays a role in cognitive, affective, and social processes.

Expanding upon the diverse examples from the Introduction (e.g., Figure 1), we delve further into the perceptual argument by focusing on a general research topic: working memory (e.g., Pasternak & Greenlee, 2005). Studies in this field often use color as the memory feature (W. Zhang & Luck, 2008), making a thorough understanding of color working memory integral to theories of working memory. Critical to this line of work is the careful characterization of color properties, which, at times, has been inadequate. Indeed, without proper display calibration, rendered colors can diverge significantly from the intended nominal colors (e.g., "red"), including in luminance (Bae et al., 2014). Failure to calibrate displays and carefully consider stimulus properties compromise the validity of working memory research, as the stimuli seen by the participants may differ from those they are intended to see; moreover, response variability across stimuli may not accurately reflect memory precision, as often assumed (Allred & Flombaum, 2014).

From this perceptual perspective, one might assume that characteristics of visual stimuli are most theoretically relevant in the context of basic perceptual and cognitive domains, with their relevance diminishing in domains such as social psychology. However, even in social or other high-level processes involving visual stimuli, the relevance of stimulus properties, diminished as it may be, is not entirely eliminated. For example, a common research approach to investigating high-level processes is to compare different categories of images and measure some behavioral and/or neural indexes. Thus, by contrasting self-face images with other-face images, researchers attempt to find out whether self-face images preferentially, automatically, and unconsciously attract attention. By pitting food images against nonfood images, researchers seek to learn how food and nonfood information is processed differently among different subgroups of participants, such as those with varying statuses in obesity, eating disorders, body image concerns, and so on. A crucial assumption of these comparisons is that differences in low-level properties (e.g., luminance and contrast) cannot explain the differences observed at the categorical level. But without reporting and controlling for these stimulus properties, the validity of the conclusions cannot be adequately established (Knebel et al., 2008; Willenbockel et al., 2010).

In general, the specific effect of stimulus properties on diverse psychological and neural phenomena should be evaluated on a case-by-case basis. Using the finding of Figure 1b as an example, if absolute performance is important, then the exact contrast would be a critical parameter to measure and specify, as it greatly affects absolute performance. Conversely, if relative performance is critical (such as comparing two array sizes), the exact contrast seems less critical, provided that the contrast remains the same between the conditions being compared. Similarly, behavioral measures (e.g., accuracy) and neural measures (e.g., contralateral delay activity in event-related potentials) may show different sensitivities to stimulus properties such as luminance and contrast, as in the case of working memory (Ikkai et al., 2010).

The perceptual argument thus distinguishes stimulus properties from the nonstimulus background of an experiment, which can range from the exact room, floor, and time of day of the experiment to the physical appearance and demeanor of the experimenter. That is, whereas stimulus properties specify the very materials used to evoke relevant processes and behaviors, the nonstimulus information typically does not. This distinction is made regardless of whether the nonstimulus information *can* affect the outcome. Even if factors such as the location and time of the experiment or characteristics of the experimenter influence the measured outcome—such as the participant's attentional bias to the visual stimuli on the screen—the effect of these nonstimulus factors must be indirect or secondary, by affecting the participant, rather than through direct perceptual means.

Epistemic Argument

That we cannot completely rule out a priori the relevance of stimulus properties in nonperceptual phenomena compels us to consider the epistemic argument for specifying these properties. That is, without testing, we cannot empirically determine whether a nonperceptual phenomenon demonstrated with visual stimuli is free from stimulus-level confounds and can generalize across variations in stimulus properties. It may be that these propositions often survive empirical testing (e.g., Olsson-Collentine et al., 2020). But we cannot definitively know this beforehand when evaluating a given

[5] In this sense, stimulus properties can be considered essential for specifying the appearance of the stimulus. This is independent of whether visual appearance is critical to the robustness of the phenomenon at hand or not. Even when visual appearance turns out to be of little importance to that phenomenon—and thus stimulus properties may be accidental properties in the structure of the phenomenon—this does not refute the intrinsic relationship between stimulus properties and stimulus appearance.

phenomenon, as the relevance of perceptual processes may surface in unexpected ways.

This epistemic argument is not just playing the possibility card. For example, when researchers compare images varying in emotion, self-identity, race, and other high-level constructs, low-level stimulus properties may be deemed theoretically irrelevant. However, to the extent that differences at the image level may pose a confound (Knebel et al., 2008; Willenbockel et al., 2010) and thus compromise the internal validity of the studied construct, this methodological threat becomes theoretically relevant.

The epistemic uncertainty is not eliminated through experimental designs such as using within-subject comparisons. Consider an experiment comparing two conditions with identical or highly similar display setups. Even without considering stimulus confounds, the finding may be limited to and dependent on the specific parameters of the display. Imagine a study that compares a dot-probe task and a visual search task in a within-subject design. Without empirical evidence, there is no guarantee that a similar finding will emerge with a different monitor or setup (e.g., due to potential interaction effects with stimulus contrast).

Thus, the epistemic argument draws an analogy between specifying stimulus properties in studies using visual stimuli and counterbalancing in studies involving condition ordering. The order of conditions may turn out to not affect the outcome in most cases, but counterbalancing helps empirically rule out a potential order effect. Similarly, demographic variables like gender may not modulate the outcome in most cases (Delios et al., 2022; Olsson-Collentine et al., 2020), but reporting demographic composition is crucial (Henrich et al., 2010)—as effects may surface in ostensibly unlikely areas, such as gender differences in visual motion processing (Murray et al., 2018).

While drawing these analogies, we do not mean to imply that reporting stimulus properties should be mandated across all research domains. Our primary purpose, rather, is to draw attention to the potential impacts of stimulus properties, possibly in unsuspected areas—including in studies of high-level processes using visual stimuli. By demonstrating how stimulus characteristics can influence replicability, internal validity, and external validity (generalizability), we hope to help researchers make more informed decisions. Ultimately, we believe it is up to the community of researchers in each field to reach a consensus on reporting standards.

Implications for Reporting Practices and Journal Policy

What are the implications of the finding? At the policy level, in addition to the rightful emphasis on sharing data, code, and research materials, journals may consider requiring transparent reporting of how materials are used and presented in research. The nature of this reporting will depend on the specific materials involved. In the case of visual stimuli, to foster transparency and reproducibility in reporting and to facilitate community discussion and consensus building, Table 3 provides reporting templates for (a) the general study setup; (b) six common types of visual displays (monitor, laptop, tablet, projector, goggle, and online/over the internet presentation); and (c) two primary types of visual stimuli (simple and complex). For *study setup*, reporting may include viewing distance, lighting conditions, and the use of a stability device for head movement. For *visual display*, reporting usually entails the display setup (model number, screen resolution, refresh rate, and γ correction if any). For *visual stimuli*, reporting may include details about stimulus properties such as luminance/contrast (or display image settings for video clips and complex images) and size (in visual angle). Additional information, such as software and hardware, can also be provided.

Note that Table 3 is intended to serve as a useful guide for authors and editors, rather than a strict prescription. Indeed, Table 3 makes it clear that reporting requirements can depend on the nature of the study. For example, when using iPads, it is beneficial to record major factors that impact visual appearance, such as battery level (Bodduluri et al., 2017), the time since the device has been active (Aslam et al., 2013), and status of special features like autobrightness, autolock, True Tone, and Night Shift. These features differ from parameters that are typically fixed, such as screen resolution, screen γ (around 2.2; Bodduluri et al., 2017), and refresh rate (excluding devices with the ProMotion feature, which adaptively changes refresh rate). Moreover, in developing Table 3, we have erred on the side of comprehensiveness. For example, including information on hardware models (e.g., Dell XPS 8700) can facilitate replication because issues such as software compatibility might arise from different hardware being used.

The majority of the properties in Table 3—such as model number, resolution, refresh rate, and size—are relatively straightforward to obtain through button clicking and simple calculations. For example, visual angle can be calculated based on stimulus size and viewing distance (Figure 2c). However, color and luminance measurements typically require specialized tools known as photometers (e.g., a colorimeter or a more advanced spectrophotometer). While these specialized photometers can be complex to operate and expensive, most research in psychology and neuroscience only requires luminance/color specification and screen calibration, which can be accomplished with relatively simple hardware—a consumer colorimeter will suffice in most cases. Therefore, we have recently developed a free, open-source software package for low-cost photometers (e.g., SpyderX from Datacolor, Inc., New Jersey, United States) that simplifies and automates screen calibration and luminance/color measurements (Lin et al., 2023).

Aside from the technical issue, there might be concerns that reporting stimulus properties could lengthen the article. While the article length was a genuine concern in the predigital era, the advent of online materials largely addresses this issue. Details on stimulus properties can be included in online materials to avoid overloading the main article. But fundamentally, when stimulus properties are integral to a study, it behooves us to report them—even at the cost of adding more ink to the article. For example, it allows easy access to such information without requiring other researchers to request it—a process that has proven to be unreliable (Houtkoop et al., 2018; Martone et al., 2018). The success rate for authors providing stimulus properties is likely to be even lower, given that such properties often need to be measured and calculated and may be irretrievably lost after the study's completion (e.g., if computer settings have been changed). Thus, measuring and reporting stimulus properties beforehand not only fulfills our responsibilities as authors but also saves us time and frustration.

To move the field forward, future research may pursue two directions. The first is to build on the current observations in the visual modality by assessing reporting practices in other modalities, each of which has unique properties to report and may follow different practices. Furthermore, validity is an empirical question, and future research should investigate the internal validity and generalizability of various findings across stimulus properties.

Table 3
Reporting Templates of Study Setup, Visual Displays, and Visual Stimuli

Type	Subtype	Reporting template	Key properties (italicized in the reporting template)
Setup	N/A	Participants viewed the stimuli from *57 cm* away in a *dimly lit* room using a *chinrest* to stabilize the head position.	Distance, lighting, and stability device for head movement
Display	Monitor	Stimuli were presented using *MATLAB* (MathWorks) with extensions from *Psychtoolbox* and on a *PC* running *Windows 7* with an *NVidia graphics card (GeForce GTX 750Ti)*. The stimuli were displayed on a *γ-corrected 19-inch CRT monitor* (*EIZO FlexScan T766*, resolution of *1,024 × 768* pixels, refresh rate of *120 Hz*).	Model number, resolution, refresh rate, γ correction, software, hardware
	Laptop screen	Stimuli were presented using *Presentation* (Neurobehavioral Systems, https://www.neurobs.com) on a *Sony VAIO (PCG-TR5MP)* laptop computer with an LCD screen (resolution of *1,280 × 768* pixels, refresh rate of *60 Hz*, *32-bit color*).	Model number, resolution, refresh rate, color depth, software
	Tablet	Stimuli were presented on an iPad (model *A2316*, Apple Inc., Cupertino, California, United States; resolution of *2,360 × 1,640* pixels, a refresh rate of *60 Hz*) after it was turned on for at least *15 min* and with a battery level of at least *10%*. It was set at *85%* of screen brightness (luminance: *250 cd/m*^{2}, measured with a Spyder4Elite colorimeter). *Autobrightness* was disabled to ensure consistent brightness irrespective of ambient light. *True Tone* and *Night Shift* were both disabled to maintain consistent color characteristics. *Autolock* was set to never.	Model number, resolution, refresh rate, time since power on, battery level, brightness, feature setups
	Projector	Stimuli were presented through a *γ-corrected* LCD projector (*EPSON EMP-710*; resolution of *1,024 × 768* pixels, a refresh rate of *60 Hz*), as controlled by a PC (*Apple Power Mac G4*).	Model number, resolution, refresh rate, γ correction, hardware
	Goggles	Stimuli were presented on a pair of *γ-corrected* head-mounted goggles (*Sony HMZ-T3*; goggle screen size of *49.4° × 27.8°*, resolution of *1,280 × 720* pixels, a refresh rate of *60 Hz*), using *MATLAB Psychtoolbox* on a PC (*Dell XPS 8700*). γ correction was conducted by a spectrophotometer (*Photo Research PR-655*).	Model number, field of view, resolution, refresh rate, γ correction, software, hardware
	Online/over the internet	Tasks were presented using *Classic ASP* and *JavaScript (version ES6)* and displayed on an external *website* managed by a server. Screen resolutions and display sizes were unknown but were estimated by placing a real credit card to match an adjustable card image on the screen. Distances to the screen were also unknown but were estimated using trigonometry based on the detection of the viewer's blind spot location (Li et al., 2020).	Software, hardware
Stimuli	Simple stimuli	Two equal luminance red (*CIE xyY space: 0.6, 0.36, 7.62 cd/m*^{2}) and green (as measured in a minimum motion technique described elsewhere) colors were used to create isoluminant chromatic flicker. The flicker consisted of two alternating horizontal chromatic Gabors that were counterphased at *30 Hz* (*4° × 4°*, red and green bars of *0.8 cycles per degree*; the luminance functions of the red and green phosphors are given by separate *equations* described elsewhere). Adapted from: F. Zhang et al. (2021)	Color, size, spatial and temporal frequencies, shape, and form
	Complex stimuli	A total of 40 grayscale frontal-view faces with neutral expressions were selected from a *database* (described elsewhere). Twenty were female and 20 were male; all were Asian. The images were first resized to maintain the same interpupil distance (*80 pixels*) and then cropped to the same size (*320 × 420 pixels*, subtending *6° × 8°* from the viewing distance of *60 cm* away). The background was rendered gray (RGB of [128, 128, 128], luminance of *80 cd/m*^{2}). The display was *γ-corrected*, with a maximum luminance of *200 cd/m*^{2} and a minimum luminance of *0.05 cd/m*^{2}.	Source, modification, size, display γ correction, display maximum luminance and minimum luminance

Note. N/A = not applicable; PC = personal computer; CRT = cathode ray tube; LCD = liquid crystal display; ASP = Active Server Pages; CIE = Commission Internationale de l'Éclairage (that is, International Commission on Illumination); RGB = red, green, and blue.

Conclusion

In summary, we have identified a systematic neglect to report visual characteristics in current research practice in psychology and neuroscience. We demonstrate how the reporting of visual characteristics can influence the evaluation of replicability, internal validity (e.g., stimulus confound), and external validity (i.e., generalizability). Specifically, the failure to report visual characteristics can pose specific impediments to the accumulation of scientific knowledge: (a) it can make direct replications unnecessarily burdensome; (b) it can complicate the assessment of potential confounds related to stimulus properties; and (c) it can lead to unverified assumptions about generalizability across different stimulus properties, while simultaneously hampering a thorough evaluation of generalizability itself. The impact of not reporting visual characteristics will likely vary across research domains, questions, and measures, and should be evaluated on a case-by-case basis. Reporting stimulus properties is particularly crucial for phenomena involving a perceptual component or some interaction with perception—phenomena studied not only in perceptual and cognitive research but also in affective and social research. Given the empirical uncertainty, it is also epistemically valuable to consider and evaluate the impact of stimulus properties on a wide variety of phenomena.

Whether and how the reporting of stimulus properties should be mandated is a decision that each research community must make. Indeed, achieving consensus in each field can better clarify nuances. To enhance reporting standards, we suggest that (a) journals consider adopting explicit policies on transparency and reproducibility regarding how materials are used and presented in research, and (b) authors consider adequately documenting stimulus properties in their research reports or provide a justification for not doing so (e.g., based on theoretical rationales or empirical findings).

To support this move toward more transparent stimulus-reporting practices and to facilitate consensus building, we (a) provide reporting templates for study setup, visual displays, and visual stimuli (Table 3); (b) address potential technical challenges and space concerns; and (c) offer a conceptual and measurement tutorial in the Supplemental Materials that complements the Appendix to explain technical terms for visual displays and stimuli.

References

Aczel, B., Szaszi, B., Sarafoglou, A., Kekecs, Z., Kucharský, Š., Benjamin, D., Chambers, C. D., Fisher, A., Gelman, A., Gernsbacher, M. A., Ioannidis, J. P., Johnson, E., Jonas, K., Kousta, S., Lilienfeld, S. O., Lindsay, D. S., Morey, C. C., Munafò, M., Newell, B. R., … Wagenmakers, E. J. (2020). A consensus-based transparency checklist. *Nature Human Behaviour*, *4*(1), 4–6. https://doi.org/10.1038/s41562-019-0772-6

Allred, S. R., & Flombaum, J. I. (2014). Relating color working memory and color perception. *Trends in Cognitive Sciences*, *18*(11), 562–565. https://doi.org/10.1016/j.tics.2014.06.002

Aslam, T. M., Murray, I. J., Lai, M. Y., Linton, E., Tahir, H. J., & Parry, N. R. (2013). An assessment of a modern touch-screen tablet computer with reference to core physical characteristics necessary for clinical vision testing. *Journal of the Royal Society, Interface*, *10*(84), Article 20130239. https://doi.org/10.1098/rsif.2013.0239

Bae, G. Y., Olkkonen, M., Allred, S. R., Wilson, C., & Flombaum, J. I. (2014). Stimulus-specific variability in color working memory with delayed estimation. *Journal of Vision*, *14*(4), Article 7. https://doi.org/10.1167/14.4.7

Baribault, B., Donkin, C., Little, D. R., Trueblood, J. S., Oravecz, Z., van Ravenzwaaij, D., White, C. N., De Boeck, P., & Vandekerckhove, J. (2018). Metastudies for robust tests of theory. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(11), 2607–2612. https://doi.org/10.1073/pnas.1708285114

Bayer, M., Sommer, W., & Schacht, A. (2012). Font size matters—Emotion and attention in cortical responses to written words. *PLOS ONE*, *7*(5), Article e36042. https://doi.org/10.1371/journal.pone.0036042

Betts, L. R., Taylor, C. P., Sekuler, A. B., & Bennett, P. J. (2005). Aging reduces center-surround antagonism in visual motion processing. *Neuron*, *45*(3), 361–366. https://doi.org/10.1016/j.neuron.2004.12.041

Bodduluri, L., Boon, M. Y., & Dain, S. J. (2017). Evaluation of tablet computers for visual function assessment. *Behavior Research Methods*, *49*(2), 548–558. https://doi.org/10.3758/s13428-016-0725-1

Bosnjak, M., Fiebach, C. J., Mellor, D., Mueller, S., O'Connor, D. B., Oswald, F. L., & Sokol, R. I. (2022). A template for preregistration of quantitative research in psychology: Report of the joint psychological societies preregistration task force. *American Psychologist*, *77*(4), 602–615. https://doi.org/10.1037/amp0000879

Butler, P. D., Abeles, I. Y., Weiskopf, N. G., Tambini, A., Jalbrzikowski, M., Legatt, M. E., Zemon, V., Loughead, J., Gur, R. C., & Javitt, D. C. (2009). Sensory contributions to impaired emotion processing in schizophrenia. *Schizophrenia Bulletin*, *35*(6), 1095–1107. https://doi.org/10.1093/schbul/sbp109

Chambers, C. D., & Tzavella, L. (2022). The past, present and future of registered reports. *Nature Human Behaviour*, *6*(1), 29–42. https://doi.org/10.1038/s41562-021-01193-7

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, *12*(4), 335–359. https://doi.org/10.1016/S0022-5371(73)80014-3

Cross, E. S., Ramsey, R., Liepelt, R., Prinz, W., & de C Hamilton, A. F. (2016). The shaping of social perception by stimulus and knowledge cues to human animacy. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *371*(1686), Article 20150075. https://doi.org/10.1098/rstb.2015.0075

De Boeck, P., & Jeon, M. (2018). Perceived crisis and reforms: Issues, explanations, and remedies. *Psychological Bulletin*, *144*(7), 757–777. https://doi.org/10.1037/bul0000154

DeKay, M. L., Rubinchik, N., Li, Z., & De Boeck, P. (2022). Accelerating psychological science with metastudies: A demonstration using the risky-choice framing effect. *Perspectives on Psychological Science*, *17*(6), 1704–1736. https://doi.org/10.1177/17456916221079611

Delios, A., Clemente, E. G., Wu, T., Tan, H., Wang, Y., Gordon, M., Viganola, D., Chen, Z., Dreber, A., Johannesson, M., Pfeiffer, T., Uhlmann, E. L., Abd Al-Aziz, A. M., Abraham, A. T., Trojan, J., Adamkovic, M., Agadullina, E., Ahn, J., Akinci, C., … the Generalizability Tests Forecasting Collaboration. (2022). Examining the generalizability of research findings from archival data. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(30), Article e2120377119. https://doi.org/10.1073/pnas.2120377119

Delplanque, S., N'diaye, K., Scherer, K., & Grandjean, D. (2007). Spatial frequencies or emotional effects? A systematic measure of spatial frequencies for IAPS pictures by a discrete wavelet analysis. *Journal of Neuroscience Methods*, *165*(1), 144–150. https://doi.org/10.1016/j.jneumeth.2007.05.030

Durgin, F. H., Klein, B., Spiegel, A., Strawser, C. J., & Williams, M. (2012). The social psychology of perception experiments: Hills, backpacks, glucose, and the problem of generalizability. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(6), 1582–1595. https://doi.org/10.1037/a0027805

Dyer, F. N. (1973). The Stroop phenomenon and its use in the stlldy of perceptual, cognitive, and response processes. *Memory & Cognition*, *1*(2), 106–120. https://doi.org/10.3758/BF03198078

Fotios, S., Yang, B., & Cheal, C. (2015). Effects of outdoor lighting on judgements of emotion and gaze direction. *Lighting Research & Technology*, *47*(3), 301–315. https://doi.org/10.1177/1477153513510311

Frank, M. C., & Saxe, R. (2012). Teaching replication. *Perspectives on Psychological Science*, *7*(6), 600–604. https://doi.org/10.1177/1745691612460686

Gernsbacher, M. A. (2018). Writing empirical articles: Transparency, reproducibility, clarity, and memorability. *Advances in Methods and Practices in Psychological Science*, *1*(3), 403–414. https://doi.org/10.1177/2515245918754485

Goffaux, V., & Rossion, B. (2006). Faces are "spatial"—Holistic face perception is supported by low spatial frequencies. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(4), 1023–1039. https://doi.org/10.1037/0096-1523.32.4.1023

Golomb, J. D., McDavitt, J. R., Ruf, B. M., Chen, J. I., Saricicek, A., Maloney, K. H., Hu, J., Chun, M. M., & Bhagwagar, Z. (2009). Enhanced visual motion perception in major depressive disorder. *The Journal of Neuroscience*, *29*(28), 9072–9077. https://doi.org/10.1523/JNEUROSCI.1003-09.2009

Hawkins, H. L., Shafto, M. G., & Richardson, K. (1988). Effects of target luminance and cue validity on the latency of visual detection. *Perception & Psychophysics*, *44*(5), 484–492. https://doi.org/10.3758/BF03210434

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2–3), 61–83. https://doi.org/10.1017/S0140525X0999152X

Herrmann, K., Montaser-Kouhsari, L., Carrasco, M., & Heeger, D. J. (2010). When size matters: Attention affects performance by contrast or response gain. *Nature Neuroscience*, *13*(12), 1554–1559. https://doi.org/10.1038/nn.2669

Houtkoop, B. L., Chambers, C., Macleod, M., Bishop, D. V. M., Nichols, T. E., & Wagenmakers, E. J. (2018). Data sharing in psychology: A survey on barriers and preconditions. *Advances in Methods and Practices in Psychological Science*, *1*(1), 70–85. https://doi.org/10.1177/2515245917751886

Ikkai, A., McCollough, A. W., & Vogel, E. K. (2010). Contralateral delay activity provides a neural measure of the number of representations in visual working memory. *Journal of Neurophysiology*, *103*(4), 1963–1968. https://doi.org/10.1152/jn.00978.2009

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, *23*(5), 524–532. https://doi.org/10.1177/0956797611430953

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology*, *103*(1), 54–69. https://doi.org/10.1037/a0028347

Knebel, J. F., Toepel, U., Hudry, J., le Coutre, J., & Murray, M. M. (2008). Generating controlled image sets in cognitive neuroscience research. *Brain Topography*, *20*(4), 284–289. https://doi.org/10.1007/s10548-008-0046-5

Kumar, D., & Srinivasan, N. (2011). Emotion perception is mediated by spatial frequency content. *Emotion*, *11*(5), 1144–1151. https://doi.org/10.1037/a0025453

Li, Q., Joo, S. J., Yeatman, J. D., & Reinecke, K. (2020). Controlling for participants' viewing distance in large-scale, psychophysical online experiments using a virtual chinrest. *Scientific Reports*, *10*(1), Article 904. https://doi.org/10.1038/s41598-019-57204-1

Lin, Z. (2023, February 26). *Pervasive failure to report properties of visual stimuli in experimental research in psychology and neuroscience*. https://osf.io/gpn2f

Lin, Z., & Li, N. (2023). Global diversity of authors, editors, and journal ownership across subdisciplines of psychology: Current state and policy implications. *Perspectives on Psychological Science*, *18*(2), 358–377. https://doi.org/10.1177/17456916221091831

Lin, Z., Ma, Q., & Zhang, Y. (2023). Psycalibrator: An open-source package for display gamma calibration and luminance and color measurement. *Advances in Methods and Practices in Psychological Science*, *6*(2). https://doi.org/10.1177/25152459221151151

Lindsay, D. S. (2017). Sharing data and materials in psychological science. *Psychological Science*, *28*(6), 699–702. https://doi.org/10.1177/0956797617704015

Martone, M. E., Garcia-Castro, A., & VandenBos, G. R. (2018). Data sharing in psychology. *American Psychologist*, *73*(2), 111–125. https://doi.org/10.1037/amp0000242

Munafo, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., du Sert, N. P., Simonsohn, U., Wagenmakers, E.-J., Ware, J. J., & Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, *1*, Article 0021. https://doi.org/10.1038/s41562-016-0021

Murray, S. O., Schallmo, M. P., Kolodny, T., Millin, R., Kale, A., Thomas, P., Rammsayer, T. H., Troche, S. J., Bernier, R. A., & Tadin, D. (2018). Sex differences in visual motion processing. *Current Biology*, *28*(17), 2794–2799.e3. https://doi.org/10.1016/j.cub.2018.06.014

Näsänen, R., Ojanpää, H., & Kojo, I. (2001). Effect of stimulus contrast on performance and eye movements in visual search. *Vision Research*, *41*(14), 1817–1824. https://doi.org/10.1016/S0042-6989(01)00056-6

Nelson, L. D., Simmons, J., & Simonsohn, U. (2018). Psychology's renaissance. *Annual Review of Psychology*, *69*(1), 511–534. https://doi.org/10.1146/annurev-psych-122216-011836

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., Buck, S., Chambers, C. D., Chin, G., Christensen, G., Contestabile, M., Dafoe, A., Eich, E., Freese, J., Glennerster, R., Goroff, D., Green, D. P., Hesse, B., Humphreys, M., … Yarkoni, T. (2015). Scientific standards. Promoting an open research culture. *Science*, *348*(6242), 1422–1425. https://doi.org/10.1126/science.aab2374

Nosek, B. A., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Dreber, A., Fidler, F., Hilgard, J., Kline Struhl, M., Nuijten, M. B., Rohrer, J. M., Romero, F., Scheel, A. M., Scherer, L. D., Schönbrodt, F. D., & Vazire, S. (2022). Replicability, robustness, and reproducibility in psychological science. *Annual Review of Psychology*, *73*(1), 719–748. https://doi.org/10.1146/annurev-psych-020821-114157

Olsson-Collentine, A., Wicherts, J. M., & van Assen, M. A. L. M. (2020). Heterogeneity in direct replications in psychology and its association with effect size. *Psychological Bulletin*, *146*(10), 922–940. https://doi.org/10.1037/bul0000294

Open Science Collaboration. (2015). Psychology. Estimating the reproducibility of psychological science. *Science*, *349*(6251), Article aac4716. https://doi.org/10.1126/science.aac4716

Pasternak, T., & Greenlee, M. W. (2005). Working memory in primate sensory systems. *Nature Reviews Neuroscience*, *6*(2), 97–107. https://doi.org/10.1038/nrn1603

Pham, M. T., & Oh, T. T. (2021). Preregistration is neither sufficient nor necessary for good science. *Journal of Consumer Psychology*, *31*(1), 163–176. https://doi.org/10.1002/jcpy.1209

Pool, E., Brosch, T., Delplanque, S., & Sander, D. (2016). Attentional bias for positive emotional stimuli: A meta-analytic investigation. *Psychological Bulletin*, *142*(1), 79–106. https://doi.org/10.1037/bul0000026

Proulx, M. J., & Egeth, H. E. (2008). Biased competition and visual search: The role of luminance and size contrast. *Psychological Research*, *72*(1), 106–113. https://doi.org/10.1007/s00426-006-0077-z

Reynolds, J. H., & Heeger, D. J. (2009). The normalization model of attention. *Neuron*, *61*(2), 168–185. https://doi.org/10.1016/j.neuron.2009.01.002

Scholl, B. J., & Tremoulet, P. D. (2000). Perceptual causality and animacy. *Trends in Cognitive Sciences*, *4*(8), 299–309. https://doi.org/10.1016/S1364-6613(00)01506-0

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359–1366. https://doi.org/10.1177/0956797611417632

Tadin, D., Kim, J., Doop, M. L., Gibson, C., Lappin, J. S., Blake, R., & Park, S. (2006). Weakened center-surround interactions in visual motion processing in schizophrenia. *The Journal of Neuroscience*, *26*(44), 11403–11412. https://doi.org/10.1523/JNEUROSCI.2592-06.2006

Tadin, D., Lappin, J. S., Gilroy, L. A., & Blake, R. (2003). Perceptual consequences of centre-surround antagonism in visual motion processing. *Nature*, *424*(6946), 312–315. https://doi.org/10.1038/nature01800

Vazire, S., Schiavone, S. R., & Bottesini, J. G. (2022). Credibility beyond replicability: Improving the four validities in psychological science. *Current Directions in Psychological Science*, *31*(2), 162–168. https://doi.org/10.1177/09637214211067779

Vuilleumier, P., Armony, J. L., Driver, J., & Dolan, R. J. (2003). Distinct spatial frequency sensitivities for processing faces and emotional expressions. *Nature Neuroscience*, *6*(6), 624–631. https://doi.org/10.1038/nn1057

Wells, G. L., & Windschitl, P. D. (1999). Stimulus sampling and social psychological experimentation. *Personality and Social Psychology Bulletin*, *25*(9), 1115–1125. https://doi.org/10.1177/01461672992512005

Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., & Tanaka, J. W. (2010). Controlling low-level image properties: The SHINE toolbox. *Behavior Research Methods*, *42*(3), 671–684. https://doi.org/10.3758/BRM.42.3.671

Wójcik, M. J., Nowicka, M. M., Bola, M., & Nowicka, A. (2019). Unconscious detection of one's own image. *Psychological Science*, *30*(4), 471–480. https://doi.org/10.1177/0956797618822971

Yarkoni, T. (2022). The generalizability crisis. *Behavioral and Brain Sciences*, *45*, Article e1. https://doi.org/10.1017/S0140525X20001685

Yeshurun, Y., & Carrasco, M. (2008). The effects of transient attention on spatial resolution and the size of the attentional cue. *Perception & Psychophysics*, *70*(1), 104–113. https://doi.org/10.3758/PP.70.1.104

Zhang, F., Lin, Z., Zhang, Y., & Zhang, M. (2021). Behavioral evidence for attention selection as entrained synchronization without awareness. *Journal of Experimental Psychology: General*, *150*(9), 1710–1721. https://doi.org/10.1037/xge0000825

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*(7192), 233–235. https://doi.org/10.1038/nature06860

Zhao, Y., & Heinke, D. (2014). What causes IOR? Attention or perception? —Manipulating cue and target luminance in either blocked or mixed condition. *Vision Research*, *105*, 37–46. https://doi.org/10.1016/j.visres.2014.08.020

Zwaan, R. A., Etz, A., Lucas, R. E., & Donnellan, M. B. (2017). Making replication mainstream. *Behavioral and Brain Sciences*, *41*, Article e120. https://doi.org/10.1017/S0140525X17001972

Appendix

A Primer on Technical Terms for Visual Displays and Stimuli

Screen Resolution

Screen or display resolution refers to the number of *pixels*—the smallest unit of area—displayed on the screen (Figure 2a). It is usually specified as pixels in width by height, such as 1,080 × 1,920. A screen may support multiple resolutions, but the actual resolution depends on the specific resolution set by the user.

A related concept is *pixel density*, which describes the size of pixels and is measured in pixels per inch (ppi) or pixels per centimeter (ppcm). It can be derived from the screen resolution (in pixels) and viewable diagonal length of the screen (in inches; e.g., 21 inches for a 21-inch liquid crystal display monitor)—namely, $\text{ppi} = \sqrt{\text{width}^2 + \text{height}^2}/\text{length}$.

Generally, these screen parameters—as well as the refresh rate described below—can be obtained from the computer. For example, on Windows personal computers, right-click on the desktop, select "Display settings," and then "Advanced display settings." On macOS, click on the Mac icon on the top left, select "About This Mac," and open the "Displays" tap to find the screen diagonal length and resolution (e.g., 13.3-inch [1,440 × 900]). However, for devices like the iPad, where the screen resolution and refresh rate are generally fixed and the parameters hidden, you may need to search online using the model number (found under Settings > General).

Refresh Rate

The refresh rate of a display describes its temporal resolution: how frequently the display updates the on-screen image (Figure 2b). Expressed in hertz (Hz), a 60-Hz screen refreshes images 60 times per second, or every 16.67 ms. This refresh duration defines the unit of presentation time—an image can only be displayed for a duration multiple of the refresh duration.

A related concept is *frame rate*, which describes the rate at which consecutive static images—called frames—appear on a display. It is measured in frames per second (fps; Figure 2b); for example, movies are typically filmed at 24 fps. Frame rate depends on computer hardware (central processing unit, graphics processing unit, and cables); software (drivers); and the media itself.

Visual Angle

Visual angle describes the angular size of a visual stimulus subtended at the eye—the plane angle formed by two end points of the stimulus as they converge on the center of the eye's pupil (Figure 2c). For example, the width of the index fingernail at arm's length subtends about one degree. To derive the visual angle of a stimulus, measure its physical size and its distance from the observer: angle = 2 × Arctan (half size/distance), in radians. Radians, mathematically defined as the arc length divided by the radius, can be converted to degrees—one radian is about $360/2\pi$ ($\approx$57.3) degrees. A MATLAB script is available in the Supplemental Materials that converts visual angle and number of pixels from each other.

The three-dimensional analog of plane angle is called *solid angle*—the angular area of an object subtended at a particular point in space (the apex, such as the eye or the light source; see Figure 2e). For example, the moon and the sun have about the same solid angle from earth, as demonstrated during a total solar eclipse. Solid angle can be expressed in either squared degrees or steradians (defined as the subtended area on a sphere divided by the radius squared); the two units are convertible—one steradian is equivalent to $(180/\pi)^2$ squared degrees.

Screen γ

Screen γ is an important display property that influences the appearance of images on the screen. Images are made of pixels, and the color and light intensity of a pixel are specified by its red, green, and blue values (RGB, each ranging from 0 to 255 in an 8-bit system). The RGB level is internally raised to a power value known as screen γ (typically 2.2 in liquid crystal display monitors). This determines the voltage and therefore the light intensity of that pixel. Due to its nonlinear effect, γ—along with ambient light and other display characteristics—significantly influences image brightness, contrast, and color appearance (Figure 2d). As a result, an image that looks "good" on one display might appear too bright, too dark, or off-color on another display.

Setting the γ value to one is convenient for manipulating stimulus luminance and contrast (see below) because it establishes a linear relation between the RGB value and its luminance. Linearization is

(*Appendix continues*)

achieved by applying a γ correction function—a multiplication coefficient that is the reciprocal of the device γ value (1/γ)—to the display's RGB values through a lookup table. γ measurement and display linearization can be accomplished using visual methods (less accurate but requiring no extra equipment) or photometer methods (more accurate; see Lin et al., 2023, for a tutorial).

Luminance

Luminance describes the amount of light that reaches the eye from a specific direction in space, adjusted by the eye's light sensitivity. It takes into account the physical light spectrum, which consists of electromagnetic waves with varying frequencies, multiplied by the photopic (daylight) spectral response of a typical human eye. Photometry is the measurement of light that accounts for the human eye's response and requires a photometer (for a tutorial, see Lin et al., 2023).

Luminance closely relates to our perception of brightness. It is connected to two concepts describing the light source: the total amount of visible light (*luminous flux*) and its intensity (*luminous intensity*). As Figure 2e illustrates, *luminous flux* is defined as the total amount of visible light emitted in all directions by the light source per second, measured in lumen (lm). *Luminous intensity*, on the other hand, is the luminous flux emitted by the light source in a particular direction per unit solid angle, measured in candela (cd, which is Latin for candle and is equivalent to lumen per steradian). For example, a common wax candle emits light with a luminous intensity of approximately one candela. Luminance, then, is the luminous intensity of the light source per unit of its projected area toward the viewer, measured in candelas per square meter (cd/m^2). The projected area is the area of the source (e.g., a screen that emits light or a surface that reflects light) that is projected onto the plane that is perpendicular to the viewer's line of sight.

It is important to distinguish luminance from *illuminance*, which refers to the amount of light falling onto a surface being illuminated—more specifically, the total amount of luminous flux per unit area. Illuminance is measured in lux (lx), which is equivalent to lumen per square meter. Both luminance and illuminance, but not luminous flux or luminous intensity, are (negatively) affected by distance according to the inverse-square law. This means that they decrease in proportion to the square of the measuring distance—specifically following the surface area of a sphere ($4\pi D^2$, with D being the distance).

Color

Color is our perception that distinguishes different kinds of light with varying spectra. The specific spectral power distribution over the visible range of wavelengths gives rise to the *chromaticity* of color, including *hue* (red, green, blue, etc.) and *saturation* (the amount of white light mixed with a hue). However, a given color can have infinite instances of light with different spectra (a phenomenon known as *metamerism*).

Colors can be specified using three numbers (called *coordinates*) within a *color space*. One common scheme is the *additive* scheme, where colors are described and created by adding specific proportions of three *primary* lights. To achieve a wide range of colors, the primaries are chosen as red (R), green (G), and blue (B). When the three coordinates have equal values, it represents the chromaticity of light with a flat spectrum, known as the *white point*. Together, these three primaries and the white point define an additive RGB system.

This additive scheme mimics color physiology in the retina of the human eye, where responses from three types of cone cells (S for short wavelength, M for medium wavelength, and L for long wavelength) are integrated to perceive color. Color matching experiments have been conducted to determine the proportions of the three primaries required to match colors of different wavelengths, mathematically related to the response curves of the three cones. The most widely used scheme is the CIE 1931 RGB color space (CIE stands for the initials of the French name for the International Commission on Illumination, Commission Internationale de l'Éclairage). In an additive color space, any color can be transformed into the coordinates of another additive color space using a 3 × 3 transformation matrix. For example, the widely used CIE 1931 XYZ color space is derived from the RGB color space.

Another common color scheme is the *luminance–chrominance* color space, where one coordinate represents luminance and the other two represent chromaticity. The CIE xyY color space, derived from the CIE XYZ color space, is a popular example: Y for luminance, x and y for chromaticity ($x = X/[X + Y + Z]$, $y = Y/[X + Y + Z]$). Plotting the chromaticity coordinates of x and y on the CIE x-y chromaticity diagram (Figure 2f) helps visualize the color *gamut*, which is the range of colors that can be mixed.

To display an image consistently across different displays, two things are important: understanding the representation of the image in its associated color space and knowing the specific color space of the display (*device color profile*). Color spaces are typically stored using International Color Consortium color profiles, which define the transformation to and from the CIE XYZ color space.

Contrast

Contrast refers to the difference between a stimulus and its background in terms of luminance or color. It is closely related to visibility, as black on black or white on white is difficult to discern, whereas black on white or white on black stands out. In the case of luminance (and also color chromaticity), contrast can be understood as the ratio of luminance difference and average luminance. The specific calculation of contrast depends on the situation.

For example, as Figure 2g illustrates, when describing small features on a large uniform background, Weber contrast is often used, which is the luminance difference divided by the background luminance. For periodic patterns with bright and dark features, Michelson contrast is employed, which is the difference between the maximum and minimum luminance divided by their sum. For complex images such as scenes, root-mean-square (RMS) contrast is calculated, which is the standard deviation of pixel intensities in an image.

In the Supplemental Materials, there is a MATLAB script available to calculate the mean luminance and RMS contrast of natural images and equalizes the mean luminance and RMS contrast between images.

Received July 21, 2022
Revision received June 2, 2023
Accepted June 13, 2023 ■