Methods and more

A brief look at the changes underway in *Nature Methods* as the journal continues to mature.

Last year marked the five year anniversary of *Nature Methods*. We are pleased with the community response to the journal: metrics and comments from our readers both show increasing impact of our content. Partially because of this success, we made only minor changes to the journal's offerings over this period, but the changes are now becoming more frequent as we strive to better serve our readers.

In 2008, we discontinued the Protocol article type after *Nature Protocols* launched as a dedicated venue for step-by-step descriptions of previously published methods. That same year, we began offering two new venues for presenting original research that we believe will prove valuable for many scientific communities: Analyses and Resources. Analysis articles contain thorough and objective performance comparisons of existing methods and tools. Resource papers describe new large collections of tools or data.

Original research, reviews and commentaries represent the majority of content in *Nature Methods*, and we are indebted to the scientists who serve as our authors and reviewers. Accompanying this extensive content, the Research Highlights in *Nature Methods* allow our editors to communicate high-interest methodological developments published in other journals. By describing the work in a short, engaging format we hope our readers will discover interesting methods they might otherwise overlook and possibly insights that might not be gleaned from the original manuscript. This year we began including a greater number of these highlights in each issue.

As of January of this year we also chose to focus on *Nature Methods* authors in a new piece called 'The Author File'. This monthly column provides a behind-the-scenes view of the work that went into a research paper appearing in that issue of the journal and of the primary author or authors responsible for it. It also includes their outlook on how methods could be used in other laboratories.

These sections are part of an effort to provide our readers with more general interest journalistic content relevant to practicing researchers. In this issue, we are happy to announce a new monthly column on visual presentation of scientific data. This past April we published a well-received Focus on visualizing biological data that covered the use of visualization techniques and software. The new column, 'Points of View', deals with fundamental aspects of visual presentation applicable to anyone who works with visual representations of data. Content in this column will be more along the lines of the information in the classic text *The Visual Display of Quantitative Information* by Edward Tufte (2nd edn.; Graphics Press, 2001) than the specialized information in the Focus.

The column author, Bang Wong, is the creative director of the Broad Institute in Cambridge, Massachusetts, USA where he works at the interface of science and art. He led the effort to design and build the Broad Institute's small museum, the DNAtrium, and established a data visualization initiative to develop methods for visually exploring large data sets. In addition to these efforts, he provides hands-on assistance to researchers on the communication of their experimental results to the scientific community through graphics and data visualization.

At *Nature Methods* we witness authors' struggles to display data clearly and with minimal bias. Everything from choice of colors, fonts, graph type and design to overall layout can influence data interpretation. Poor choices can have a range of effects. Figures or presentations may be unattractive but otherwise serve their purpose. Alternatively, figures may be difficult to interpret—or more worryingly—misleading. Each month Wong will focus on a particular aspect of data presentation or visualization and provide easy-to-apply tips on how to overcome common pitfalls. This month's column is on the use of color in scientific figures, a commonly mishandled aspect of figure design.

Admittedly, some journals' style rules may preclude researchers from following all advice. We expect that we may learn a thing or two ourselves. But many of the tips will be easily implementable, and published papers are not the only place where researchers present their data. PowerPoint presentations and posters often suffer the most from poor presentation. Interestingly, Tufte contends that distributing a brief written report—readable in five to ten minutes—followed by open discussion, is a superior method of transferring knowledge. But PowerPoint and poster presentations will probably be with us for the foreseeable future, and tips to make the information presented in them easier to grasp in less time would clearly benefit the scientific community.

We hope our readers find these additions to *Nature Methods* useful, and we urge you to let us know what you think, including suggestions of additions you would like to see. Your feedback will help us in our goal of providing valuable and engaging content.

POINTS OF VIEW

Color coding

Color can add dimensionality and richness to scientific communications. In figures, color is typically used to differentiate information into classes. The challenge is picking colors that are discriminable. A systematic approach to choosing colors can help us find a lineup effective for color coding.

Occasionally, authors use a sequence of colors, such as the 'rainbow' color scheme, to represent a range of values. Color, however, is not ideal for encoding quantitative data because of the inherent ambiguity in how the different colors should be ordered. For instance, does yellow represent a smaller value than blue? One could pattern the sequence after the ordering of visible light by wavelength (remembered by the mnemonic ROYGBIV), but use of this color spectrum is inherently problematic. The transitions from red to yellow to green and so on are uneven, breaking the correspondence between color and numerical value. Visually, certain colors in the rainbow spectrum seem to run on, whereas others are short lived. Even when we limit the spectrum to just a few colors, the incremental change in mapped value still might not translate to the magnitude of change we see.

In contrast, color is well suited to represent categorical data when it is used properly—for example, to distinguish between experimental conditions. If used improperly, such as by assigning intense or weak colors to specific categories, color can bias the reader. Because color is such a potent differentiator, the appropriate strategy is to choose colors that are discernible from one another but comparable in visibility.

Color is a relative medium, and neighboring colors can affect visual perception. For example, it is possible to make the same color look different or different colors appear the same (or nearly the same) by changing only the background color (**Fig. 1a**,**b**). The perception of color depends on context, and manipulating the attributes of neighboring colors affects how we see the original color[1]. A heat map requires us to judge the relative brightness of colors in a matrix. The interaction of color can cause a profound effect that makes this graphical representation suffer (**Fig. 1c**).

Every color is described by three properties: hue, saturation and lightness. Hue is the attribute we use to classify a color as red or yellow. Saturation describes the neutrality of a color; a red object with little or no white is said to be very saturated. The lightness of a color tells us about its relative ordering on the dark-to-light scale.

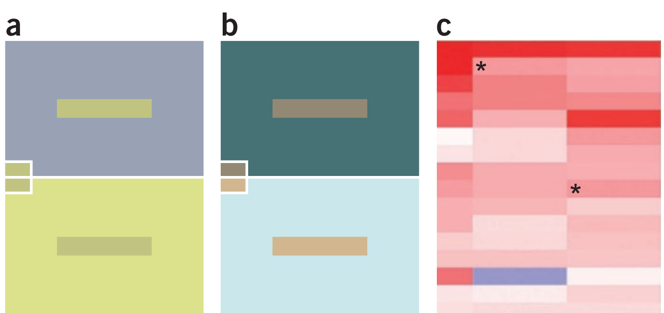


Figure 1 | Perception of color can vary. (**a**,**b**) The same color can look different (**a**), and different colors can appear to be nearly the same by changing the background color (**b**)[1]. (**c**) The rectangles in the heat map indicated by the asterisks (*) are the same color but appear to be different.

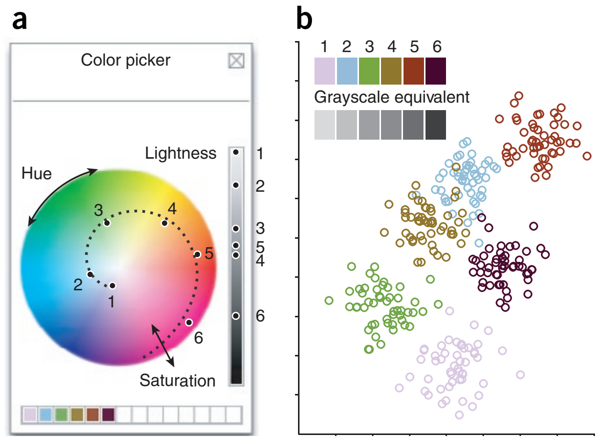


Figure 2 | Color has hue, saturation and brightness. (**a**,**b**) Colors can be tuned using a color picker (**a**). Spiraling through hue and saturation while varying lightness can generate a discernible color set distinguishable even in grayscale (points labeled 1–6).

On a computer, we can tune color attributes using the color picker (**Fig. 2a**). On a Mac or PC and in software such as Adobe Illustrator and Photoshop, the color picker is based on the traditional color wheel. In this system, hues are arranged around a circle with saturation increasing from the center outward. The 'true' color (hue) is near the ring midway from the center. On a PC and in Adobe products, the color wheel is transformed into a square with hue arrayed across the top and saturation decreasing from top to bottom. In all cases, lightness is controlled by a separate slider.

To pick colors easily discernible from each other, whether in color or converted to grayscale, spiral through the color wheel while varying the lightness (**Fig. 2**). We can achieve wide dynamic range by adjusting all three attributes of color. Our perceptual system is highly sensitive to grayscale, and the lightness property makes it possible to differentiate colors when photocopied to black and white. In this way, we can define a group of 6–8 colors. Beyond this number, the task of picking distinctive colors becomes difficult. To show more categories, we can rely on textural differences in addition to color. For example, we can encode data for two categories as red crosses and red circles.

Just picking suitable colors is not always sufficient, though. The size of the 'visual objects' in the figure also matters; the smaller the objects (or the thinner the lines) the greater the variations in hue, saturation and lightness that are needed. Finally, to test for comparable visibility of the selected colors, squint at the graphic and look for general evenness.

Color is a familiar and widely used design element. Poor color choices can introduce bias and unwanted artifacts into our presentations. Careful consideration when choosing colors will help us make the most of the communication and enable readers to discern the encoded information. Next month, we will focus on the design of data graphs.

Bang Wong

1. Albers, J. *Interaction of Color* (Yale University Press, New Haven, Connecticut, USA, 1975).

Bang Wong is the creative director of the Broad Institute of the Massachusetts Institute of Technology and Harvard and an adjunct assistant professor in the Department of Art as Applied to Medicine at The Johns Hopkins University School of Medicine.

POINTS OF VIEW

Design of data figures

Data figures or graphs are essential to life-science communication. Using these tools authors encode information that readers later decode. It is imperative that graphs are interpreted correctly. Despite the importance and widespread use of graphs, we primarily rely on our intuition, common sense and precedent in published material when creating them—a largely unscientific approach.

Because accurately interpreting visual variables is such a vital step in understanding graphs, a rational framework for creating effective graphs would accommodate the needs of the reader and focus on the strengths of human perception. Conversely, we want to avoid displays of data that are misleading or difficult to discern. For example, it can be tough to accurately judge the differences between two curves (**Fig. 1a**). The disparity is actually constant but our perceptual system is attuned to detecting minimal distances so the divergence appears to decrease. Another shortcoming limits our ability to accurately judge relative area. This diminishes the usefulness of bubble charts. For example, the larger circle in **Figure 1a** is 14 times larger than the smaller circle.

In 1967, the French cartographer Jacques Bertin provided a wide theoretical framework for information visualization[1]. His analysis focused on the visual properties of graphical elements such as shape, orientation, color, texture, volume and size for displaying quantitative variation. He defined several visual operations needed to extract information stored in graphs. Cleveland and McGill were one of the first to measure people's ability to efficiently and accurately carry out these elementary perceptual tasks[2] (**Table 1**).

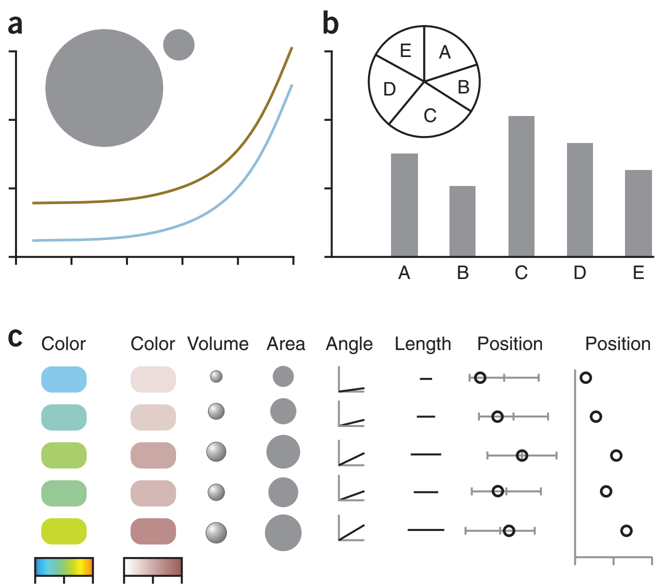


Figure 1 | Some visual estimations are more easily carried out than others. (**a**) Examples illustrating the difficulty in interpreting graphs and charts accurately. (**b**) Same data presented in a bar chart and in a pie chart. (**c**) Different visual variables encoding the same five values.

Table 1 | Elementary perceptual tasks

Rank	Aspect to compare
1	Positions on a common scale
2	Positions on the same but nonaligned scales
3	Lengths
4	Angles, slopes
5	Area
6	Volume, color saturation
7	Color hue

Tasks are ordered from most to least accurate. Information adapted from ref. 2.

When communicating with graphs, we want readers to perceive patterns and trends. This is distinct from conveying information through tables in which we report precise names and numbers. Cleveland and McGill's study assessed people's ability to judge the relative magnitude between two values encoded with a particular visual variable (for example, length, angle and others). In other words, they asked people to estimate how many times bigger *A* is when compared to *B*. Accuracy in their study does not imply reading out precise values from data points in graphs.

Different graph types depend on different visual assessments to uncover underlying trends. Pie charts are a common way to show parts of a whole. Most readers will likely judge angle when extracting information from pie charts, but they could also compare areas and arc length of the slices (**Fig. 1b**). Each of these perceptual tasks ranks low in efficiency and accuracy (**Table 1**). Plotting the same data as a bar chart effectively shows relative values (**Fig. 1b**).

When we occasionally need to invent new ways to graph data, we ideally want to use perceptual tasks that rank high in efficiency and accuracy (**Table 1**). In **Figure 1c**, I plotted the same five values using different encoding. In some cases, identifying magnitude and direction of change is laborious. In other cases, the trends are readily apparent. Encodings on the right more efficiently and accurately display the magnitude and direction of change. Though we can detect slight shifts in color hue, the relationship between hue and quantitative value is not obvious (see also ref. 3), making color hue one of the weaker methods to illustrate relative values.

Communicating with graphs depends on authors encoding information for readers to decode. Graphs' effectiveness can benefit from attention to their visual design. Composing figures with strong visual cues and relying on accurate perceptual tasks supports the visual assessment critical for interpreting information from graphs. Next month we will explore salience, the use of visual properties as differentiators.

Bang Wong

1. Bertin, J. *Semiology of Graphics*, English translation by W.J. Berg (University of Wisconsin Press, Madison, Wisconsin, USA, 1983).
2. Cleveland, W.S. & McGill, R. *Science* **229**, 828–833 (1985).
3. Wong, B. *Nat. Methods* **7**, 573 (2010).

Bang Wong is the creative director of the Broad Institute of the Massachusetts Institute of Technology and Harvard and an adjunct assistant professor in the Department of Art as Applied to Medicine at The Johns Hopkins University School of Medicine.

POINTS OF VIEW

Gestalt principles (Part 1)

Gestalt principles of perception are theories proposed by German psychologists in the 1920s to explain how people organize visual information[1]. Gestalt is a German word meaning shape or form. The principles describe the various ways we tend to visually assemble individual objects into groups or 'unified wholes'. They are highly relevant to the design of charts and graphs as well as the reports that contain them.

Gestalt is the interplay between the parts and the whole. Kurt Koffka, one of the founding fathers of Gestalt psychology, made a statement about this. He said, "The whole is 'other' than the sum of its parts." This phrase has been translated to the familiar saying, 'the whole is greater than the sum of its parts'. A classic example of subjective contour is illustrated in **Figure 1a**. We clearly see edges of a white triangle that does not exist. Koffka insisted that the emergent entity is 'other' (not greater or lesser) than the sum of the parts. By composing elements on the page according to specific principles, we can add additional layers of meaning.

In the following discussion, to be continued in next month's column, we will explore several Gestalt principles. Here we will examine the principles of similarity, proximity, connection and enclosure. The fundamental concept behind these principles is grouping; we tend to perceive objects that look alike, are placed close together, connected by lines or enclosed in a common space as belonging together. These are simple but powerful ways to build context for information.

The principle of similarity is likely familiar to many. We often use color, size and shape to organize data objects into categories. As readers, we tend to see things that are similar to be more related than things that are dissimilar (**Fig. 1b**). We can apply this observation to all elements on the page; by repeating graphical treatments including font, type size, orientation and white space, we can design elements so they appear more related.

Another quality that inclines us to make associations between objects is proximity. We tend to group objects placed close together. We can apply this principle when organizing figure panels. In a grid of evenly spaced panels, it can be unclear at first glance how one should dissect the information contained within (**Fig. 1c**). Are we to compare the panels or read them in succession? If the reader is to make two pairwise comparisons, then grouping the four panels as two pairs reinforces our natural tendency to relate proximal objects (**Fig. 1c**). If, however, we want readers to review the panels one after another, then arranging the panels in a row provides a natural order that supports reading them sequentially (**Fig. 1c**).

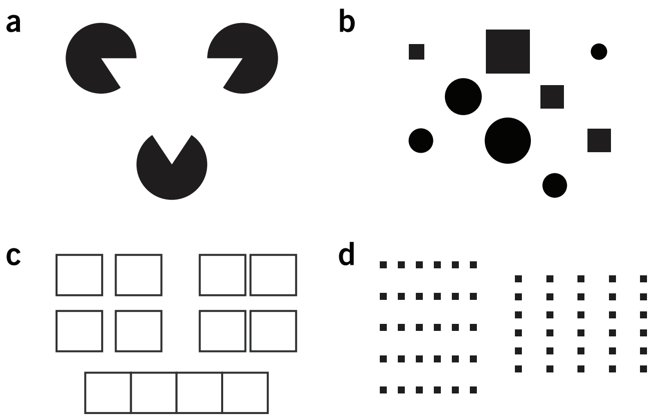


Figure 1 | Gestalt principles. (**a**) An illustration of subjective contour. (**b**) Similar objects are visually grouped. (**c**) Objects placed close to one another are seen as going together. (**d**) Relative proximity elicits vertical or horizontal correlations between objects.

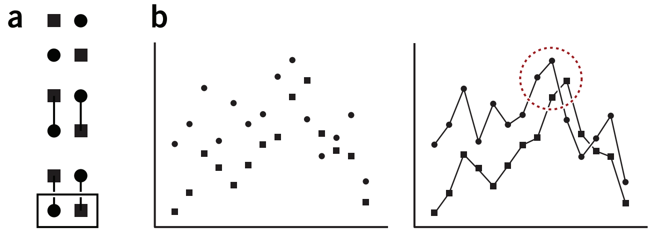


Figure 2 | Principles of grouping. (**a**) Relative strength of grouping by similarity, proximity, connection and enclosure. (**b**) Lines in graphs create clear connection. Enclosure is an effective way to draw attention to a group of objects.

Proximity could be considered a special case of grouping by similarity because of the underlying spacing between objects. Relative spacing between columns and rows can dramatically affect whether we group the components vertically or horizontally (**Fig. 1d**).

Whereas objects grouped by similarity and proximity are seen as loose confederations, grouping by connection and enclosure leads us to associate them as a unified whole. The relative strength each principle exerts on perceptual grouping is illustrated in **Figure 2a**. Lines create clear connection and bring out the overall shape of the data (**Fig. 2b**). They provide a useful method for encoding information in graphs and network diagrams. Finally, grouping by enclosure resulting in elements bounded in a common region is powerful enough to overcome similarity, proximity and connection (**Fig. 2**).

The Gestaltists described phenomena about how we organize bits and pieces of visual information into larger units. This perceptual organization is deeply ingrained in the visual experience. When we present visual information, including blocks of text projected on screen, it is helpful to arrange the elements into a meaningful structure. One framework is simply to group related information. The principles of similarity, proximity, connection and enclosure provide simple rules to draw correlations between visual elements.

Next month, we will examine the principles of visual completion and continuity, which describe our tendency to fill in missing information to perceive shapes as being complete to the greatest degree possible.

Bang Wong

1. Palmer, S.E. *Vision Science: Photons to Phenomenology* (Massachusetts Institute of Technology Press, Cambridge, Massachusetts, USA, 1999).

Bang Wong is the creative director of the Broad Institute of MIT & Harvard and an adjunct assistant professor in the Department of Art as Applied to Medicine at The Johns Hopkins University School of Medicine.

POINTS OF VIEW

Gestalt principles (Part 2)

Our visual system attempts to structure what we see into patterns to make sense of information. The Gestalt principles describe different ways we organize visual data. Last month, we looked at four principles that incline us to group objects when they are made to look alike, are placed near one another, are connected by lines or are enclosed in a common space[1]. This month, we will examine the principles of visual completion and continuity. These principles are useful in page layout work and when we compose figures and slides.

Visual interpolation creates interesting illusions in which we see contours that do not actually exist. The Kanizsa triangle[2] we looked at last month is a famous example of illusory or subjective contours (**Fig. 1a**). The 'Pac-Man' shapes align to form what appears to be well-defined edges of a triangle.

Another example of visual completion is shown in **Figure 1b**. We automatically and spontaneously perceive a full circle behind the square. In reality, several shapes are possible in the occluded area. This disparity between the actual visual stimulus and what we think (or know) we should be seeing points to the psychology involved in seeing. It is likely that we complete the object behind the square as a circle because it produces a simple and familiar shape.

Because we have a strong tendency to see shapes as continuous to the greatest degree possible, we fill in voids with visual cues found elsewhere on the page. This means every element on a page affects how we perceive every other element. Visual completion enables us to forgo the extraneous lines, boxes, bullets and other graphical elements that tend to clutter our presentations.

Graphics and text can be considered shapes with vertices and edges. To construct unified compositions, align these constituent parts to form meaningful blocks of information (**Fig. 2a**). Simple geometric shapes provide a base structure on which to organize and build content (**Fig. 2b**). It is helpful to actually draw these background shapes and use them as alignment guides. I have shown examples of guides as dotted lines in **Figure 2**, which would not exist in the final figure. Placing components on the guide's path anchors the information and helps the audience identify patterns. Curvilinear guides are useful in sequencing information because they create a clear path through the material. Such alignment produces invisible lines that connect content.

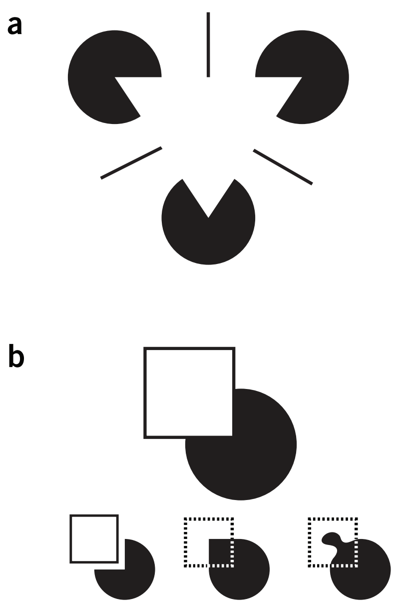


Figure 1 | Visual completion. (**a**) The Kanizsa triangle and illusory contour. (**b**) Spontaneous and automatic completion of occluded surfaces as a simple and familiar circle.

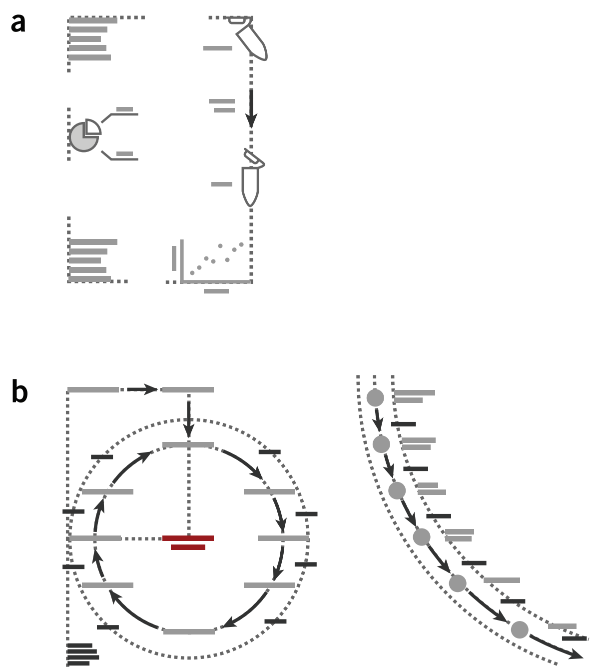


Figure 2 | Alignment. (**a**) Graphics and text used as vertices and edges of geometric shapes. (**b**) Geometric and curvilinear shapes used as flexible guides to align content.

Our eyes are acutely aware of small misalignments; compositions that use guides tend to look clean and professional. We can create different alignment guides for different information. For example, labels that describe an action can be distinguished from those for names. Moreover, we can combine alignment with the Gestalt principles of similarity, proximity, connection and enclosure to group information and structure the content. The action labels can be distinguished from the name labels with color or typographical treatment.

Our goal is to lay out information in a way that enhances its message. In structuring the components of a slide or figure, we inevitably affect the surrounding white space. White space is a vital part of design; it frames the content and gives our eyes a place to rest. Next month, we will look at 'negative space' to complete our exploration of composition.

Bang Wong

1. Wong, B. *Nat. Methods* **7**, 863 (2010).
2. Kanizsa, G. *Organization in Vision: Essays on Gestalt Perception* (Praeger Publishers, New York, 1979).

Bang Wong is the creative director of the Broad Institute of MIT & Harvard and an adjunct assistant professor in the Department of Art as Applied to Medicine at The Johns Hopkins University School of Medicine.

POINTS OF VIEW

Negative space

Negative space, also known as whitespace, refers to the unmarked areas of the page. Collectively, it is the margins and the gaps between text blocks and images. Whitespace is as much a part of a composition as the titles, words and pictures. The Swiss typographer Jan Tschichold calls whitespace 'the lungs of a good design'[1]. In addition to giving elements breathing room, judicious use of whitespace can dramatically improve the visual appeal and effectiveness of figures, posters and slides.

The term whitespace stems from the printing practice in which white paper is generally used. Margins and gaps that separate blocks of text make it easier to access written material because they provide a visual structure. Well-planned negative space balances the positive (nonwhite) space and is key to aesthetic. Asian art makes wide use of negative space to create harmony and to add dimension to flat silkscreen prints.

The openings in and between objects can inform us about the objects themselves. A protein and the negative space masked in black are shown in **Figure 1**. Note how the reverse image implies and defines the shape of the protein (**Fig 1b**). It gives us almost as much information as the original image.

In science communication, unfettered empty space is rare. Presentations tend to be densely packed. Whitespace is a commodity we need to put to good use. Some people see whitespace as expendable and even as an indication that there is insufficient content to fill the page. After all, whitespace carries no information, so what is the harm in filling it up? The harm is that overcrowded slides and posters are taxing to comprehend. Usually this is due to the irregularity of the negative space.

A focus on the spacing of elements can help us create layouts with meaningful structure. One approach I find useful is to enclose images and text in boxes either literally or by visual estimation. Doing so makes the distribution of positive and negative spaces clear. A typical scientific poster not dissimilar to those we see at conferences is shown in **Figure 2a**. A study of spaces reveals a labyrinth of shapes (**Fig. 2b**). The goal is to unify the whitespaces into regularly shaped contiguous blocks. This can be achieved by aligning the boxes vertically or horizontally to create visual divides that inform the grouping of information. For example, we might use larger gaps to differentiate sections but thinner gutters to separate items within a section (**Fig. 2c**). In this way, the negative space can telegraph to readers the hierarchy and organization of content.

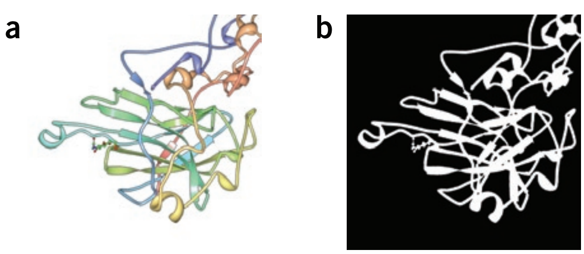


Figure 1 | Empty space defines the shape of an object. (**a**,**b**) Ribbon diagram of a protein (**a**) and with the negative space masked in black (**b**).

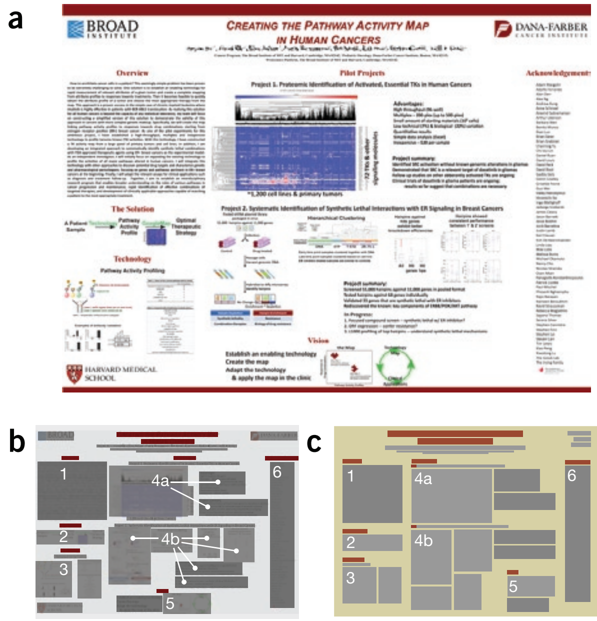


Figure 2 | Whitespace can be used to structure content. (**a**) An example of a scientific poster. (**b**) A space study reveals that contents in sections 1–6 are scattered and whitespace is fragmented. (**c**) An example of consolidated whitespace organizing contents.

The approach described above requires us to manipulate many elements. It can be a challenge to size and tile the parts to fit a prescribed layout. Luckily modern software makes layout work fluid. We are constrained to scale images proportionally. However, we can radically alter the shapes and sizes of text blocks to make them conform to the available space. Text allows us to adjust the spacing between letters, the length of the lines and the spacing between those lines.

Additionally, whitespace offers one of the most effective ways to attract readers' attention. In congested environments, applying brighter colors or special typographical styles such as capitalization or boldface may not be enough to get certain content noticed. In these situations, try surrounding the content to be emphasized with relatively more of the available whitespace. The generous framing will usually draw the eyes to that part of the page.

In the last six columns, I have discussed ways to visually encode data (color coding, design of data figures and salience) and methods for organizing elements on the page (Gestalt principles and negative space). Next month, I will review these ideas and apply the concepts to real-world examples.

Bang Wong

1. Ambrose, G. & Harris, P. *The Layout Book* (AVA Publishing, Lausanne, Switzerland, 2007).

Bang Wong is the creative director of the Broad Institute of the Massachusetts Institute of Technology and Harvard and an adjunct assistant professor in the Department of Art as Applied to Medicine at The Johns Hopkins University School of Medicine.

POINTS OF VIEW

Points of review (part 1)

My goal over the next two months is to show concretely how scientific figures can benefit from design principles. I will review concepts from past columns by applying them to several published figures.

In the design of common objects, such as a door, when a handle is used many people will mistakenly pull even if the door is to be opened by pushing. When the handle is replaced with a flat plate, which affords pushing, people will know to push. When dealing with figures, we depend on visual cues. We want our figure's layout to express its underlying meaning.

The diagram shown in **Figure 1a** is intended to illustrate three parts of a microscopy system[1]. We could redraw the figure so that the threefold nature of the system is apparent even at a glance. The Gestalt principles (November 2010 column)[2] impart trends in visual organization; we tend to organize objects into groups, for instance, when they are placed near one another, connected by lines or contained in a common space. Using the principles of proximity, connection and enclosure we could sketch the general form of the microscopy system as shown in **Figure 1b**. By grouping the components related to each part of the system and placing those groupings in compartments, we create a visual structure that strongly reflects the message. The prominent horizontal feature links the system together.

In arranging elements on the page, we inextricably affect the negative space (January 2011 column)[3]. Similar to the Gestalt principles, white space is another mechanism to organize content. For example, wider gaps can be used to separate major groupings whereas narrower spaces are left between more related objects. In **Figure 1a** there are large unused areas on the top right and on the left. Consolidating the empty spaces into more regularly shaped areas creates uniformity and helps to further delineate our defined groupings (**Fig. 1b**).

Meaningful compositions become more challenging to create when figures have many independent parts. A helpful strategy is to let the intent of the figure guide the layout. In **Figure 2a** a protocol for analyzing gene expression is illustrated[4]. The details of the process are presented in several steps. But the even distribution of graphical elements provides neither an intuitive path through the information nor visual cues for us to relate the parts to one another. One fitting structure is horizontal groupings strung together vertically (**Fig. 2b**). We can rely on the principle of visual completion (December 2010 column)[5] and line up the arrows between steps to connect and order the process. To differentiate the central path that traces the gene of interest from additional reagents, I used orientation and alignment to create salience (October 2010 column)[6] and set them apart. The added reagents are either misaligned or placed at an angle from the central molecules.

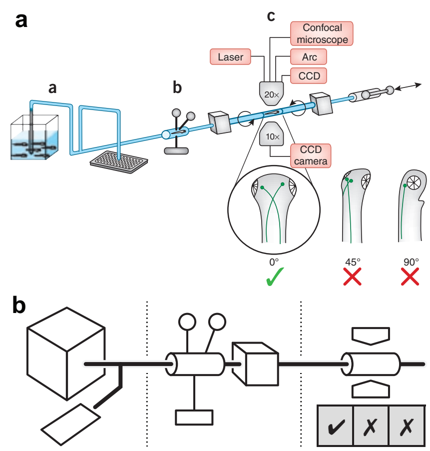


Figure 1 | Layouts can express meaning. (**a**) Diagram of a microscopy system. Reprinted from *Nature Methods*[1]. (**b**) A sketch using grouping and white space to make the three parts of the system being illustrated more apparent.

When showing sequential information, it is also helpful to use consistent language and representations so readers can more easily follow the story. In **Figure 2a**, the identifying barcode represented by the color green at the beginning is not the one captured at the end. These inconsistencies may require readers to redouble their steps when working through the figures.

Conceptual figures like the ones described above have an important purpose; they provide context for readers to understand the experimental design and research results.

Bang Wong

1. Tamplin, O. & Zon, L. *Nat. Methods* **7**, 600 (2010).
2. Wong, B. *Nat. Methods* **7**, 863 (2010).
3. Wong, B. *Nat. Methods* **8**, 5 (2011).
4. Peck, D. *et al. Genome Biol.* **7**, R61 (2006).
5. Wong, B. *Nat. Methods* **7**, 941 (2010).
6. Wong, B. *Nat. Methods* **7**, 773 (2010).

Bang Wong is the creative director of the Broad Institute of the Massachusetts Institute of Technology and Harvard and an adjunct assistant professor in the Department of Art as Applied to Medicine at The Johns Hopkins University School of Medicine.

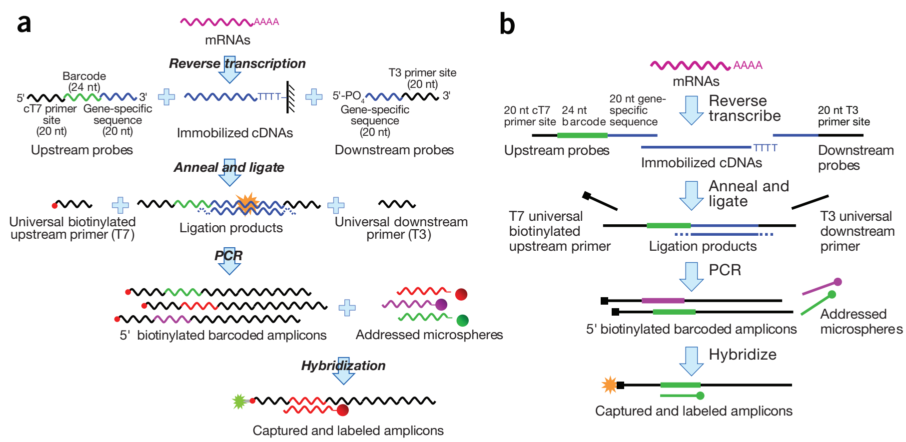


Figure 2 | Visual structure that matches the message. (**a**) Illustration showing a gene expression analysis technique. Reprinted from *Genome Biology*[4]. (**b**) The same elements organized according to the purpose of the illustration, which is to show a sequence of steps.

POINTS OF VIEW

Points of review (part 2)

I will continue to demonstrate how judicious choice of graphical representations can improve visual communication. Here I will focus on data figures.

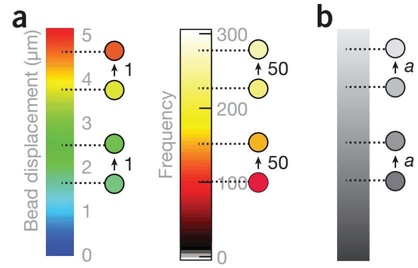
The power and primary purpose of graphs is to reveal connections in data. As opposed to tables, in which there is little visual association between individual values, graphs and charts depend on readers to form patterns. In reading graphs, we observe individual data points, keep each of them in memory and construct an image from the constituents. The entire process can be exceedingly fast and attest to the power of visual perception. Graphical encoding needs to support the detection and assembly process of reading graphs.

We are more accurate at certain types of visual estimation than others (September 2010 column)[1]. For example, to understand relative differences between categories, a standard bar chart might be easier to read than a pie chart, particularly to appreciate the direction and magnitude of change (**Fig. 1**). Small differences are more readily apparent when we compare length of bars (**Fig. 1c**) than sizes of pie slices (**Fig. 1a**)[2].

Pie charts can be useful. Although they are not intended to show complex relationships, pie charts do well to depict parts of a whole. *The Wall Street Journal Guide to Information Graphics*[3] suggests an ordering of slices to aid reading: place the largest wedge to the right of 12 o'clock, the second largest to the left of 12 o'clock and the remainder counter-clockwise descending in size (**Fig. 1d**). In this way, the largest (and presumably most important) wedges end up at the top. With the two largest slices sharing a vertical edge, we can rely on reading angles to estimate proportion.

When we need to show several dimensions of data at once, the multivariate scatter plot is one solution. With these displays of data, the challenge is in choosing representations that allow us to distinguish the qualities within and between parameters. In an example published figure that relies on position, color, color value and size to represent different aspects of the data (**Fig. 1b**)[2], it is difficult to pick out the eight sizes of data points, 11 shades of yellow and 13 shades of blue. One way to reduce the busyness is to limit the color value and size scales to several ranges (for example, 0–3, 4–7 and others). Additionally, only plotting the parameters that matter most to convey the intended message will also reduce visual complexity. In the graph in **Figure 1c**, color value actually has a very limited role; it is not explicitly keyed in the original figure legend. But because of the severe data occlusion problem, it might be most helpful to separately plot the former yellow and blue categories each in gray (**Fig. 1e**).

Figure 2 | Color is not ideal for presenting quantitative data. (**a**) Shifts in color scales (circles) are not visually commensurate with change in value. Reprinted from *Nature Methods*[2,5]. (**b**) A gradation from 10–90% black produces even transitions.



Color is not ideal for representing quantitative information. In the above example, yellow is particularly problematic. It has an extremely restricted value range so there is not much difference between the lightest and deepest yellow. With color scales such as the rainbow spectrum, uneven transitions in color can break the correspondence between color and numerical value (August 2010 column)[4]. In **Figure 2a**, two color scales from recent journal articles are shown[1,3]. In each instance, I sampled colors equal distance apart at two locations. The same incremental change in value does not equate to the qualitative difference between the pairs of color spots (**Fig. 2a**). Color can introduce considerable biases in data presentation. When we must represent values with color, a gradient of 10–90% black produces a consistent visual scale (**Fig. 2b**).

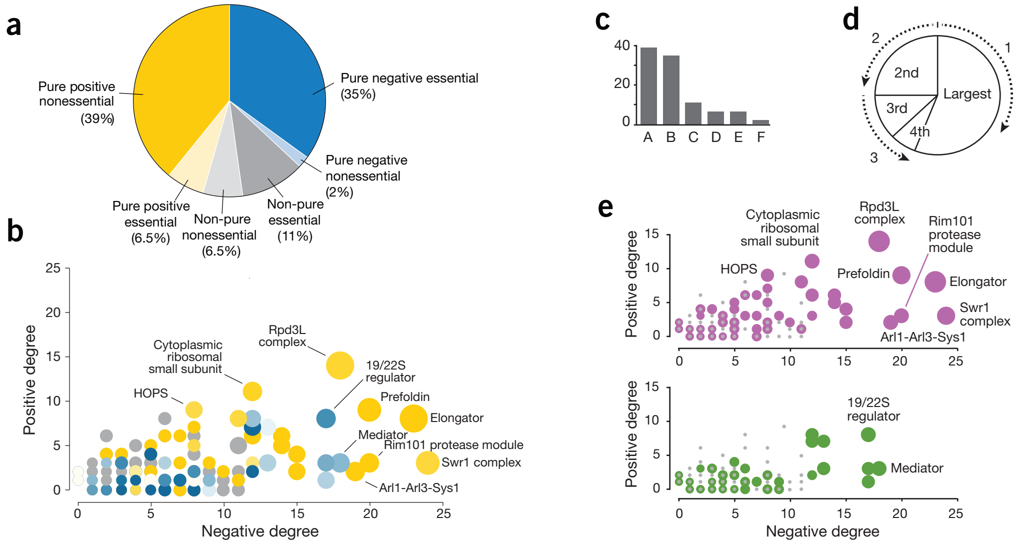
Next month I will cover another fundamental of design: typography.

Bang Wong

1. Wong, B. *Nat. Methods* **7**, 665 (2010).
2. Baryshnikova, A. *Nat. Methods* **7**, 1017–1024 (2010).
3. Wong, D. *The Wall Street Journal Guide to Information Graphics* (W.W. Norton and Company, New York, New York, USA, 2010).
4. Wong, B. *Nat. Methods* **7**, 573 (2010).
5. Legant, W. *Nat. Methods* **7**, 969–971 (2010).

Bang Wong is the creative director of the Broad Institute of the Massachusetts Institute of Technology and Harvard and an adjunct assistant professor in the Department of Art as Applied to Medicine at The Johns Hopkins University School of Medicine.

Figure 1 | Certain visual encodings are easier to read. (**a**,**b**) Analysis of genetic interactions. Adapted and reprinted from *Nature Methods*[2]. (**c**) A bar chart showing data from the pie chart in **a**. (**d**) A method for ordering slices of a pie chart. (**e**) Multiple views to show overlapping data from **b**. Former 'yellow' and 'blue' categories are shown in purple and green, respectively.



POINTS OF VIEW

Typography

Typography is the art and technique of arranging type. Like a person's speaking style and skill, the quality of our treatment of letters on a page can influence how people respond to our message. It is an essential act of encoding and interpretation, linking what we say to what people see.

Typography has been known to affect perception of credibility. In one study, identical job resumes printed using different typefaces were sent out for review. Resumes with typefaces deemed appropriate for a given industry resulted in applicants being considered more knowledgeable, mature, experienced, professional, believable and trustworthy than when less appropriate typefaces were used[1]. In this case, picking the right typeface can help someone's chances of landing a job.

The term typeface is frequently conflated with font; Arial is a 'typeface' that may include roman, bold and italic 'fonts'. Most generally we categorize letterforms as serif or sans serif. Primary characteristics of a letterform are illustrated in **Figure 1a**. Serif typefaces tend to be thinner, more formal and easier to read in multiline blocks of text because the 'feet' help our eyes follow the line. Sans serif typefaces have simpler letterforms, are informal and, according to some, less readable in long stretches, so are appropriate for short bursts of text such as headings and labels. In general, sans serif fonts work well for slides and serif fonts for posters and printed documents.

Picking type is a matter of personal taste, but typography exists to honor content. The four most common typefaces are Baskerville, Helvetica, Palatino and Times New Roman (**Fig. 1b**), and a good rule is: when limited to the palette of type preinstalled on our computers, pick one and ignore the rest. The acclaimed poet and typographer Robert Bringhurst eloquently states that these four typefaces are "faces with nothing to offer one another except for public disagreement"[2]. If nothing else, the single typeface approach ensures consistency. Uniformity is one form of beauty; contrast is another. Of course, typefaces can be combined, but the operation requires care and craft.

Typography can reveal the tone of the document and clarify the structure and meaning of the text. Perhaps more than any other formatting options, our selection of fonts shows readers at a glance whether the document is stately or humble, formal or informal, creative or technical. Words, phrases, sentences and blocks of text should be spaced according to their underlying meaning. The space between paragraphs should be greater than between lines; items of a list should be spaced so they appear related to each other but separate from adjacent text. As I previously described in my columns on Gestalt principles[3,4], objects that are aligned or placed near one another are seen as belonging together. In **Figure 2**, I show sample text with spacing established simply with carriage returns (**Fig. 2a**), in contrast to the spacing made by adjusting line and paragraph settings (**Fig. 2b**). The relative scale of white space in **Figure 2b** makes the hierarchy of the content apparent. Differentially aligning the paragraph text and bulleted list, when allowed, differentiates the content.

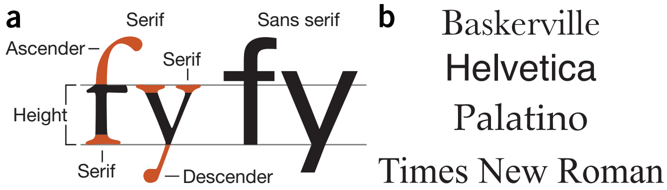


Figure 1 | Typefaces. (**a**) The anatomy of letterform for serif (Garamond) and sans serif (Univers) type both set at 58 point. (**b**) Four of the most readily available fonts.

To achieve meaningfully spaced text, use the 'space before' and 'space after' settings instead of extra carriage returns. Find the settings under Font menu > Paragraphs (PowerPoint) or Format menu > Paragraphs (Word). The paragraph text in **Figure 2b** is set with 5 point space after it; the bulleted list has 3 point space after it. Furthermore, left justified text leaves a ragged right edge that can be made more regular by adjusting the size of the text box and using soft returns (shift and return) to manually break lines.

Most documents can be set perfectly well with one typeface using no more than two or three type sizes, with judicious use of bold and italics if necessary. By limiting the variation in type and type treatment, we can unify the tapestry of visual information to be presented on scientific slides or posters. In these formats, we often need to combine a disparate array of information taken from different sources, including text, images and figures. A consistent typographical program unifies the elements and makes documents easier to read. Typography must draw our attention before it is read but not interfere with reading. The goal is to achieve a balance between text and all other elements on the page.

Bang Wong

1. Shaikh, D. & Fox, D. *Usability News* **10** (2008).
2. Bringhurst, R. *The Elements of Typographic Style* (Hartley & Marks Publishers, Point Roberts, Washington, USA, 2005).
3. Wong, B. *Nat. Methods* **7**, 863 (2010).
4. Wong, B. *Nat. Methods* **7**, 941 (2010).

Bang Wong is the creative director of the Broad Institute of the Massachusetts Institute of Technology and Harvard and an adjunct assistant professor in the Department of Art as Applied to Medicine at The Johns Hopkins University School of Medicine.

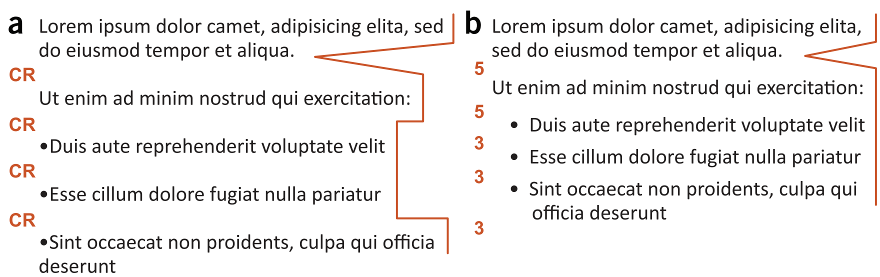


Figure 2 | Spacing can reveal structure and give meaning to text. (**a**) Uniform carriage return (CR) spacing is incongruous with hierarchical content. (**b**) Relative spacing using paragraph formatting expresses relationships in the text. Numbers are 'space after' values given in point sizes.

POINTS OF VIEW

The overview figure

Our goal when writing research papers is to convey information as clearly as possible. In past columns I have suggested several graphic design techniques to improve the clarity of figures. In addition to refining data figures, including overview figures in a research paper provides a framework for readers to understand the experimental design and reported findings.

Illustrative schematics in overview figures can make publications accessible to a wider audience. They give context to the data presented. An example of such a figure is one I illustrated (**Fig. 1**)[1]. It depicts technology called Hi-C used to determine how cells organize the billions of DNA base pairs. This opening figure is effective because it constructs a mental model for understanding the technology and primes readers to expect DNA sequence information as the primary data type.

Typical overview figures illustrate a procedure (**Fig. 1**) or compare conditions such as 'control' and 'experimental' (**Fig. 2**)[2]. These figures portray a continuous process as discrete steps. As such, it is imperative that we create continuity through imagery and written descriptions. Each step in the progression is understood by relating it to the previous and subsequent step. For comparisons, differences in the corresponding steps between processes should also be highlighted (**Fig. 2**).

In the design of procedural schematics, it is useful to adopt an '*A* to *B*' structure in which *A* and *B* are states connected by an action. The states are often depicted graphically, and the action is text describing the transformation from *A* to *B* (for example, cut with restriction enzyme). To create good visual linkage between steps, redraw the elements from the previous step highlighting only the effective change. Because readers need to follow a series of events, it is helpful to account for all graphical elements introduced and removed from the figure. When the numbers of elements do not match from one step to the next, it can confuse readers and compromise the utility of overview figures.

With visual communications, it is essential that symbols have minimal overlapping meanings. For example, arrows can be used to point and to indicate motion. When the same graphical representation is used to mean different things, it impedes efficient and accurate decoding of information. In designing **Figure 2**, I used arrows to indicate progression and leaders—lines without arrowheads—for labeling. In **Figure 1**, I used arrows to represent and indicate the directionality of sequencing primers. Clear delineation in meaning enables readers to quickly learn the visual vocabulary and group information into hierarchy. Similarly, using language consistently makes it easier for readers to follow the word story. One sentence structure could be used to describe actions and another to label objects (that is, 'cut with restriction enzyme' and 'restriction fragments').

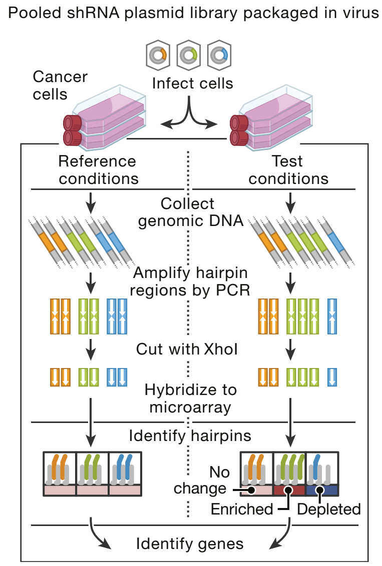


Figure 2 | Well-ordered compositions and clear visual encodings make schematics easy to follow. Schematic comparing experimental conditions in a pooled RNA interference screen. Reprinted from reference 2.

Fundamentally, overview figures are intended to convey general concepts and not to present data. When selecting graphics to represent each step, consider how a reader might interpret the imagery. In **Figure 2,** the authors initially selected a heatmap taken from elsewhere in the manuscript to illustrate the 'identify hairpins' step. Although the researchers did identify hairpins by analyzing heatmaps, a schematic representation (as shown) better demonstrates the experimental strategy. Research data in the context of an overview figure are disconcerting. Are we supposed to read them as graphs or see them as stand-ins for something else?

Despite their general usefulness, overview figures are usually the first to be eliminated when space becomes limited. One strategy to have them included in the final publication is to design the illustrations with an economy of marks and to make them as compact as possible. I designed the overview of Hi-C (**Fig. 1**) without intervening arrows and used the action labels as headers to save space. The horizontal layout provides a natural left-to-right ordering. Space-efficient designs can be achieved by fully using available whitespace[3] and organizing visual elements into groups according to the Gestalt principles[4,5].

Bang Wong

1. Lieberman-Aiden, E. *et al. Science* **326**, 289–293 (2009).
2. Luo, B. *et al. Proc. Natl. Acad. Sci. USA* **105**, 20380–20385 (2008).
3. Wong, B. *Nat. Methods* **8**, 5 (2011).
4. Wong, B. *Nat. Methods* **7**, 863 (2010).
5. Wong, B. *Nat. Methods* **7**, 941 (2010).

Bang Wong is the creative director of the Broad Institute of the Massachusetts Institute of Technology and Harvard and an adjunct assistant professor in the Department of Art as Applied to Medicine at The Johns Hopkins University School of Medicine.

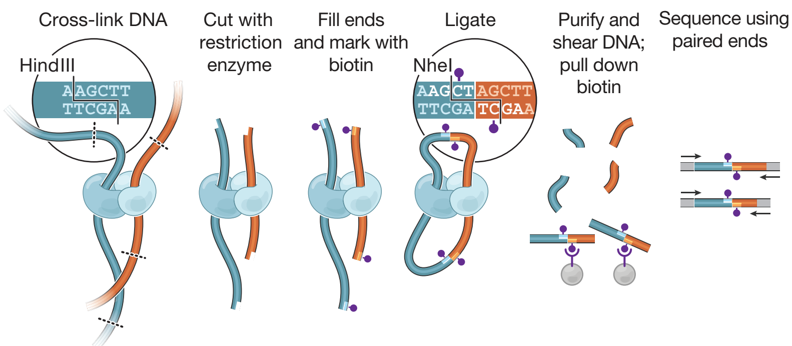


Figure 1 | Overview figures can clarify concepts. Outline of the Hi-C technique used to decipher the three-dimensional structure of the human genome. Reprinted from reference 1.

POINTS OF VIEW

Color blindness

Since my first column on color coding[1] appeared, we have received a number of e-mails asking us to highlight the issue of color blindness. One of those correspondences was published in the October 2010 issue[2]. Here I offer guidelines to make graphics accessible to those with color vision deficiencies.

Color blindness affects a substantial portion of the human population. Protanopia and deuteranopia, the two most common forms of inherited color blindness, are red-green color vision defects caused by the absence of red or green retinal photoreceptors, respectively. In individuals of Northern European ancestry, as many as 8 percent of men and 0.5 percent of women experience the common form of red-green color blindness[3]. If a submitted manuscript happens to go to three male reviewers of Northern European descent, the chance that at least one will be color blind is 22 percent.

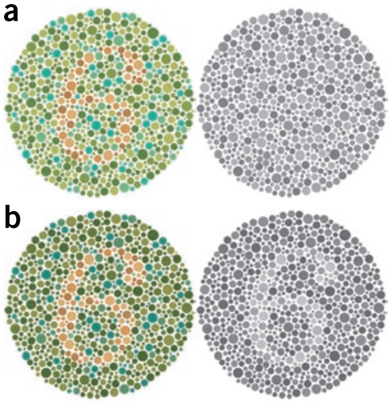


Figure 1 | Ishihara color-vision test plate. (**a**) Viewers with normal color vision should see the numeral '6'. (**b**) Changing lightness of background improves contrast.

Picking colors suitable for color-blind readers not only enhances accessibility but also is good graphic design practice. For example, the Ishihara color vision test intentionally relies only on color hue to create contrast, as evident when the image is converted to grayscale (**Fig. 1a**). In general, colors will be easier to distinguish when they vary in lightness and saturation as well as hue (**Fig. 1b**). The palette of eight colors shown in **Figure 2** has good overall variability and can be differentiated by individuals with red-green color blindness.

It is useful to remember that pure red and pure green are not the only culprits in color confusion—rather, any color with components of red and green can cause trouble. Authors can rely on software to simulate how images might appear to individuals with red-green color blindness. In Adobe Illustrator and Photoshop, first convert the document to RGB color space for accurate simulation and create a soft proof (View > Proof Setup > Color Blindness). Simultaneously viewing the original and the soft proof (Window > Arrange > New Window in Photoshop) makes it convenient to adjust colors in order to make them universally accessible. Web-based tools such as Vischeck (www.vischeck.com) can also produce simulated images.

Color	Color name	RGB (1–255)	CMYK (%)	P	D
	Black	0, 0, 0	0, 0, 0, 100		
	Orange	230, 159, 0	0, 50, 100, 0		
	Sky blue	86, 180, 233	80, 0, 0, 0		
	Bluish green	0, 158, 115	97, 0, 75, 0		
	Yellow	240, 228, 66	10, 5, 90, 0		
	Blue	0, 114, 178	100, 50, 0, 0		
	Vermillion	213, 94, 0	0, 80, 100, 0		
	Reddish purple	204, 121, 167	10, 70, 0, 0		

Figure 2 | Colors optimized for color-blind individuals. P and D indicate simulated colors as seen by individuals with protanopia and deuteranopia, respectively.

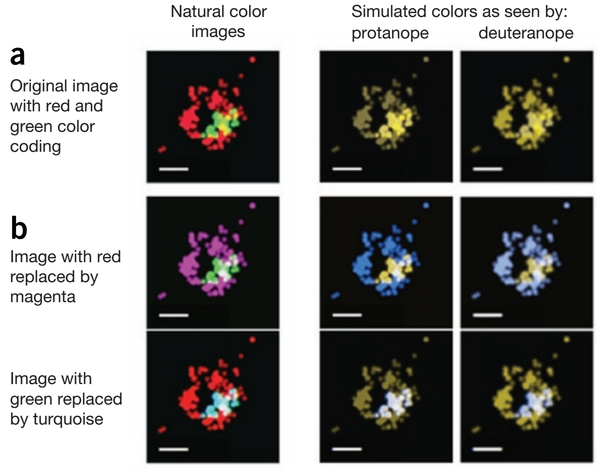


Figure 3 | Red-green color coding in an immunofluorescent image. (**a**) Conventional color coding is difficult for individuals with red-green color blindness (protanopia or deuteranopia) to discriminate. (**b**) Replacing red with magenta (top) or green with turquoise (bottom) improves visibility for such individuals. Source image from reference 4.

Perhaps the most widespread use of red-green color coding in the life sciences is in immunofluorescent images (**Fig. 3a**). To make this and other artificial color schemes accessible to readers with red-green color blindness, replace red with magenta (**Fig. 3b**, top). This can be easily accomplished using Photoshop. Because red mixes with blue to produce magenta, copy the contents from the red channel (Window > Channels) and paste them into the blue channel. This unconventional magenta-green color coding may require a key indicating that the overlap of these colors produces white. Alternatively, some individuals with red-green color blindness find that replacing green with turquoise provides the most visible difference (**Fig. 3b**, bottom).

For color-blind individuals viewing existing images with colors that are difficult to discriminate, there are several tools for computers and mobile devices that may be helpful. The DanKam app for iPhone and Android takes information coming into the phone's camera and shifts the color spectrum so that colors fall within the range that people who are color blind can see. eyePilot (www.colorhelper.com) and Visolve Deflector (www.ryobi-sol.co.jp/visolve/en/deflector.html) each use a 'lens' to enable users to manipulate colors of any content on the screen. People with typical color vision may also find these computer tools useful. For example, eyePilot permits one to isolate specific colors against a gray background, facilitating in-depth analysis of presentations with complex color-coding schemes.

Bang Wong

1. Wong, B. *Nat. Methods* **7**, 573 (2010).
2. Albrecht, M. *Nat. Methods* **7**, 775 (2010).
3. Deeb, S.S. *Clin. Genet.* **67**, 369–377 (2005).
4. Jones, S.A *et al. Nat. Methods* **8**, 499–505 (2011).

Bang Wong is the creative director of the Broad Institute of the Massachusetts Institute of Technology & Harvard and an adjunct assistant professor in the Department of Art as Applied to Medicine at The Johns Hopkins University School of Medicine.

POINTS OF VIEW

Avoiding color

Last month I wrote about color blindness and ways to make information accessible to individuals with color vision deficiencies. I would like to continue by considering graphical alternatives to color that could improve the overall clarity and utility of data displays.

The primary use of color in research is to convey information. When used effectively, color can simplify a complex analysis task. When misused, it can bias a reader's perception of the underlying data. For example, when color gradients indicating relative quantity contain abrupt transitions, specific numerical ranges can be preferentially accentuated (**Fig. 1a**). Edward Tufte advises us that color used poorly is worse than no color at all; his motto is: "Above all, do no harm"[1]. Color can cause the wrong information to stand out and make meaningful information difficult to see. Furthermore, the overuse of color can produce visual clutter akin to signage in Times Square or Piccadilly Circus with countless elements competing for our attention.

In addition to limiting accessibility, there are several other disadvantages to using color to present data. I showed how the visual phenomenon resulting from the interaction of color can cause the same color in heatmaps to appear different[2]. Color is a relative medium. When we pair hues varying greatly in saturation or value (lightness), we can unintentionally produce presentations that are lopsided. In **Figure 1b**, the light blue bands appear under-represented partially because they are lighter than the other colors as evident by looking at the key in grayscale (**Fig. 1b**). Color can also elicit size biases; some people find equal areas filled with vibrant colors seem to be more dissimilar than when less saturated colors are used.

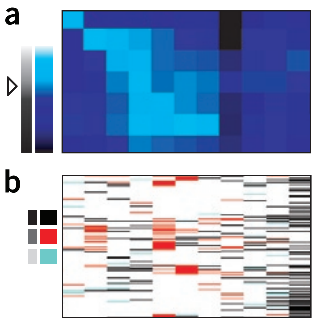


Figure 1 | Color can mask data. (**a**) Color scale with sharp transition in hue and value (arrow) can exaggerate specific data ranges. (**b**) Juxtaposing colors highly varying in saturation and value can make aspects of the data appear under-represented (light blue).

Although color is an attractive choice for conveying information, it may not be the best visual cue to bring out relevant trends. Color hue can be such a potent differentiator that using size, shape, texture, length, width, orientation, curvature and intensity to encode information may enable more aspects of the data to be discriminable. Our choice of graphical cues should depend on what we and others need to see to reliably pick out patterns.

In one project at the Broad Institute, researchers wanted to understand the evolution of molecular networks by studying gene expression in yeast. They had time course data for about a dozen species. The researchers were interested in comparing expression profiles across genes and species. With their data displayed as heatmaps, it is difficult to characterize the differences between profiles (**Fig. 2a**). Redrawing the data as line graphs and shading the area under the curve better support the visual task of comparing patterns for mirror symmetry and peak shift (**Fig. 2b**). To gauge conservation across metabolic pathways, the researchers calculate a correlation score accounting for all species for each node in the network and assign color to score (**Fig. 2c**). As it is difficult to sequence color hues, mapping the data to length and position makes it easier to see points of high and low correlation (**Fig. 2d**). The compact format allowed data for both metabolites and genes to be displayed (**Fig. 2d**). The visual complexity that comes from too many colors makes it difficult to also show the metabolite data in the original scheme (**Fig. 2c**).

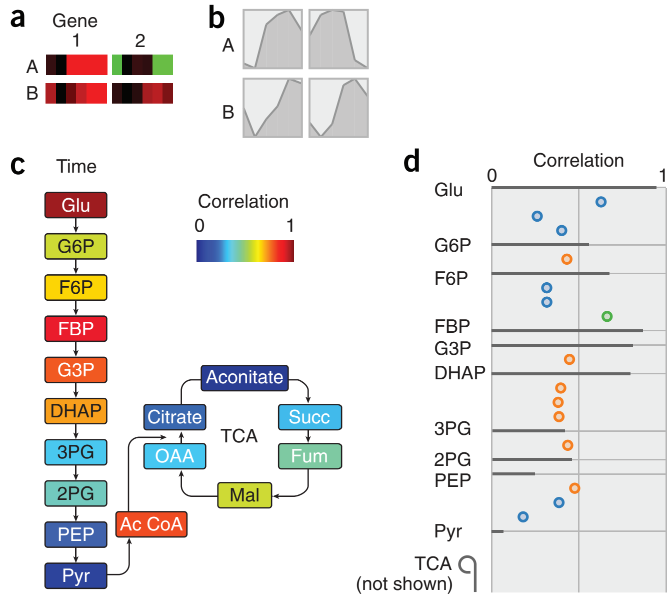


Figure 2 | Color can limit accessibility and hinder analysis. (**a**) Heatmap representation of time series data for species A and B. (**b**) Filled line charts of data from **a** facilitate profile comparison. (**c**) Color hue indicates correlation score for metabolites in glycolysis (boxes). Enzymes are shown as arrows. (**d**) Replacing color encoding from **c** with bar length for metabolites and position of circles on the *x* axis for enzymes increases data density and makes rank ordering easy. Color indicates directionality of enzymatic activity. Visualization technique is from reference 3.

Color is often our first choice when it comes to showing data. Depending on the fundamental visual task required for analysis, basic diagrammatic marks may do a better job of revealing data structures. I have seen squiggly lines used effectively to denote several data dimensions at once. Although color is inextricably tied to what many of us consider to have high visual impact, expressiveness relies primarily on one's graphical selection, whereas effectiveness also depends on the capabilities of the perceiver.

Bang Wong

1. Tufte, E. *Envisioning Information* (Graphics Press, Cheshire, Connecticut, USA, 1990).
2. Wong, B. *Nat. Methods* **7**, 665 (2010).
3. Meyer, M. *et al. Proc. EuroVis* **29**, 1043–1052 (2010).

Bang Wong is the creative director of the Broad Institute of the Massachusetts Institute of Technology & Harvard and an adjunct assistant professor in the Department of Art as Applied to Medicine at The Johns Hopkins University School of Medicine.

POINTS OF VIEW

Simplify to clarify

In the past two columns I have focused on making information accessible. I discussed ways to avoid color and shift color hues to make them discernible by individuals with color vision deficiencies. In this column I focus on ways to make information apparent by simplifying its presentation.

Simplification can lead to greater clarity. In the marketplace, simplicity is the capital used to develop clear brand identity. Apple prides itself on making things simple and on offering products that are easy to use. In science, value is placed on communications that are accurate and concise. Edward Tufte wrote about the data:ink ratio as a call to reduce the proportion of a graphic that is used for decorative purposes or that can be erased without loss of data information[1].

The best way to simplify is to reduce the number of elements on the page. Every picture and bit of text stimulates the visual senses and contributes to the intricacy of the presentation. The aim is to use the fewest possible 'marks' to convey the message without sacrificing sophistication. Our general tendency is to fill white space with more information. Thus, the judicious removal of material is typically not a natural part of the authoring workflow. But the opportunity lost from including less is gained in greater emphasis on what is shown.

I find it helpful to focus on the primary goal of a figure or slide as a guide to pare it down to its constituent parts. I assess every component against this measure to create a hierarchy of information, eliminating extraneous elements and refining the remainder to support the message. In **Figure 1**, an inversion event that results in two fusion genes is shown. The process as initially illustrated is unnecessarily complicated (**Fig. 1a**). The diagram can be simplified by combining the first two steps of the process and using fewer arrows to indicate movement (**Fig. 1b**). These modifications effectively improve the communication by simplifying the design.

Simple should not be mistaken for simplistic. By simplifying, we take advantage of the way people see and process information. The Gestalt psychologists favored the theoretical approach that

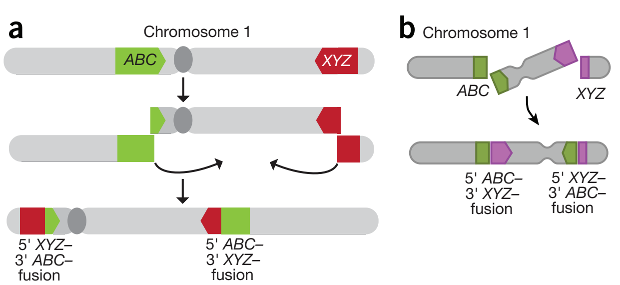


Figure 1 | Simplifying illustrations. (**a**) Initial diagram shows chromosomal inversion in three steps with the distal chromosomal ends exchanging places as indicated by arrows. (**b**) A simplified version of the diagram in **a** with fewer steps and a single arrow depicting the rotation of the center part of the chromosome.

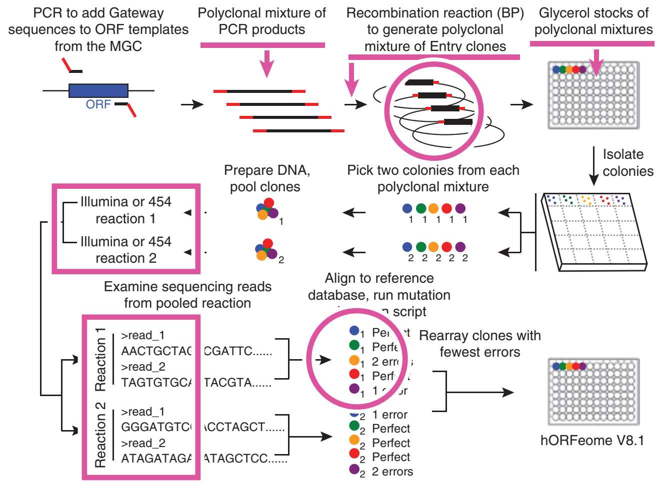


Figure 2 | Reducing redundant elements. Words repeated in several labels (magenta boxes) can be pulled out as headers. Using the smallest number of examples to convey a concept will make ideas easier to understand (magenta circles). Grouping labels that describe transformations between steps with arrows and starting or ending products with images (magenta arrows) will add meaningful structure to layouts. Reprinted from *Nature Methods*[2].

explains phenomena of perceptual organization in terms of maximizing simplicity. Simplified presentations with well-ordered layouts and clean lines are more engaging to read and are likely better understood.

Eliminating redundant elements is another way to trim extra material from a presentation. It is common to see repetition in figure labels indicating a series, for example, 'reaction 1' and 'reaction 2' (**Fig. 2**). In these cases, extracting the word in common between the labels to use as a header will generally tidy the appearance. Moreover, authors will occasionally show a variety of experimental constructs to capture the underlying diversity (**Fig. 2**). In these situations, try to use the minimum number of examples required to demonstrate the concept. Including more examples than necessary may actually confuse readers.

Simplicity can also be achieved by systematically organizing the elements that remain. By grouping we can make a system of many independent parts appear to have fewer elements. Deciding what goes with what is the first step to create structure. Labels that describe an action or transformation from one step to the next should be placed with the progression arrows; object descriptions should be placed next to the images (**Fig. 2**). Also, layouts that are neat and orderly appear simpler. In addition to grouping, align elements to a few imaginary horizontal and vertical lines appropriate to the presentation, paying attention to the negative space to create clear boundaries between groups.

Bang Wong

1. Tufte, E. *The Visual Display of Quantitative Information* (Graphic Press, Cheshire, Connecticut, USA, 2007).
2. Yang, X. *et al. Nat. Methods* **8**, 659–661 (2011).

Bang Wong is the creative director of the Broad Institute of the Massachusetts Institute of Technology & Harvard and an adjunct assistant professor in the Department of Art as Applied to Medicine at The Johns Hopkins University School of Medicine.

POINTS OF VIEW

Arrows

Arrows are one of the most commonly used graphical devices in scientific figures. In the July 2011 issue of *Nature Methods* alone I counted nearly 300 instances of arrows; more than half of the figures contain them. Given the widespread use of arrows, it is worthwhile to take a closer look at this privileged class of diagrammatic form and how we might benefit from its use.

Arrows can be highly efficient instruments of visual communication because they guide us through complex information. Typically arrows are used to point out relevant features, order sequences of events, connect elements and indicate motion. In molecular biology, there are several conventions involving the arrow that are generally recognized (**Fig. 1a**). For example, an arrow with a right-angle line segment is understood as a transcription start site or promoter, and a short arrow placed parallel to a line usually indicates a PCR primer. Several other common conventions are shown in **Figure 1a**. But authors also use arrows to illustrate other concepts, some of which are easily understood, whereas others may be less intuitive.

In his thorough survey of diagrams Robert E. Horn documented hundreds of meanings for arrows, including metaphorical uses such as increases and decreases[1]. An arrow's geometric shape can tell us something about its purpose (**Fig. 1b**), but its meaning is refined and interpreted in context. Arrows are a special class of symbols that can have multiple meanings even when used in the same figure. A recently published figure has many arrows that are used to label parts, convey mechanical motion and show reagent flow (**Fig. 1c**).

When arrows are added to diagrams, they are most readily interpreted as conveying change, movement or causality (**Fig. 2a**). In one study, researchers asked college students to evaluate mechanical diagrams with and without arrows. Participants who saw diagrams

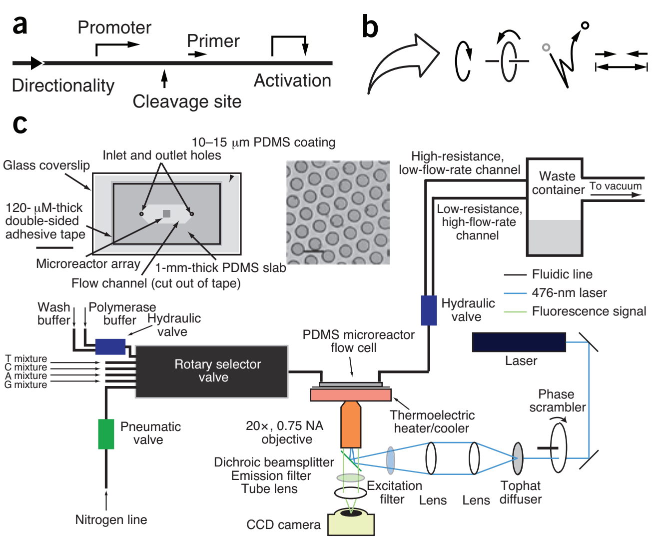


Figure 1 | Arrows in scientific diagrams. (**a**) Well-understood conventions in molecular biology indicated by arrows. (**b**) Arrows are defined loosely by their geometric shapes and more definitely in context. (**c**) A diagram with 19 arrows used as leaders, to indicate reagent flow and to show mechanical movement. Reprinted from *Nature Methods*[3].

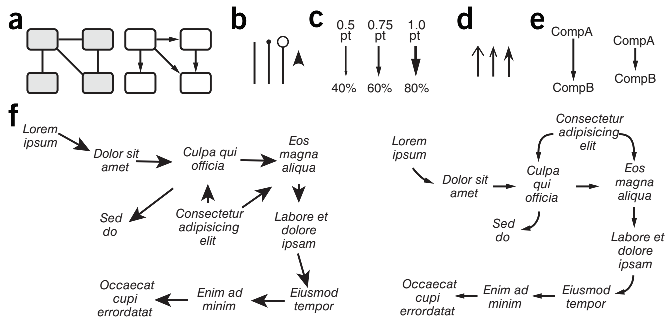


Figure 2 | Functional qualities of arrows. (**a**) The use of arrows versus lines as connectors suggests a certain functional relationship. (**b**) Alternatives to arrows as leader lines. (**c**) Reasonably sized arrows clearly indicate direction without being a distraction. (**d**) Trapped whitespace in 'open' arrowheads creates optical illusions that can attract unwanted attention. (**e**) Whitespace at the ends of the arrows makes them easy to discriminate from other content. (**f**) Orienting arrows in similar directions creates natural visual flow.

with arrows included twice as much functional information in their descriptions as those who saw diagrams without arrows[2]. Arrows are therefore most effectively used to focus attention on the functional relationships between elements rather than the elements themselves.

A goal in producing effective figures might be to use arrows sparingly and clearly. One way to do this is to reserve the use of lines with heads shaped like arrows for indicating direction or sequence and use other well-known graphical marks for other purposes. To emphasize the structure of a system—that is, spatial, as opposed to functional, inter-relatedness of the parts—we should use lines instead of arrows to connect the elements (**Fig. 2a**). For example, leaders are lines used to point at, or lead to, labeled or important parts of an illustration. Leaders used for labels should have either no head or only a bullet: either a small ball or open circle (**Fig. 2b**). One exception is the well-understood arrowhead commonly used in micrographs or other imaging to indicate salient features.

The arrow's distinctiveness comes from its asymmetric form. As such, arrows should be well-proportioned so that their directionality is easy to recognize but not be so big as to distract us from reading the content they intend to illuminate. I prefer Adobe Illustrator for drawing arrows because the software offers fine control of size and shape. For print publication, an arrow with a stem weight of 0.75 points and arrowhead scaled to 60% produces a balanced arrow (**Fig. 2c**). Also, I avoid open arrowheads (that is, the letter V on a stick) and those with sweeping wings because the trapped whitespace produces the optical illusion of 'sparkle', adding visual noise (**Fig. 2d**). Finally, arrows should be strung together as a continuous wireframe upon which to hang content. This can be achieved by avoiding sharp opposing arrow orientation and allowing for whitespace at the ends of the arrows (**Fig. 2e,f**).

Used most effectively, arrows are the 'verbs' of visual communication, describing processes and functional relationships. Next month, I will focus on layout.

Bang Wong

1. Horn, R.E. *Visual language: global communication for the 21st century* (MacroVU, Inc., Bainbridge Island, Washington, USA, 1998).
2. Hesier, J. & Tversky, B. *Cogn. Sci.* **30**, 581–592 (2006).
3. Sims, P. *et al. Nat. Methods* **7**, 575–580 (2011).

Bang Wong is the creative director of the Broad Institute of the Massachusetts Institute of Technology & Harvard and an adjunct assistant professor in the Department of Art as Applied to Medicine at The Johns Hopkins University School of Medicine.

POINTS OF VIEW

Layout

Layout is the act of arranging text and images on the page according to an overall aesthetic scheme and for the purpose of clarifying a presentation. In graphic arts, it is the elephant in the room; layout underlies everything we do when we communicate visually. Well-structured content can guide readers through complex information, but when the material we present lacks order, it can confuse or, worse yet, agitate readers trying to make sense of the material.

Many artists and architects achieve balanced outcomes by proportioning their work to approximate the golden section. The golden section is a special mathematical relationship that comes from dividing a line into two segments where the ratio of the total length $(x + y)$ to the length of the longer segment (x) is the same as that of the length of the longer segment (x) to the length of the shorter segment (y) (**Fig. 1a**), or 13:8. Many celebrated paintings since at least the Renaissance exhibit these proportions (**Fig. 1b**).

Compositional aesthetics may serve a fundamentally different purpose from designs aimed to communicate. However, the Fibonacci numbers, which are also linked to the golden ratio, heavily influence graphic design. This sequence of numbers starts with 0 and 1 and each subsequent integer is the sum of the previous two (that is, 0, 1, 1, 2, 3, 5, 8, 13 and so on). The quotient of successive pairs of numbers, with the exception of the first few, is approximately 1.6180 (or 13:8). The harmonious relationships of the Fibonacci integers are often used as measurements for font sizes and determining page layouts in books.

A practical application of the golden section is to incorporate their congruous proportions into slides and posters we create, and not just for artistic reasons: the placement of objects on a page can carry meaning. A simplified version of the golden section is the 'rule of thirds', which suggests dividing a page into nine equal parts (**Fig. 1c**). Elements placed along the lines and especially where the lines intersect (the so-called power points) become more visually prominent. Eye-tracking studies have shown that our gaze lingers in the regions marked by the lines when we scan an image.

Using a grid to aid layout (**Fig. 2a**) can dramatically streamline the design process by taking the guesswork out of sizing and placing content. Try creating a set of strategically placed guides in Microsoft

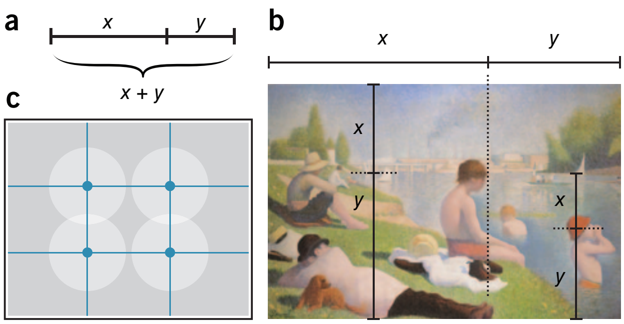


Figure 1 | Infallible proportions. (**a**) The golden section is a line segment divided by the golden ratio 13:8 such that $(x + y)$ is to x as x is to y. (**b**) In *Bathers at Asnières,* Georges-Pierre Seurat used the golden section to position the horizon and subjects in the composition (http://en.wikipedia.org/wiki/File:Seurat_bathers.png). (**c**) The 'rule of thirds' is a simplified version of the golden section used to form interesting compositions.

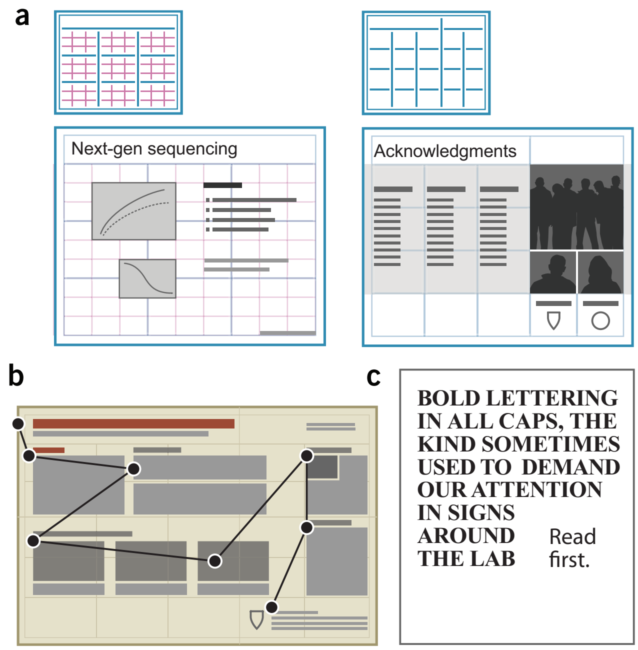


Figure 2 | Gridlines help to structure layouts. (**a**) Examples of gridline systems for presentation slides. (**b**) Arrange elements according to the order in which they should be read. (**c**) Surrounding an element in ample white space helps it get noticed first.

PowerPoint or Adobe Illustrator before you work. Grids help to anchor content and create stability within a design. They also build consistency between slides that allows the audience to anticipate where content will appear.

Layout is more than adhering to lines of a grid system: it is the process of planning out exactly the journey we want the eyes to travel across the arrangements (**Fig. 2b**). The goal is to reveal the hierarchical relationship in the information and make clear what is to be read first, second and so on. This can be done by developing dominance with some elements and practicing restraint with others. Two ways to draw a reader's attention to a compositional element is to make it visually different from its surroundings[1] or to frame the object in ample white space[2] (**Fig. 2c**). The Gestalt principles[3,4] also offer useful operational guidance to describe relationships between objects based on certain graphical cues.

We all have seen slides and posters packed full of content where the presenters have assigned equal visual weight to all the material. In these situations, it is difficult to know where to begin reading. The legendary American graphic designer Paul Rand said, "Without contrast, you're dead." Layout is the foundation of graphic design, and it should not be overlooked. How we arrange elements on the page can help or hinder whether the information is understood.

Next month, I will focus on the importance of aligning 'salience' and 'relevance'.

Bang Wong

1. Wong, B. Salience. *Nat. Methods* **7**, 773 (2010).
2. Wong, B. Negative space. *Nat. Methods* **8**, 5 (2011).
3. Wong, B. Points of View: Gestalt principles (part 1). *Nat. Methods* **7**, 863 (2010).
4. Wong, B. Points of View: Gestalt principles (part 2). *Nat. Methods* **7**, 941 (2010).

Bang Wong is the creative director of the Broad Institute of the Massachusetts Institute of Technology & Harvard and an adjunct assistant professor in the Department of Art as Applied to Medicine at The Johns Hopkins University School of Medicine.

POINTS OF VIEW

Salience to relevance

In science communication, it is critical that visual information be interpreted efficiently and correctly. The discordance between components of an image that are most noticeable and those that are most relevant or important can compromise the effectiveness of a presentation. This discrepancy can cause viewers to mistakenly pay attention to regions of the image that are not relevant. Ultimately, the misdirected attention can negatively impact comprehension.

Salience is the physical property that sets an object apart from its surroundings. It is particularly important to ensure that salience aligns with relevance in visuals used for slide presentations. In these situations, information transmission needs to be efficient because the audience member is expected to simultaneously listen and read. By highlighting relevant information on a slide, we can direct a viewer's attention to the right information. For example, coloring a row or column of a table will preferentially direct attention to the selected material (**Fig. 1a**). As information presented as tables typically appears homogenous, it is especially helpful to define what is most important. The same approach can be applied to plots and graphs to delineate segments of data (**Fig. 1b**). Whereas these techniques are not appropriate for all journal publications, annotating information presented in slides can be an effective mechanism to enable the audience to better grasp what is being said and shown.

Human vision is highly selective. When multiple stimuli are in a scene they compete for our visual attention. We make sense of the visual field by selecting, in turn, one or few objects for detailed analysis at the expense of all others. Cognitive scientists create 'salience maps' to describe the relative visibility of objects in an image that explain what we might look at first, second and so on[1].

Using the concept of a salience map, we can rely on relative visibility to order content on the page and help us design better graphics. There are several graphical variables—including color, shape, size and position—we can use to create salience (see October 2010 column)[2]. Salience is a relative property that depends on the relationship of one object to other objects on the page. Information that is presented physically larger is usually easier to see and is likely to be read first. In a composition where most of the parts are oriented vertically and horizontally, elements placed at a diagonal stand out. On a backdrop of predominantly black-and-white elements, colored information is highly conspicuous (**Fig. 1**).

a

Color name	RGB (1–255)
Black	0, 0, 0
Orange	230, 159, 0
Sky blue	86, 180, 233
Bluish green	0, 158, 115
Blue	0, 114, 178
Vermillion	213, 94, 0

b

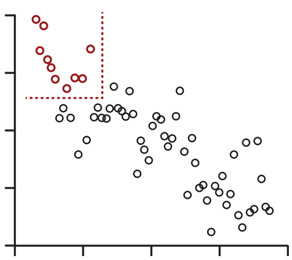


Figure 1 | Matching salience to relevance draws visual attention to important information. (**a**) Table with a row highlighted. (**b**) Segments of data in a plot emphasized with color.

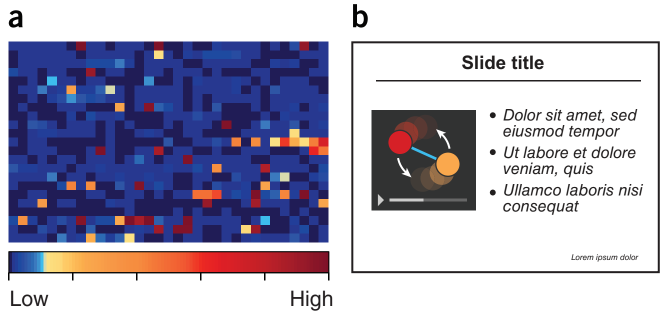


Figure 2 | Discordances between salience and relevance can be harmful. (**a**) The relative visibility of hues in the color scale is asymmetric, making higher values (represented by deep red) less apparent. (**b**) Continuously moving images can be distracting and can compromise the viewer's ability to concentrate on other content.

In contrast, unintentional and inadvertent assignment of salience can be harmful to the communicative potential of images. In the sample heatmap shown in **Figure 2**, the authors chose a color scale that makes common sense, using deep red to represent high values. But in this case lower values are actually more salient than higher ones because deep red is hard to see against the deep blue background of the lowest values.

What stands out is often taken as most important or relevant. In one study, researchers assessed the effects of salience on the ability of test subjects to accurately answer questions that required interpreting weather maps. By alternating the relative visibility of task-relevant and task-irrelevant information (in this case, information about pressure and temperature, respectively) they found that display factors such as salience had large effects on task performance[3]. For example, a question about wind direction was supposed to elicit an answer about air pressure, but when data on temperature were made most apparent, subjects incorrectly responded with a reference to temperature, having been influenced by the salience of the temperature data presented.

In presentations, a potential source of misalignment between salience and relevance is in the use of moving images. Presenters may include short movies (for example, a rotating three-dimensional structure). When these movies are allowed to loop continuously, this powerful competing stimulus makes it nearly impossible to concentrate on other content, as motion is one of the most potent mechanisms for attracting attention. For this reason, animation in PowerPoint slides should be used judiciously. The element being animated should direct our attention to the most relevant content that supports the primary message of the slide. An oscillating arrow will draw more of our attention than the objects it is intended to highlight.

It is well recognized that how the same information is presented can dramatically affect comprehension. Making relevant information visually obvious will ensure that viewers notice the right content. To get a sense of what is most salient on the screen, stand back and squint.

Next month, I will conclude this segment of 'design principles' by discussing the value of 'design' itself.

Bang Wong

1. Fecteau, J. & Munoz, D. *Trends Cogn. Sci.* **10**, 382–390 (2006).
2. Wong, B. *Nat. Methods* **7**, 773 (2010).
3. Hegarty, M. *et al. J. Exp. Psychol. Learn. Mem. Cogn.* **36**, 37–53 (2010).

Bang Wong is the creative director of the Broad Institute of the Massachusetts Institute of Technology & Harvard and an adjunct assistant professor in the Department of Art as Applied to Medicine at The Johns Hopkins University School of Medicine.