80/20 Rule

A high percentage of effects in any large system are caused by a low percentage of variables.[1]

The 80/20 rule asserts that approximately 80 percent of the effects generated by any large system are caused by 20 percent of the variables in that system. The 80/20 rule is observed in all large systems, including those in economics, management, user-interface design, quality control, and engineering, to name a few. The specific percentages are not important, as measures of actual systems indicate that the proportion varies between 10 percent and 30 percent. The universality of the 80/20 rule suggests a link to normally distributed systems, which limits its application to variables that are influenced by many small and unrelated effects—e.g., systems that are used by large numbers of people in a variety of ways. A few examples of the 80/20 rule include:[2]

80 percent of a product's usage involves *20* percent of its features.
80 percent of a town's traffic is on *20* percent of its roads.
80 percent of a company's revenue comes from *20* percent of its products.
80 percent of innovation comes from *20* percent of the people.
80 percent of progress comes from *20* percent of the effort.
80 percent of errors are caused by *20* percent of the components.

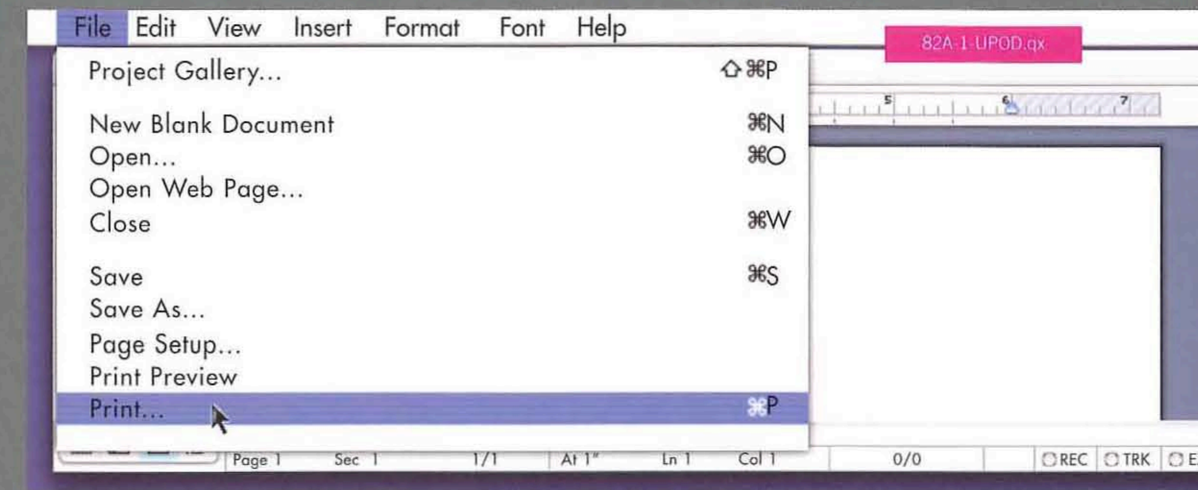
The 80/20 rule is useful for focusing resources and, in turn, realizing greater efficiencies in design. For example, if the critical 20 percent of a product's features are used 80 percent of the time, design and testing resources should focus primarily on those features. The remaining 80 percent of the features should be reevaluated to verify their value in the design. Similarly, when redesigning systems to make them more efficient, focusing on aspects of the system beyond the critical 20 percent quickly yields diminishing returns; improvements beyond the critical 20 percent will result in less substantial gains that are often offset by the introduction of errors or new problems into the system.

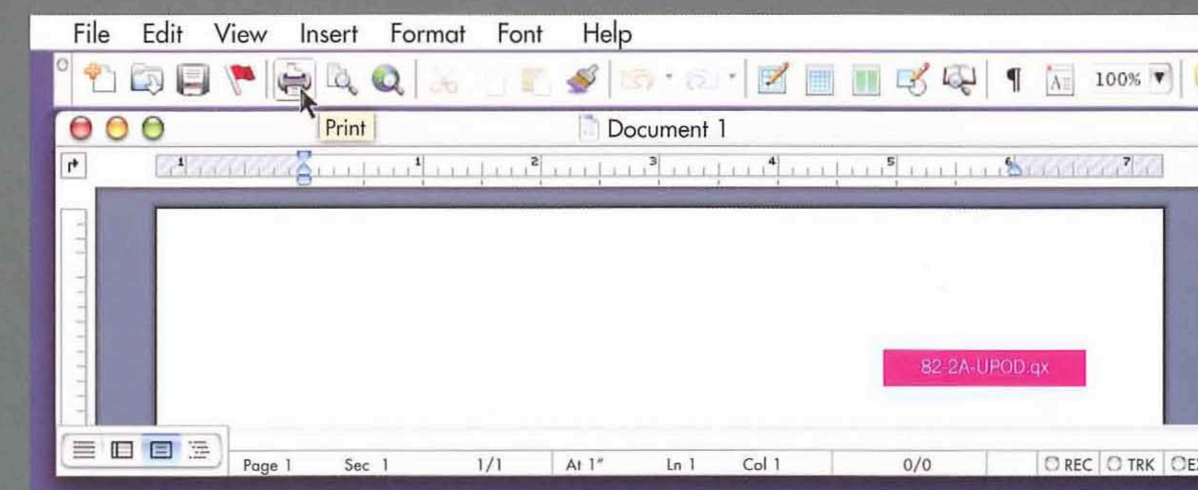
All elements in a design are not created equal. Use the 80/20 rule to assess the value of elements, target areas for redesign and optimization, and focus resources efficiently. Noncritical functions that are part of the less-important 80 percent should be minimized or even removed altogether from the design. When time and resources are limited, resist efforts to correct and optimize designs beyond the critical 20 percent, as such efforts yield diminishing returns. Generally, limit the application of the 80/20 rule to variables in a system that are influenced by many small and unrelated effects.

See also Cost-Benefit, Form Follows Function, Highlighting, and Normal Distribution.

[1] Also known as *Pareto's Principle, Juran's Principle,* and *Vital Few and Trivial Many Rule.*

[2] The first recognition of the 80/20 rule is attributed to Vilfredo Pareto, an Italian economist who observed that 20 percent of the Italian people possessed 80 percent of the wealth. The seminal work on the 80/20 rule is *Quality Control Handbook* by Joseph M. Juran (Ed.), McGraw-Hill, 1951.





Graphical user interfaces conceal most of their functions in drop-down menus (top image). This reduces the complexity of the display, but also makes frequently used functions more difficult to access. Identifying the critical 20 percent of the functions and making them readily available in toolbars solves the problem (bottom image).

Alignment

The placement of elements such that edges line up along common rows or columns, or their bodies along a common center.

Elements in a design should be aligned with one or more other elements. This creates a sense of unity and cohesion, which contributes to the design's overall aesthetic and perceived stability. Alignment can also be a powerful means of leading a person through a design. For example, the rows and columns of a grid or table can be used to make explicit the relatedness of elements sharing those rows and columns, and to lead the eyes left-to-right and top-to-bottom accordingly. Edges of the design medium (e.g., edge of a page or screen) and the natural positions on the design medium (e.g., centerlines) should also be considered alignment elements.

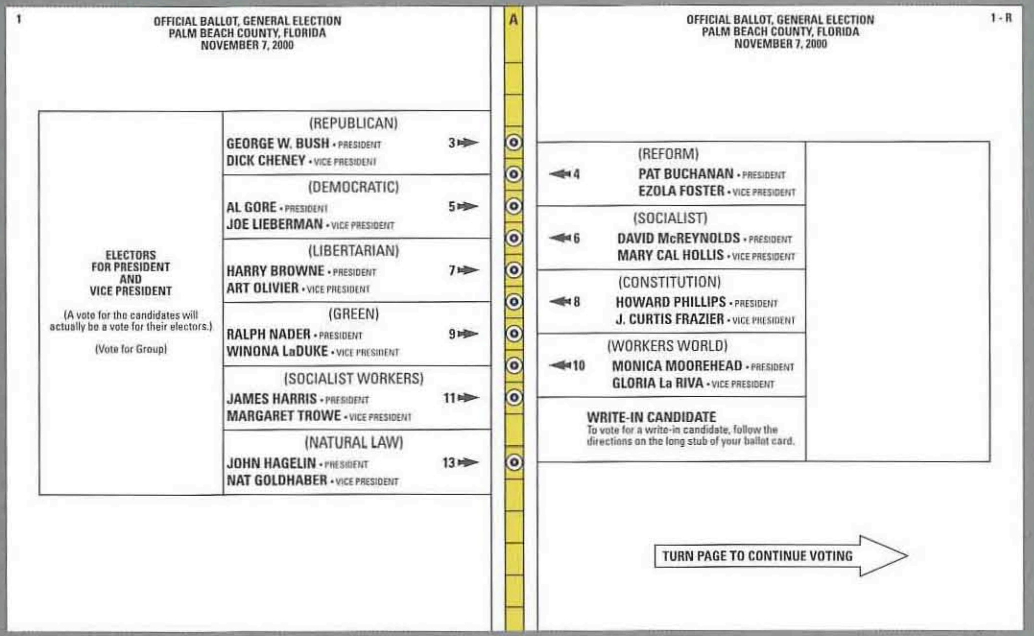
In paragraph text, left-aligned and right-aligned text blocks provide more powerful alignment cues than do center-aligned text blocks. The invisible column created by left-aligned and right-aligned text blocks presents a clear, visual cue against which other elements of the design can be aligned. Center-aligned text blocks, conversely, provide more visually ambiguous alignment cues, and can be difficult to connect with other elements. Justified text provides more alignment cues than unjustified text, and should be used in complex compositions with many elements.

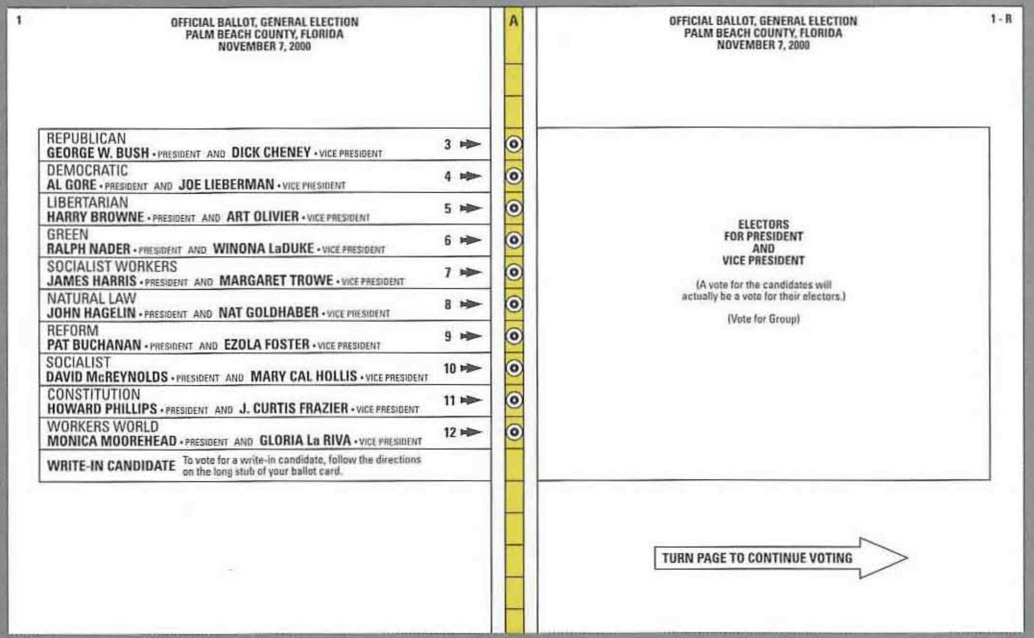
Although alignment is generally defined in terms of rows and columns, more complex forms of alignment exist. In aligning elements along diagonals, for example, the relative angles between the invisible alignment paths should be 30 degrees or greater because separation of less than 30 degrees is too subtle and difficult to detect.[1] In spiral or circular alignments, it may be necessary to augment or highlight the alignment paths so that the alignment is perceptible; otherwise the elements may not seem to be placed according to any particular pattern. As with all such principles of this type, there are exceptions (e.g., the misalignment of elements to attract attention or create tension). However, these exceptions are rare, and alignment should be considered the general rule.

For most designs, align elements into rows and columns or along a centerline. When elements are not arranged in a row/column format, consider highlighting the alignment paths. Use left- or right-justified text to create the best alignment cues, and consider justified text for complex compositions.

See also Aesthetic-Usability Effect and Good Continuation.

[1] See, for example, *Elements of Graph Design* by Stephen M. Kosslyn, W. H. Freeman and Company, 1994, p. 172.





The design of the "butterfly ballot" of Palm Beach County, Florida, may have decided the presidential election of 2000. Although there are several problems with the design of the butterfly ballot, most confusion likely resulted from the misalignment of the row and punch-hole lines. This conclusion is supported by the improbable number of votes for Patrick Buchanan in Palm Beach County, as well as the number of double votes that occurred for candidates that were adjacent on the ballot. A simple adjustment to the ballot design would have dramatically reduced the error rate.

Chunking

A technique of combining many units of information into a limited number of units or chunks, so that the information is easier to process and remember.

The term *chunk* refers to a unit of information in short-term memory—a string of letters, a word, or a series of numbers. The technique of chunking seeks to accommodate short-term memory limits by formatting information into a small number of units. The maximum number of chunks that can be efficiently processed by short-term memory is four, plus or minus one. For example, most people can remember a list of five words for 30 seconds, but few can remember a list of ten words for 30 seconds. By breaking the list of ten words into multiple, smaller chunks (e.g., two groups of three words, and one group of four words), recall performance is essentially equivalent to the single list of five words.[1]

Chunking is often applied as a general technique to simplify designs—a potential misapplication of the principle. The limits specified by this principle deal specifically with tasks involving memory. For example, it is unnecessary and counterproductive to restrict the number of dictionary entries on a page to four or five. Reference-related tasks consist primarily of scanning for a particular item; chunking in this case would dramatically increase the scan time and effort, and yield no benefits.

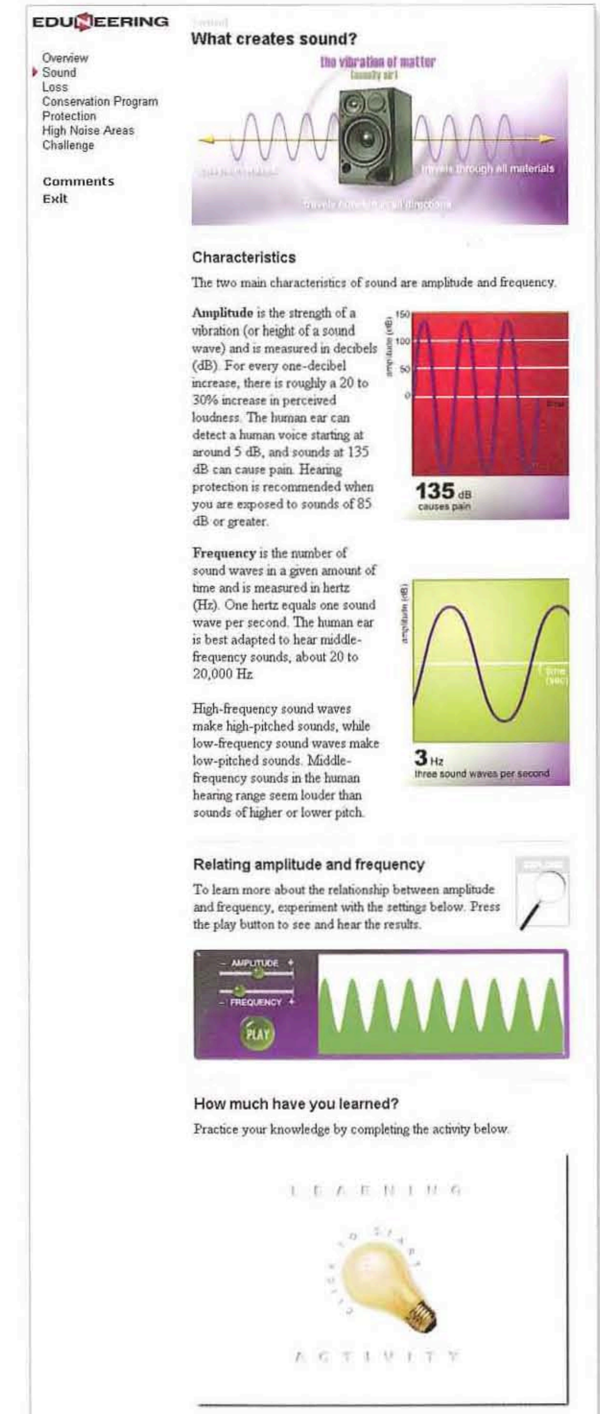
Chunk information when people are required to recall and retain information, or when information is used for problem solving. Do not chunk information that is to be searched or scanned. In environments where noise or stress can interfere with concentration, consider chunking critical display information in anticipation of diminished short-term memory capacity. Use the contemporary estimate of four, plus or minus one chunks when applying this technique.[2]

See also Errors, Mnemonic Device, Performance Load, and Signal-to-Noise Ratio.

[1] The seminal work on short-term memory limits is "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information" by George Miller, *The Psychological Review*, 1956, vol. 63, p. 81–97. As made evident by the title of Miller's paper, his original estimate for short-term memory capacity was 7 ± 2 chunks.

[2] A readable contemporary reference is *Human Memory: Theory and Practice* by Alan Baddeley, Allyn & Bacon, 1997. Regarding short-term memory limits, see, for example, "The Magical Number Four in Short-Term Memory: A Reconsideration of Mental Storage Capacity" by Nelson Cowan, *Behavioral and Brain Sciences*, 2001, vol. 24, p. 87–114.

This e-learning course by EduNeering makes excellent use of chunking. Note that the number of content topics (left gray panel) observes the appropriate limits, as do the information chunks on the topics themselves. Overview and Challenge are not counted because they contain organizing information and quizzes only.



List 1	List 2
angry	thrunced
hoarse	rooped
snuggle	croodle
search	poosk
fatigue	quanked
stutter	maffle
scorch	brizzle
warning	gardyloo
teenager	haspenald
anxious	cark

Familiar words are easier to remember and chunk together than unfamiliar words. Of the two lists, list 1 is easier to recall.

292635732	7045556791
292-63-5732	(704) 555-6791

Large strings of numbers are difficult to recall. Chunking large strings of numbers into multiple, smaller strings can help. Most people can remember their Social Security number and frequently called phone numbers.

Closure

A tendency to perceive a set of individual elements as a single, recognizable pattern, rather than multiple, individual elements.

The principle of closure is one of a number of principles referred to as *Gestalt principles of perception.* It states that whenever possible, people tend to perceive a set of individual elements as a single, recognizable pattern, rather than multiple, individual elements. The tendency to perceive a single pattern is so strong that people will close gaps and fill in missing information to complete the pattern if necessary. For example, when individual line segments are positioned along a circular path, they are first perceived holistically as a circle, and then as comprising multiple, independent elements. The tendency to perceive information in this way is automatic and subconscious; it is likely a function of an innate preference for simplicity over complexity, and pattern over randomness.[1]

Closure is strongest when elements approximate simple, recognizable patterns, such as geometric forms, and are located near one another. When simple, recognizable patterns are not easily perceived, designers can create closure through transitional elements (i.e., subtle visual cues that help direct the eye to find the pattern). Generally, if the energy required to find or form a pattern is greater than the energy required to perceive the elements individually, closure will not occur.

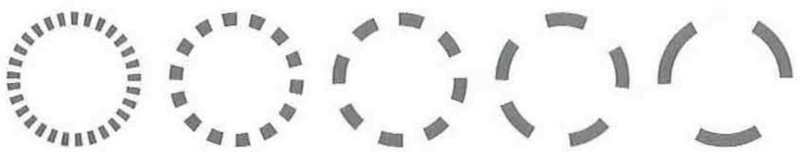
The principle of closure enables designers to reduce complexity by using a smaller number of elements needed to organize and communicate information. For example, a logo design that is composed of recognizable elements does not need to complete many of its lines and contours to be clear and effective. Reducing the number of lines in the logo not only reduces its complexity, but it makes the logo more interesting to look at because viewers subconsciously participate in the completion of its design. Many forms of storytelling leverage closure in a similar way. For example, in comic books, the illustrator presents discrete scenes to the reader, who then supplies what happens in between. The storyline is a unique combination of information provided by the storyteller, and information provided by the reader.[2]

Use closure to reduce the complexity and increase the interestingness of designs. When designs involve simple and recognizable patterns, consider removing or minimizing the elements of a design that can then be supplied by viewers. When designs involve more complex patterns, consider the use of transitional elements to assist viewers in finding or forming the pattern.

See also Good Continuation, Law of Prägnanz, and Proximity.

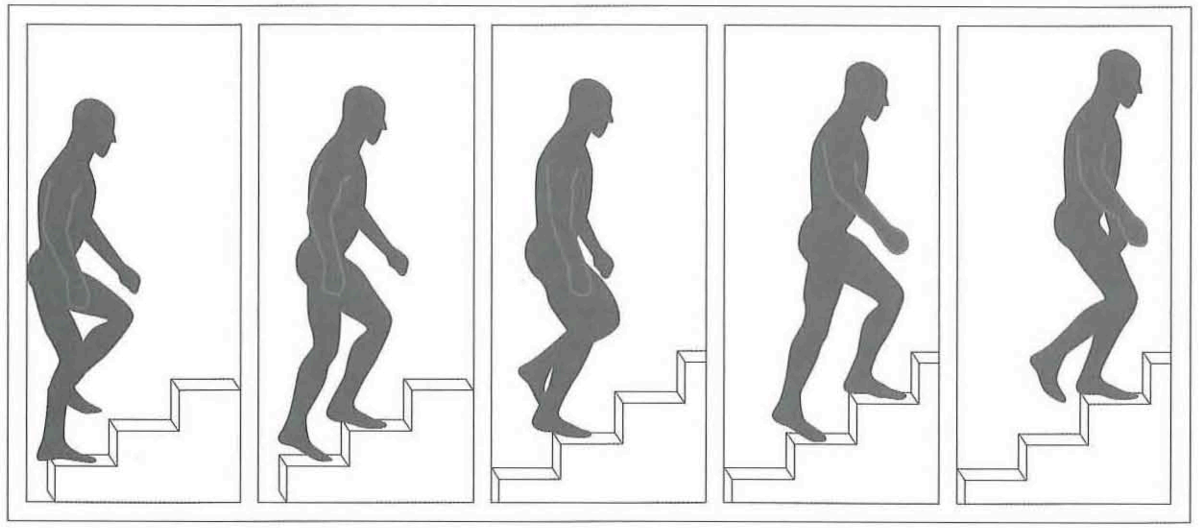
[1] The seminal work on closure is "Untersuchungen zür Lehre von der Gestalt, II" [Laws of Organization in Perceptual Forms] by Max Wertheimer, *Psychologische Forschung,* 1923, vol. 4, p. 301–350, reprinted in *A Source Book of Gestalt Psychology* by Willis D. Ellis (ed.), Routledge & Kegan Paul, 1999, p. 71–88.

[2] See, for example, *Understanding Comics: The Invisible Art* by Scott McCloud, Kitchen Sink Press, 1993.



The elements are perceived holistically as a single pattern first (circle), and then as individual elements.

Elements in text and graphics can be minimized to allow viewers to participate in the completion of the pattern. The result is a more interesting design.



Series images are understood as representing motion because people supply the information in between the images.

Common Fate

Elements that move in the same direction are perceived to be more related than elements that move in different directions or are stationary.

The principle of common fate is one of a number of principles referred to as *Gestalt principles of perception.* It asserts that elements that move together in a common direction are perceived as a single group or chunk, and are interpreted as being more related than elements that move at different times or in different directions. For example, a row of randomly arranged *X*s and *O*s that is stationary is naturally grouped by similarity, *X*s with *X*s, and *O*s with *O*s. However, if certain elements in the row move in one direction, and other elements move in the opposite direction, elements are grouped by their common motion and direction.[1]

Perceived relatedness is strongest when the motion of elements occurs at the same time and velocity, and in the same direction. As any of these factors vary, the elements are decreasingly related. One exception is when the motion exhibits an obvious pattern or rhythm (e.g., wave patterns), in which case the elements are seen as related. Although common fate relationships usually refer to moving elements, they are also observed with static objects that flicker (i.e., elements that alternate between brighter and darker states). For flickering elements, perceived relatedness is strongest when the elements flicker at the same time, frequency, and intensity, or when a recognizable pattern or rhythm is formed.[2]

Common fate relationships influence whether elements are perceived as figure or ground elements. When certain elements are in motion and others are stationary, the moving objects will be perceived as figure elements, and stationary ones will be perceived as ground elements. When elements within a region move together with the bounding edge of the region, the elements and the region will be perceived as the figure. When elements within a region move together, but the bounding edge of the region remains stationary or moves opposite to the elements, the elements within the region will be perceived as the ground.[3]

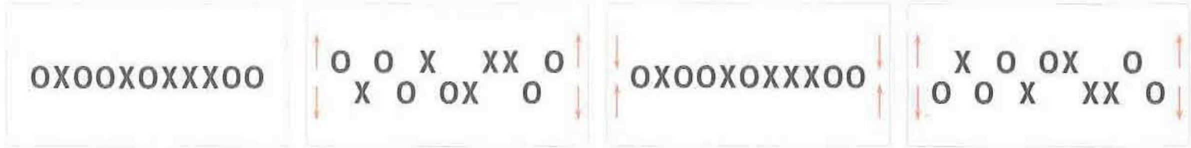
Consider common fate as a grouping strategy when displaying information with moving or flickering elements. Related elements should move at the same time, velocity, and direction, or flicker at the same time, frequency, and intensity. It is possible to group elements when these variables are dissimilar, but only if the motion or flicker forms a recognizable pattern. When moving elements within bounded regions, move the edges of the region in the same direction as the elements to achieve a figure relationship or in the opposite direction as the elements to achieve a ground relationship.

See also Figure-Ground Relationship and Similarity.

[1] The seminal work on common fate is "Untersuchungen zür Lehre von der Gestalt, II" [Laws of Organization in Perceptual Forms] by Max Wertheimer, *Psychologische Forschung*, 1923, vol. 4, p. 301–350, reprinted in *A Source Book of Gestalt Psychology* by Willis D. Ellis (ed.), Routledge & Kegan Paul, 1999, p. 71v88.

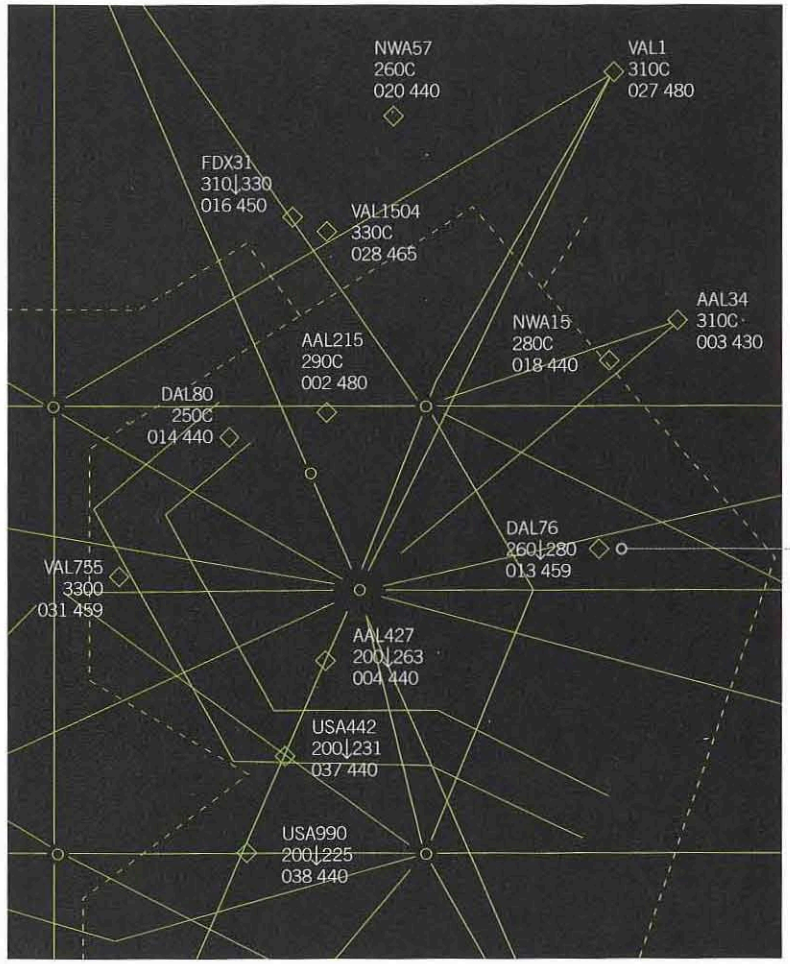
[2] See, for example, "Generalized Common Fate: Grouping by Common Luminance Changes" by Allison B. Sekuler and Patrick J. Bennett, *Psychological Science*, 2001, Vol. 12(6), p. 437–444.

[3] "Common Fate as a Determinant of Figure-Ground Organization" by Joseph Lloyd Brooks, Stanford-Berkeley Talk, 2000, Stanford University, May 16, 2000.



The *X*s and *O*s group by similarity when stationary, such as *X*s with *X*s, *O*s with *O*s. However, when a mix of the *X*s and *O*s move up and down in a common fashion, they are grouped primarily by common fate.

Radar tracking displays use common fate to group tracked aircraft with key information about their identities and headings.



This object is being tracked by radar. The object and its label are visually grouped because they are moving at the same speed and in the same direction.

Confirmation

A technique for preventing unintended actions by requiring verification of the actions before they are performed.[1]

Confirmation is a technique used for critical actions, inputs, or commands. It provides a means for verifying that an action or input is intentional and correct before it is performed. Confirmations are primarily used to prevent a class of errors called *slips*, which are unintended actions. Confirmations slow task performance, and should be reserved for use with critical or irreversible operations. When the consequences of an action are not serious, or when actions are completely and easily reversible, confirmations are not needed. There are two basic confirmation techniques: dialog and two-step operation.[2]

Confirmation using a dialog involves establishing a verbal interaction with the person using the system. It is most commonly represented as a dialog box on a software display (e.g., "Are you sure you want to delete all files?"). In this method, dialog boxes directly ask the user if the action was intended and if they would like to proceed. Confirmations should be used sparingly, or people will learn to ignore them, and become frustrated at the frequent interruption. Dialog messages should be concise but detailed enough to convey accurately the implications of the action. The message should end with one question that is structured to be answered *Yes* or *No*, or with an action verb that conveys the action to be performed (the use of *OK* and *Cancel* should be avoided for confirmations). For less critical confirmations that act more as reminders, an option to disable the confirmation should be provided.

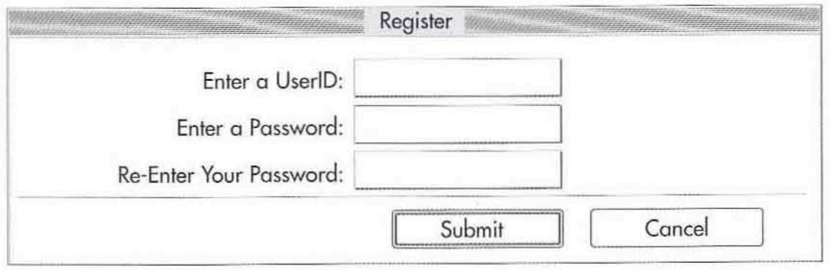
Confirmation using a two-step operation involves a preliminary step that must occur prior to the actual command or input. This is most often used with hardware controls, and is often referred to as an *arm/fire* operation—first you arm the component, and then you fire (execute) it. For example, a switch cover might have to be lifted in order to activate a switch, two people might have to turn two unique keys in order to launch a nuclear weapon, or a control handle in a spacecraft might have to be rotated and then pushed down in order to be activated. The purpose of the two-step operation is to prevent accidental activation of a critical control. If the operation works only when the two-step sequence has been completed, it is unlikely that the operation will occur accidentally. Two-step operations are commonly used for critical operations in aircraft, nuclear power plants, and other environments involving dangerous operations.

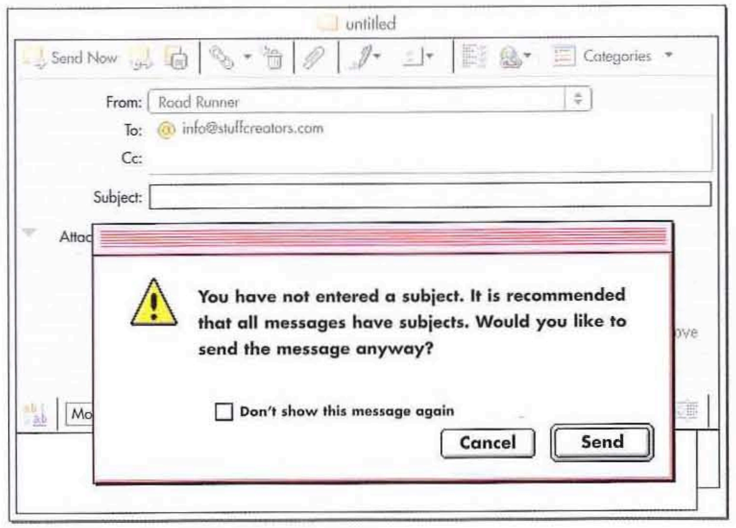
Use confirmations to minimize errors in the performance of critical or irreversible operations. Avoid overusing confirmations to ensure users do not begin to ignore them. Use a two-step operation for hardware confirmations, and a dialog box for software confirmations. Permit less critical confirmations to be disabled after an initial confirmation.

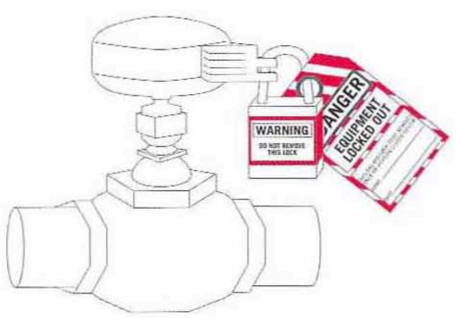
See also Constraint, Errors, Forgiveness, and Garbage In–Garbage Out.

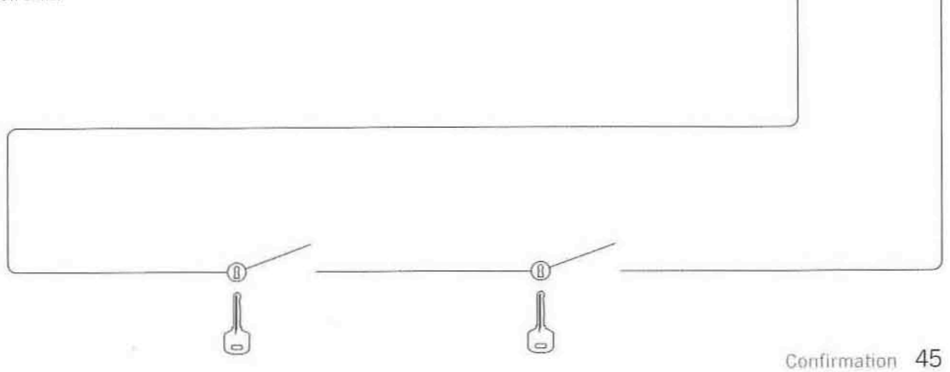
[1] Also known as *verification principle* and *forcing function*.

[2] See, for example, *The Design of Everyday Things* by Donald Norman, Doubleday, 1990; and *To Err Is Human: Building a Safer Health System* edited by Linda T. Kohn, Janet M. Corrigan, and Molla S. Donaldson, National Academy Press, 2000.









Common examples of confirmation strategies include: typing in a password twice to confirm spelling; confirming the intent of an action by clicking the action button (*Send*) with an option to disable future confirmations; having to remove a lock to open a valve; and having to turn two unique keys to complete a launch circuit.

Consistency

The usability of a system is improved when similar parts are expressed in similar ways.

According to the principle of consistency, systems are more usable and learnable when similar parts are expressed in similar ways. Consistency enables people to efficiently transfer knowledge to new contexts, learn new things quickly, and focus attention on the relevant aspects of a task. There are four kinds of consistency: aesthetic, functional, internal, and external.[1]

Aesthetic consistency refers to consistency of style and appearance (e.g., a company logo that uses a consistent font, color, and graphic). Aesthetic consistency enhances recognition, communicates membership, and sets emotional expectations. For example, Mercedes-Benz vehicles are instantly recognizable because the company consistently features its logo prominently on the hood or grill of its vehicles. The logo has become associated with quality and prestige, and informs people how they should feel about the vehicle—i.e., respected and admired.

Functional consistency refers to consistency of meaning and action (e.g., a traffic light that shows a yellow light before going to red). Functional consistency improves usability and learnability by enabling people to leverage existing knowledge about how the design functions. For example, videocassette recorder control symbols, such as for rewind, play, forward, are now used on devices ranging from slide projectors to MP3 music players. The consistent use of these symbols on new devices enables people to leverage existing knowledge about how the controls function, which makes the new devices easier to use and learn.

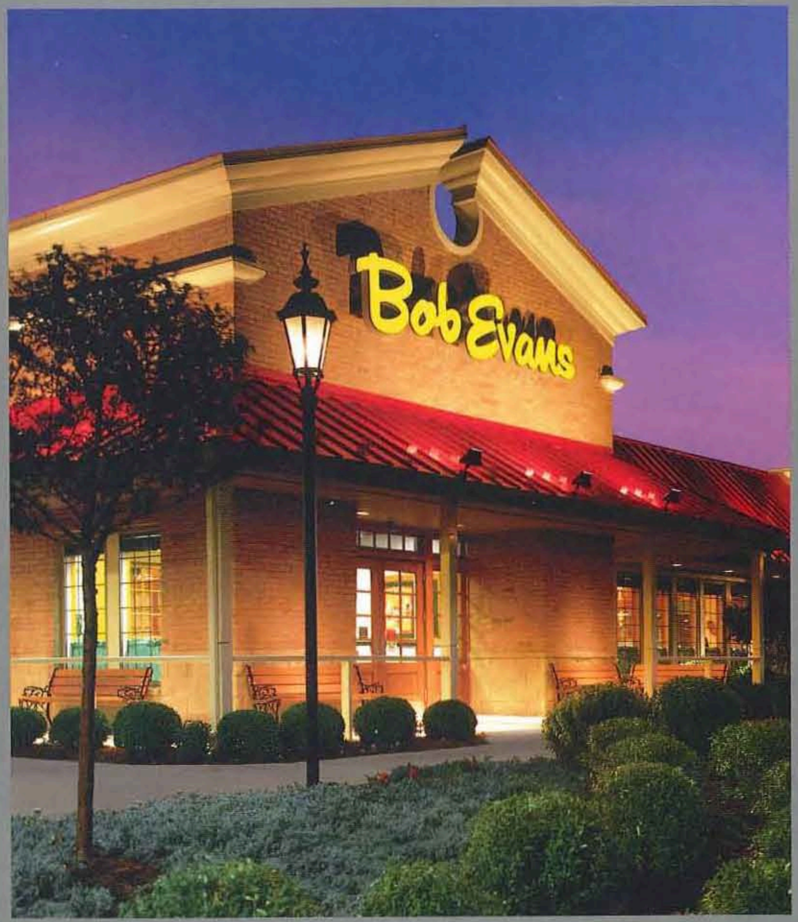
Internal consistency refers to consistency with other elements in the system (e.g., signs within a park are consistent with one another). Internal consistency cultivates trust with people; it is an indicator that a system has been designed, and not cobbled together. Within any logical grouping elements should be aesthetically and functionally consistent with one another.

External consistency refers to consistency with other elements in the environment (e.g., emergency alarms are consistent across different systems in a control room). External consistency extends the benefits of internal consistency across multiple, independent systems. It is more difficult to achieve because different systems rarely observe common design standards.

Consider aesthetic and functional consistency in all aspects of design. Use aesthetic consistency to establish unique identities that can be easily recognized. Use functional consistency to simplify usability and ease of learning. Ensure that systems are always internally consistent, and externally consistent to the greatest degree possible. When common design standards exist, observe them.

See also Modularity, Recognition Over Recall, and Similarity.

[1] Use consistent approaches when possible, but do not compromise clarity or usability for consistency. In the words of Emerson, "A foolish consistency is the hobgoblin of little minds ..."



Restaurant chains frequently use consistency to provide customers with the same experience across many locations. For example, Bob Evans uses the same logo, typefaces, color schemes, menus, staff uniforms, interior design, and architecture across its restaurants. This consistency improves brand recognition, reduces costs, and establishes a relationship with customers that extends beyond any single restaurant.

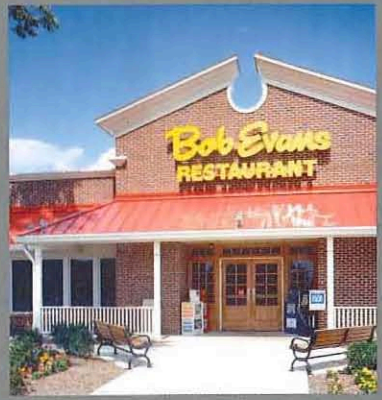


Figure-Ground Relationship

Elements are perceived as either figures (objects of focus) or ground (the rest of the perceptual field).

The figure-ground relationship is one of several principles referred to as *Gestalt principles of perception*. It asserts that the human perceptual system separates stimuli into either figure elements or ground elements. Figure elements are the objects of focus, and ground elements compose an undifferentiated background. This relationship can be demonstrated with both visual stimuli, such as photographs, and auditory stimuli, such as soundtracks with dialog and background music.

When the figure and ground of a composition are clear, the relationship is stable; the figure element receives more attention and is better remembered than the ground. In unstable figure-ground relationships, the relationship is ambiguous and can be interpreted in different ways; the interpretation of elements alternates between figure and ground.

The visual cues that determine which elements will be perceived as figure and which as ground are:

- The figure has a definite shape, whereas the ground is shapeless.
- The ground continues behind the figure.
- The figure seems closer with a clear location in space, whereas the ground seems farther away and has no clear location in space.
- Elements below a horizon line are more likely to be perceived as figures, whereas elements above a horizon line are more likely to be perceived as ground.
- Elements in the lower regions of a design are more likely to be perceived as figures, whereas elements in the upper regions are more likely to be perceived as ground.[2]

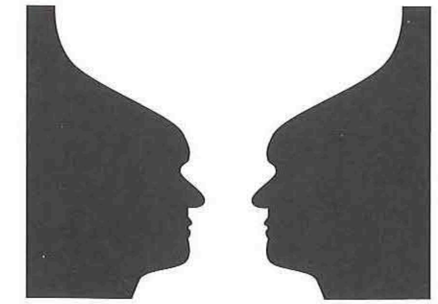
Clearly differentiate between figure and ground in order to focus attention and minimize perceptual confusion. Ensure that designs have stable figure-ground relationships by incorporating the appropriate visual cues listed above. Increase the probability of recall of key elements by making them figures in the composition.

See also Gutenberg Principle, Law of Prägnanz, and Top-Down Lighting Bias.

[1] The seminal work on the figure-ground relationship is "Synoplevede Figurer" [Figure and Ground] by Edgar Rubin, Gyldendalske, 1915, translated and reprinted in *Readings in Perception* by David C. Beardslee and Michael Wertheimer, D. Van Nostrand, 1958, p. 194–203.

[2] "Lower Region: A New Cue for Figure-Ground Assignment" by Shaun P. Vecera, Edward K. Vogel, and Geoffrey F. Woodman, *Journal of Experimental Psychology: General*, 2002, vol. 131(2), p. 194–205.

The Rubin vase is unstable because it can be perceived as a white vase on a black background or two black faces looking at each other on a white background.



Initially, there is no stable figure-ground relationship in this image. However, after a moment, the Dalmatian pops out and the figure-ground relationship stabilizes.

Placing the spa name below the horizon line in the logo makes it a figure element—it will receive more attention and be better remembered than the design that places the name at the top of the logo.

Placing the logo at the bottom of the page makes it a figure element—it will receive more attention and will be better remembered than the logo at the top of the page.

Five Hat Racks

There are five ways to organize information: category, time, location, alphabet, and continuum.[1]

The organization of information is one of the most powerful factors influencing the way people think about and interact with a design. The five hat racks principle asserts that there are a limited number of organizational strategies, regardless of the specific application: category, time, location, alphabet, and continuum.[2]

Category refers to organization by similarity or relatedness. Examples include areas of study in a college catalog, and types of retail merchandise on a Web site. Organize information by category when clusters of similarity exist within the information, or when people will naturally seek out information by category (e.g., a person desiring to purchase a stereo may seek a category for electronic appliances).

Time refers to organization by chronological sequence. Examples include historical timelines and *TV Guide* schedules. Organize information by time when presenting and comparing events over fixed durations, or when a time-based sequence is involved (e.g., a step-by-step procedure).

Location refers to organization by geographical or spatial reference. Examples include emergency exit maps and travel guides. Organize information by location when orientation and wayfinding are important or when information is meaningfully related to the geography of a place (e.g., an historic site).

Alphabet refers to organization by alphabetical sequence. Examples include dictionaries and encyclopedias. Organize information alphabetically when information is referential, when efficient nonlinear access to specific items is required, or when no other organizing strategy is appropriate.

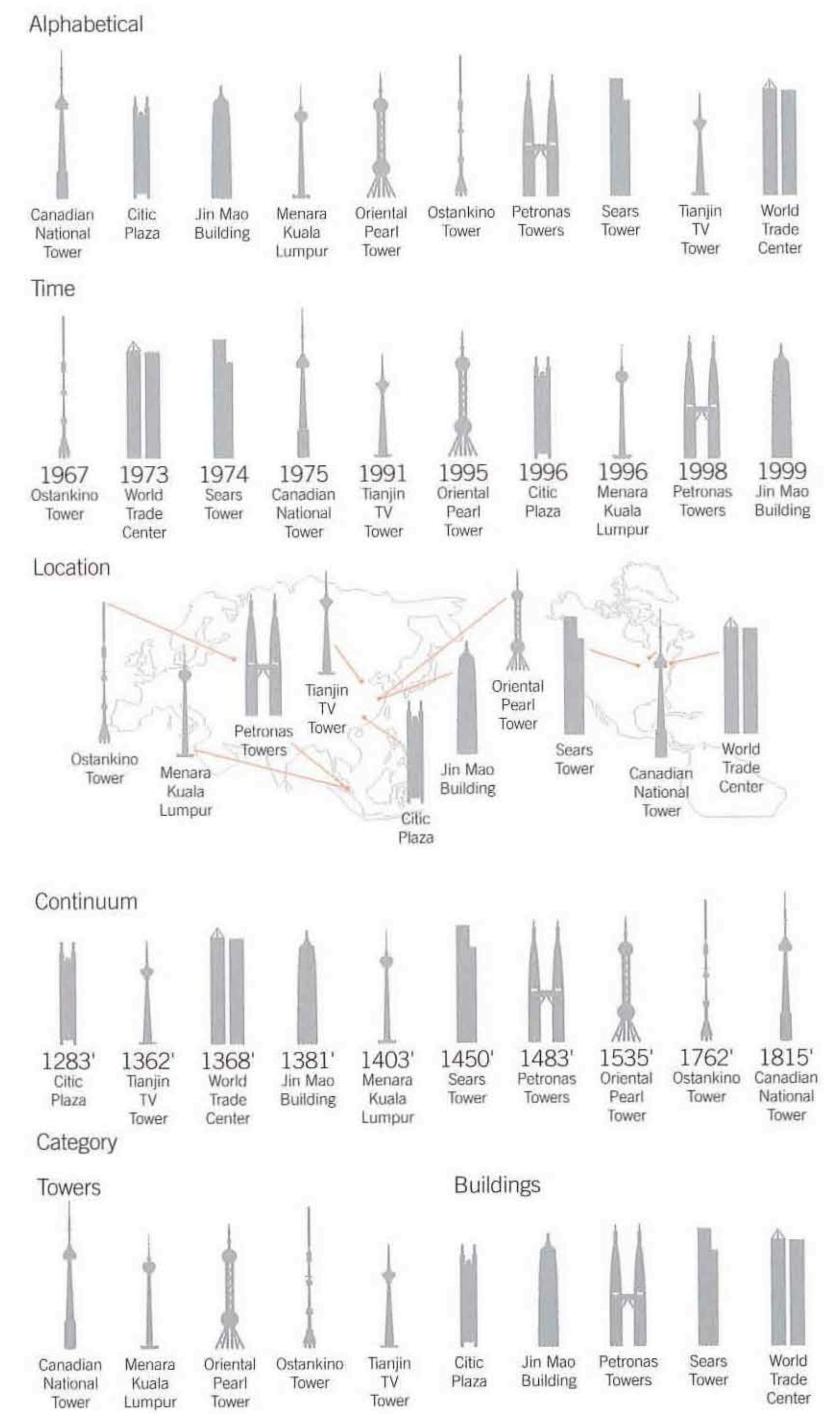
Continuum refers to organization by magnitude (e.g., lowest to highest, worst to best). Examples include baseball batting averages and Internet search engine results. Organize information by continuum when comparing things using a common measure.

See also Advance Organizer, Consistency, and Framing.

[1] The term *hat racks* is built on an analogy—*hats* as information and *racks* as the ways to organize information. Also known as *five ways of organizing information.*

[2] The seminal work on the five hat racks is *Information Anxiety* by Richard Saul Wurman, Bantam Books, 1990. Note that Wurman changed the hat rack title of *continuum* to *hierarchy* in a later edition of the book, which permits the acronym LATCH. The original title *continuum* is presented here, as the authors believe it to be a more accurate description of the category.

The five hat racks are applied here to the tallest structures in the world. Although the same information is presented in each case, the different organizations dramatically influence which aspects of the information are emphasized.



Flexibility-Usability Tradeoff

As the flexibility of a system increases, its usability decreases.

The flexibility-usability tradeoff is exemplified in the well-known maxim, *jack of all trades, master of none*. Flexible designs can perform more functions than specialized designs, but they perform the functions less efficiently. Flexible designs are, by definition, more complex than inflexible designs, and as a result are generally more difficult to use. For example, a Swiss Army Knife has many attached tools that increase its flexibility. These tools are less usable and efficient than corresponding individual tools, but taken together provide a flexibility of use not available from any single tool. The flexibility-usability tradeoff exists because accommodating flexibility entails satisfying a larger set of design requirements, which invariably means not only more compromises, but also more complexity in the design.[1]

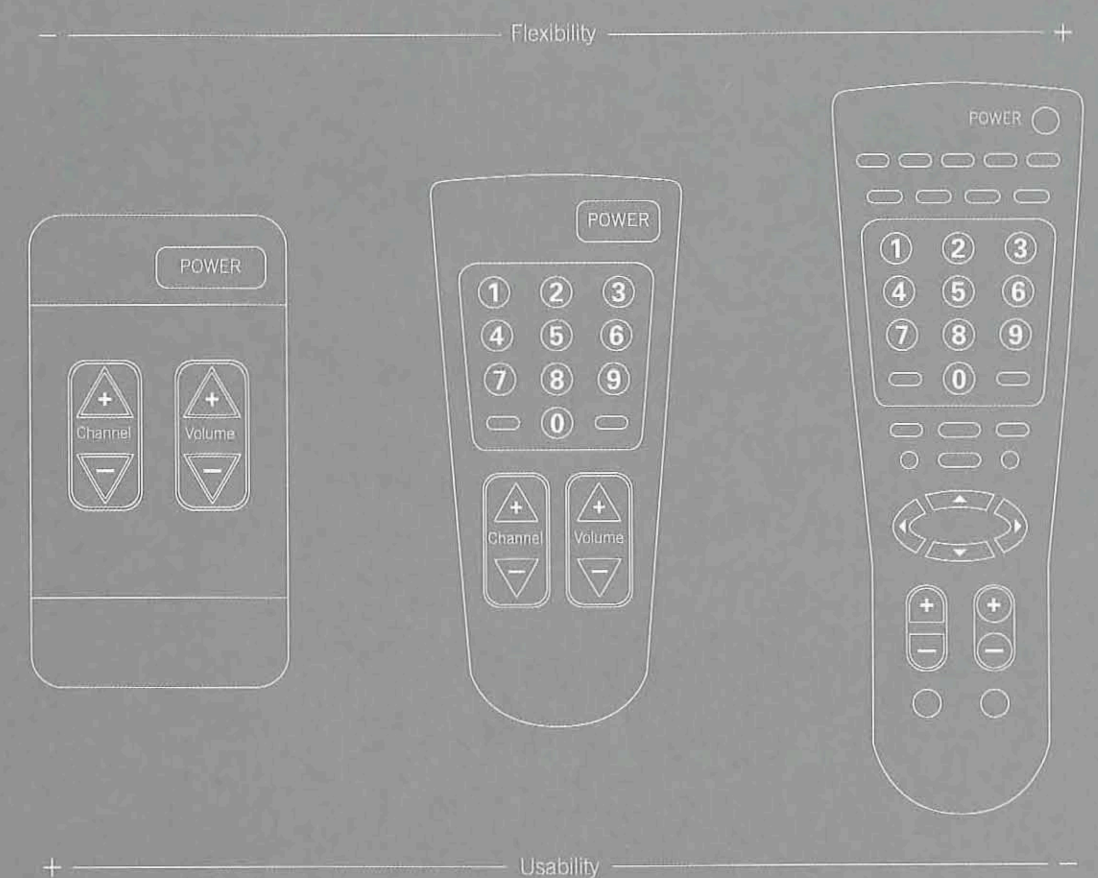
It is a common design mistake to assume that designs should always be made as flexible as possible. In reality, however, flexibility has costs in terms of decreased efficiency, added complexity, increased time, and money for development. Flexibility generally pays dividends only when an audience cannot clearly anticipate its future needs. For example, personal computers are flexible devices that are difficult to use, relative to more specialized devices like video game players. However, the primary value of a personal computer is strength as a multipurpose device that can be used in many different ways: word processing, tax preparation, e-mail. People purchase video game players to play games, but they purchase personal computers to satisfy a variety of needs, many of which are unknown at the time of purchase.

The ability of an audience to anticipate future uses of a product is a key indicator of how they will value flexibility versus usability in design. When an audience can clearly anticipate its needs, specialized designs that target those needs will be more successful. When an audience cannot clearly define its needs, more flexible designs that enable people to address future contingencies will be more successful. The degree to which an audience can or cannot define future needs should correspond to the degree of specialization or flexibility in the design. As an audience comes to understand the range of possible needs that can be satisfied, their needs become better defined and, consequently, the designs need to become more specialized. This shift from flexibility toward specialization over time is a general pattern observed in the evolution of all systems, and should be considered in the life cycle of products.

The tradeoff between flexibilty and usability has implications for weighing the relative importance of flexibility versus usability in a design. When an audience has a clear understanding of its needs, favor specialized designs that target those needs as efficiently as possible. When an audience has a poor understanding of its needs, favor flexible designs to address the broadest possible set of future applications. When designing multiple generations of products, consider a general shift toward specialization as audience needs become more defined.

See also 80/20 Rule, Convergence, Hierarchy of Needs, Life Cycle, Modularity, and Progressive Disclosure.

[1] See, for example, *The Invisible Computer* by Donald A. Norman, MIT Press, 1999; and "The Visible Problems of the Invisible Computer: A Skeptical Look at Information Appliances" by Andrew Odlyzko, First Monday, 1999, vol. 4 (9), http://www.firstmonday.org.



There is a basic tradeoff between flexibility and usability, as demonstrated by these remote control designs. The simple remote control is the easiest to use, but not very flexible. Conversely, the universal remote control is very flexible but far more complex and difficult to use.

Forgiveness

Designs should help people avoid errors and minimize the negative consequences of errors when they do occur.

Human error is inevitable, but it need not be catastrophic. Forgiveness in design helps prevent errors before they occur, and minimizes the negative consequences of errors when they do occur. Forgiving designs provide a sense of security and stability, which in turn, fosters a willingness to learn, explore, and use the design. Common strategies for incorporating forgiveness in designs include:

Good Affordances—physical characteristics of the design that influence its correct use (e.g., uniquely shaped plug that can only be inserted into the appropriate receptacle).

Reversibility of Actions—one or more actions can be reversed if an error occurs or the intent of the person changes (e.g., undo function in software).

Safety Nets—device or process that minimizes the negative consequences of a catastrophic error or failure (e.g., pilot ejection seat in aircraft).

Confirmation—verification of intent that is required before critical actions are allowed (e.g., lock that must be opened before equipment can be activated).

Warnings—signs, prompts, or alarms used to warn of imminent danger (e.g., road signs warning of a sharp turn ahead).

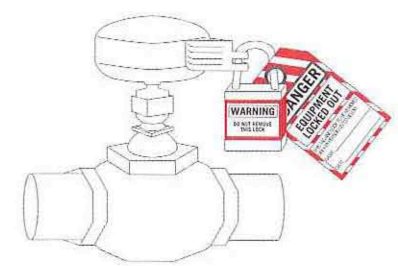
Help—information that assists in basic operations, troubleshooting, and error recovery (e.g., documentation or help line).

The preferred methods of achieving forgiveness in a design are affordances, reversibility of actions, and safety nets. Designs that effectively use these strategies require minimal confirmations, warnings, and help—i.e., if the affordances are good, help is less necessary; if actions are reversible, confirmations are less necessary; if safety nets are strong, warnings are less necessary. When using confirmations, warnings, and help systems, avoid cryptic messages or icons. Ensure that messages clearly state the risk or problem, and also what actions can or should be taken. Keep in mind that too many confirmations or warnings impede the flow of interaction and increase the likelihood that the confirmation or warning will be ignored.

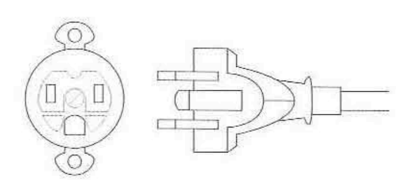
Create forgiving designs by using good affordances, reversibility of actions, and safety nets. If this is not possible, be sure to include confirmations, warnings, and a good help system. Be aware that the amount of help necessary to successfully interact with a design is inversely proportional to the quality of the design—if a lot of help is required, the design is poor.

See also Affordance, Confirmation, Errors, and Factor of Safety.

Road signs make roads more forgiving by warning drivers of impending hazards.

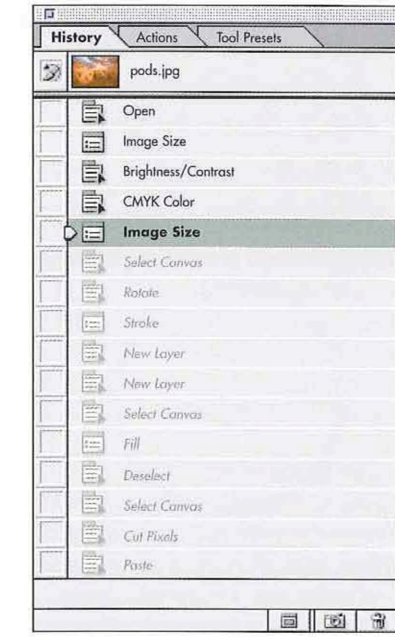


Locking and tagging equipment is a common confirmation strategy to ensure that people do not accidentally engage systems under repair.



The good affordance of this plug prevents it from being inserted into the socket improperly.

The Adobe Photoshop History palette enables users to flexibly undo and redo their previous actions.



In case of a catastrophic failure, the ballistic recovery system acts as a safety net, enabling the pilot and craft to return safely to earth.

Form Follows Function

Beauty in design results from purity of function.

The *form follows function* axiom is interpreted in one of two ways—as a description of beauty or a prescription for beauty. The descriptive interpretation is that beauty results from purity of function and the absence of ornamentation. The prescriptive interpretation is that aesthetic considerations in design should be secondary to functional considerations. The axiom was adopted and popularized by modernist architects in the early 20th century, and has since been adopted by designers in a variety of disciplines.[1]

The *descriptive* interpretation—i.e., that beauty results from purity of function—was originally based on the belief that form follows function in nature. In reality, however, it is quite the opposite—function follows form in nature, if it follows anything at all. (Evolution by natural selection transmits no *intention* from one generation to the next; genetic patterns are simply passed on and it is left to each organism to find use of the form that it has inherited.) Nonetheless, functional aspects of a design are less subjective than purely aesthetic aspects and, therefore functional criteria present a clearer and more objective criteria for judgement of quality. The result is designs that are more timeless and enduring, but also frequently perceived by general audiences as simple and uninteresting.[2]

The *prescriptive* interpretation—i.e., that aesthetic considerations in design should be secondary to functional considerations—was likely derived from the descriptive interpretation. The use of *form follows function* as a prescription or design guideline is problematic in that it focuses the designer on the wrong question. The question should not be, "What aspects of form should be omitted or traded for function?" but rather, "What aspects of the design are critical to success?" These success criteria, not a blind allegiance to form or function, should drive design specifications and decisions. When time and resources are limited, design tradeoffs should be based on what does the least harm to the probability of success, however *success* is defined. In certain circumstances, primary aesthetic considerations will be sacrificed, and in others, primarily functional ones will be

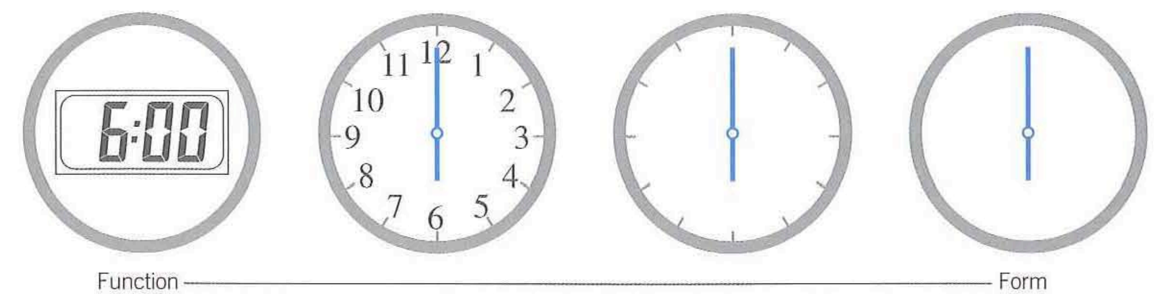
Use the descriptive interpretation of *form follows function* as an aesthetic guide, but do not apply the prescriptive interpretation as a strict design rule. When making design decisions, focus on the relative importance of all aspects of the design—form and function—in light of the success criteria.

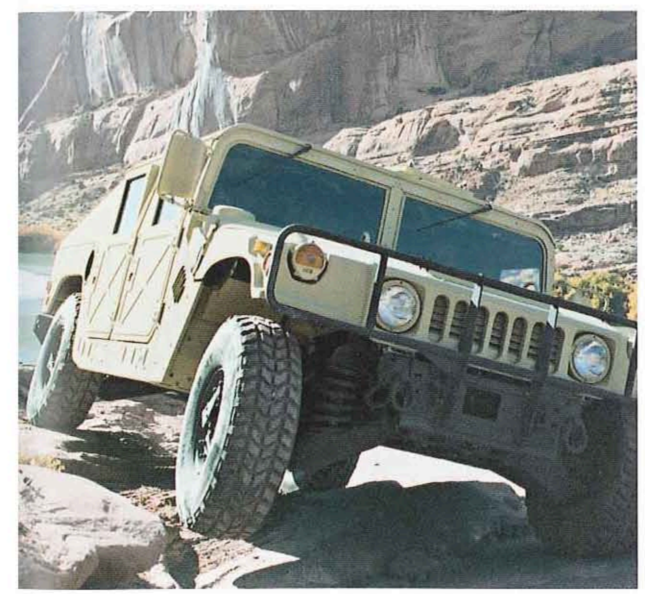
See also Aesthetic-Usability Effect, Exposure Effect, and Ockham's Razor.

[1] The origin of the concept is attributed to the 18th century Jesuit monk Carlo Lodoli. His theories on architecture likely influenced later designers like Horatio Greenough and Louis Sullivan who then articulated the concept in popular form. The seminal works on *form follows function* are "The Tall Office Building Artistically Considered" by Louis H. Sullivan, *Lippincott's Magazine*, March 1896; and *Form Follows Fiasco: Why Modern Architecture Hasn't Worked* by Peter Blake, Little, Brown, and Company, 1977.

[2] The tendency of general audiences to resist the *new* is a function of their familiarity with the *old*. It often takes several generations to erode population biases sufficiently such that the merits of a new design can be objectively considered.

Defining success criteria is essential to good design. For example, if the success criteria for a watch are defined in terms of speed and accuracy, the digital display is superior. If the success criteria are defined in terms of pure aesthetics, the minimalist analog display is superior (the pure function of the digital display has not yet translated to a popular aesthetic for general audiences). In all cases, the success criteria should direct design decisions and trade-offs, and should be the primary consideration in determining the specifications for a design.





Perhaps no purer functional form exists than the original Humvee. Born out of military specifications, the success of the Humvee in combat led to the commercial successors—Hummer H1 and H2. Each represents a unique and compelling aesthetic that results from purity of function and minimal ornamentation.

Gutenberg Diagram

A diagram that describes the general pattern followed by the eyes when looking at evenly distributed, homogeneous information.[1]

The Gutenberg diagram divides a display medium into four quadrants: the *primary optical area* at the top left, the *terminal area* at the bottom right, the *strong fallow area* at the top right, and the *weak fallow area* at the bottom left. According to the diagram, Western readers naturally begin at the primary optical area and move across and down the display medium in a series of sweeps to the terminal area. Each sweep begins along an *axis of orientation*—a horizontal line created by aligned elements, text lines, or explicit segments—and proceeds in a left-to-right direction. The strong and weak fallow areas lie outside this path and receive minimal attention unless visually emphasized. The tendency to follow this path is metaphorically attributed to *reading gravity*—the left-right, top-bottom habit formed from reading.[2]

Designs that follow the diagram work in harmony with reading gravity, and return readers to a logical axis of orientation, purportedly improving reading rhythm and comprehension. For example, a layout following the Gutenberg diagram would place key elements at the top left (e.g., headline), middle (e.g., image), and bottom right (e.g., contact information). Though designs based directly or indirectly on the Gutenberg diagram are widespread, there is little empirical evidence that it contributes to improved reading rates or comprehension.

The Gutenberg diagram is likely only predictive of eye movement for heavy text information, evenly distributed and homogeneous information, and blank pages or displays. In all other cases, the weight of the elements of the design in concert with their layout and composition will direct eye movements. For example, if a newspaper has a very heavy headline and photograph in its center, the center will be the primary optical area. Familiarity with the information and medium also influences eye movements. For example, a person who regularly views information presented in a consistent way is more likely to first look at areas that are often changing (e.g., new top stories) than areas that are the same (e.g., the title of a newspaper).

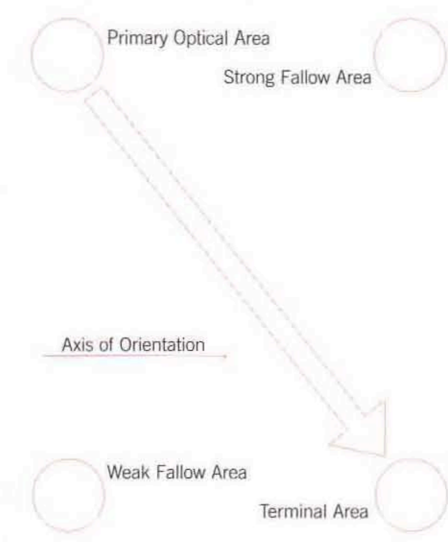
Consider the Gutenberg diagram to assist in layout and composition when the elements are evenly distributed and homogeneous, or the design contains heavy use of text. Otherwise, use the weight and composition of elements to lead the eye.

See also Alignment, Entry Point, Progressive Disclosure, and Serial Position Effects.

[1] Also known as the *Gutenberg rule* and the *Z pattern of processing*.

[2] The seminal work on the Gutenberg diagram is attributed to the typographer Edmund Arnold, who is said to have developed the concept in the 1950s. See, for example, *Type & Layout: How Typography and Design Can Get Your Message Across or Get in the Way*, by Colin Wheildon, Strathmoor Press, 1995.

Reading gravity pulls the eyes from the top-left to the bottom-right of the display medium. In homogeneous displays, the Gutenberg diagram makes compositions interesting and easy to read. However, in heterogeneous displays, the Gutenberg diagram does not apply, and can constrain composition unnecessarily.




THE WALL STREET JOURNAL.

What's News—

Bush Says Rising Price of Oil Could Hurt Pace of Recovery

Saudis Welcome Foreigners to Work —But Not to Die

Insurer's Tactic: If You Get Sick, The Premium Rises

What's New in Today's Journal

PERSONAL JOURNAL

The redesign of the *Wall Street Journal* leads the eyes of readers, and does not follow the Gutenberg diagram. Additionally, recurring readers of the *Wall Street Journal* tend to go to the section they find most valuable, ignoring the other elements of the page.

The composition of these pages illustrates the application of the Gutenberg diagram. The first page is all text, and it is, therefore, safe to assume readers will begin at the top-left and stop at the bottom-right of the page. The pull quote is placed between these areas, reinforcing reading gravity. The placement of the image on the second page similarly reinforces reading gravity, which it would not do if it were positioned at the top-right or bottom-left of the page.


CHAPTER 1: Down the Rabbit-Hole

Alice was beginning to get very tired of sitting by her sister on the bank, and of having nothing to do: once or twice she had peeped into the book her sister was reading, but it had no pictures or conversations in it, 'and what is the use of a book,' thought Alice 'without pictures or conversation?'

So she was considering in her own mind (as well as she could, for the hot day made her feel very sleepy and stupid), whether the pleasure of making a daisy-chain would be worth the trouble of getting up and picking the daisies, when suddenly a White Rabbit with pink eyes ran close by her.

There was nothing so very remarkable in that; nor did Alice think it so very much out of the way to hear the Rabbit say to itself, 'Oh dear! Oh dear! I shall be late!' (when she thought it over afterwards, it occurred to her that she ought to have wondered at this, but at the time it all seemed quite natural); but when the Rabbit actually took a watch out of its waistcoat-pocket, and looked at it, and then hurried on, Alice started to her feet, for it flashed across her mind that she had never before seen a rabbit with either a waistcoat-pocket, or a watch to take out of it, and burning with curiosity, she ran across the field after it, and fortunately was just in time to see it pop down a large rabbit-hole under the hedge.

In another moment down went Alice after it, never once considering how in the world she was to get out again.

The rabbit-hole went straight on like a tunnel for some way, and then dipped suddenly down, so suddenly that Alice had not a moment to think about stopping herself before she found herself falling down a very deep well.

Either the well was very deep, or she fell very slowly, for she had plenty of time as she went down to look about her and to wonder what was going to happen next. First, she tried to look down and make out what she was coming to, but it was too dark to see anything; then she looked at the sides of the well, and noticed that they were filled with cupboards and book-shelves; here and there she saw maps and pictures hung upon pegs. She took down a jar from one of the shelves as she passed; it was labelled 'ORANGE MARMALADE', but to her great disappointment it was empty: she did not like to drop the jar for fear of killing somebody, so managed to put it into one of the cupboards as she fell past it.

'Well!' thought Alice to herself, 'after such a fall as this, I shall think nothing of tumbling down stairs! How brave they'll all think me at home! Why, I wouldn't say anything about it, even if I fell off the top of the house!' (Which was very likely true.)

Down, down, down. Would the fall never come to an end!

'Well!' thought Alice to herself, 'after such a fall as this, I shall think nothing of tumbling down stairs! How brave they'll all think me at home! Why, I wouldn't say anything about it, even if I fell off the top of the house!'

Alice's Adventures in Wonderland, by Lewis Carroll, first appeared in print in 1865.


CHAPTER 1: Down the Rabbit-Hole

Alice was beginning to get very tired of sitting by her sister on the bank, and of having nothing to do: once or twice she had peeped into the book her sister was reading, but it had no pictures or conversations in it, 'and what is the use of a book,' thought Alice 'without pictures or conversation?'

So she was considering in her own mind (as well as she could, for the hot day made her feel very sleepy and stupid), whether the pleasure of making a daisy-chain would be worth the trouble of getting up and picking the daisies, when suddenly a White Rabbit with pink eyes ran close by her.

There was nothing so very remarkable in that; nor did Alice think it so very much out of the way to hear the Rabbit say to itself, 'Oh dear! Oh dear! I shall be late!' (when she thought it over afterwards, it occurred to her that she ought to have wondered at this, but at the time it all seemed quite natural); but when the Rabbit actually took a watch out of its waistcoat-pocket, and looked at it, and then hurried on, Alice started to her feet, for it flashed across her mind that she had never before seen a rabbit with either a waistcoat-pocket, or a watch to take out of it, and burning with curiosity, she ran across the field after it, and fortunately was just in time to see it pop down a large rabbit-hole under the hedge.

In another moment down went Alice after it, never once considering how in the world she was to get out again.

The rabbit-hole went straight on like a tunnel for some way, and then dipped suddenly down, so suddenly that Alice had not a moment to think about stopping herself before she found herself falling down a very deep well.

Either the well was very deep, or she fell very slowly, for she had plenty of time as she went down to look about her and to wonder what was going to happen next. First, she tried to look down and make out what she was coming to, but it was too dark to see anything; then she looked at the sides of the well, and noticed that they were filled with cupboards and book shelves; here and there she saw maps and pictures hung upon pegs.

Hick's Law

The time it takes to make a decision increases as the number of alternatives increases.[1]

Hick's Law states that the time required to make a decision is a function of the number of available options. It can be used to estimate how long it will take for people to make a decision when presented with multiple choices. For example, when a pilot has to press a particular button in response to some event, such as an alarm, Hick's Law predicts that the greater the number of alternative buttons, the longer it will take to make the decision and select the correct one. Hick's Law has implications for the design of any system or process that requires simple decisions to be made based on multiple options.[2]

All tasks consist of four basic steps: (1) identify a problem or goal, (2) assess the available options to solve the problem or achieve the goal, (3) decide on an option, and (4) implement the option. Hick's Law applies to the third step: decide on an option. However, the law does not apply to decisions that involve significant levels of searching, reading, or complex problem solving. For example, a task that has only three options, but requires reading sentences and intense concentration can easily take longer than a simple stimulus-response task with six options. Therefore, Hick's Law is most applicable to simple decision-making tasks in which there is a unique response to each stimulus. (For example, if A happens, then push button 1, If B happens, then push button 2.) The law becomes less applicable as the complexity of tasks increases.[3]

Designers can improve the efficiency of a design by understanding the implications of Hick's Law. For example, the law applies to the design of software menus, control displays, wayfinding layout and signage, and emergency response training—anywhere simple decisions are involved. Hick's Law does not apply to complex menus or hierarchies of options. Menu selection of this type is not a simple decision-making task since it typically involves reading sentences, searching and scanning for options, and some level of problem solving.

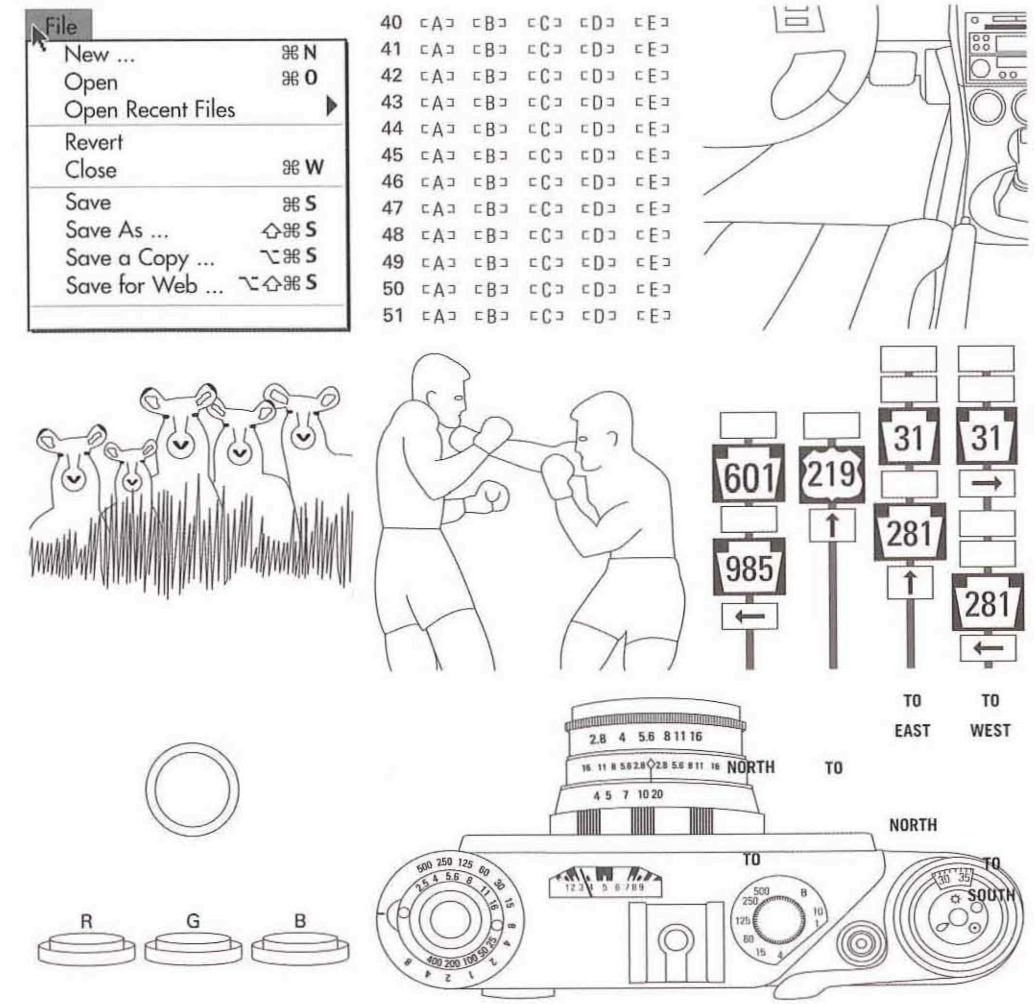
Keep Hick's Law in mind when designing systems that involve decisions based on a set of options. When designing for time-critical tasks, minimize the number of options involved in a decision to reduce response times and minimize errors. When designs require complex interactions, do not rely on Hick's Law to make design decisions; rather, test designs on the target population using realistic scenarios. In training people to perform time-critical procedures, train on the fewest possible responses for a given scenario. This will minimize response times, error rates, and training costs.

See also Errors, Fitts' Law, Progressive Disclosure, and Wayfinding.

[1] Also known as *Hick-Hyman Law*.

[2] The seminal work on Hick's Law is "On the Rate of Gain of Information" by W. E. Hick, *Quarterly Journal of Experimental Psychology*, 1952, vol. 4, p. 11–26; and "Stimulus information as a determinant of reaction time" by Ray Hyman, *Journal of Experimental Psychology*, 1953, vol. 45, p. 188–196.

[3] The Hick's Law equation is $RT = a + b\log^2(n)$, where RT = response time, a = the total time that is not involved with decision making, b = an empirically derived constant based on the cognitive processing time for each option (in this case ≈ 0.155 seconds for humans), n = number of equally probable alternatives. For example, assume it takes 2 seconds to detect an alarm and understand it's meaning. Further, assume that pressing one of five buttons will solve the problem caused by the alarm. The time to respond would be RT = (2 sec) + (0.155 sec)(log²(5)) = 2.36 sec.



Menus
The time for a person to select an item from a simple software menu increases with the number of items. However, this may not be the case for more complex menus involving a lot of text or submenus.

Predatory Behavior
The time for a predator to target a prey increases with the number of potential prey.

Simple Tasks
The time for a person to press the correct button (R, G, or B) depending on the color of the light (red, green, or blue) increases with the number of possible colors.

Test Options
Hick's Law does not apply to tasks involving significant levels of reading and problem solving, as in taking an exam.

Martial Arts
The time for a martial artist to block a punch increases with the number of known blocking techniques.

Device Settings
The time for a person to make simple decisions about adjustments on a device increases with the number of controls. This may not be the case for more complex decisions or combinations of settings.

Braking
The time for a driver to press the brake to avoid hitting an unexpected obstacle increases if there is a clear opportunity to steer around the obstacle.

Road Signs
As long as road signs are not too dense or complex, the time for a driver to make a turn based on a particular road sign increases with the total number of road signs.

Ockham's Razor

Given a choice between functionally equivalent designs, the simplest design should be selected.[1]

Ockham's razor asserts that simplicity is preferred to complexity in design. Many variations of the principle exist, each adapted to address the particulars of a field or domain of knowledge. A few examples include:

- "Entities should not be multiplied without necessity."—William of Ockham
- "That is better and more valuable which requires fewer, other circumstances being equal."—Robert Grosseteste
- "Nature operates in the shortest way possible."—Aristotle
- "We are to admit no more causes of natural things than such as are both true and sufficient to explain their appearances."—Isaac Newton
- "Everything should be made as simple as possible, but not simpler." —Albert Einstein

Implicit in Ockham's razor is the idea that unnecessary elements decrease a design's efficiency, and increase the probability of unanticipated consequences. Unnecessary weight, whether physical, visual, or cognitive, degrades performance. Unnecessary design elements have the potential to fail or create problems. There is also an aesthetic appeal to the principle, which likens the "cutting" of unnecessary elements from a design to the removal of impurities from a solution—the design is a cleaner, purer result.

Use Ockham's razor to evaluate and select among multiple, functionally equivalent designs. Functional equivalence here refers to comparable performance of a design on common measures. For example, given two functionally equivalent displays—equal in information content and readability—select the display with the fewest visual elements. Evaluate each element within the selected design and remove as many as possible without compromising function. Finally, minimize the expression of the remaining elements as much as possible without compromising function.[2]

See also Form Follows Function, Mapping, and Signal-to-Noise Ratio.

[1] Also known as *Occam's razor, law of parsimony, law of economy,* and *principle of simplicity.* The term "Ockham's razor" references William of Ockham, a 14th century Franciscan friar and logician who purportedly made abundant use of the principle. The principle does not actually appear in any of his extant writings and, in truth, little is known about either the origin of the principle or its originator. See, for example, "The Myth of Occam's Razor" by W. M. Thorburn, *Mind*, 1918, vol. 27, p. 345–353.

[2] "Make all visual distinctions as subtle as possible, but still clear and effective." *Visual Explanations* by Edward R. Tufte, Graphics Press, 1998, p. 73.

While other Internet search services were racing to add advertising and ad hoc functions to their Web sites, Google kept its design simple and efficient. The result is the best performing and easiest to use search service on the Web.

The Yamaha Compact Silent Electric Cello is a minimalist cello with only those portions touched by the player represented. Musicians can hear concert-quality cello sound through headphones while creating little external sound, or through an amplifier and speakers for public performances. The cello can also be collapsed for easy transport and storage.

The Taburet M Stacking Stool is strong, comfortable, and stackable. It is constructed from a single piece of molded wood and has no extraneous elements.

Picture Superiority Effect

Pictures are remembered better than words.[1]

It is said that a picture is worth a thousand words, and it turns out that in most cases, this is true. Pictures are generally more easily recognized and recalled than words, although memory for pictures and words together is superior to memory for words alone or pictures alone. For example, instructional materials and technical manuals that present textual information accompanied by supporting pictures enable information recall that is better than that produced by either the text or pictures alone. The picture superiority effect is commonly used in instructional design, advertising, technical writing, and other design contexts requiring easy and accurate recall of information.[2]

When information recall is measured immediately after the presentation of a series of pictures or words, recall performance for pictures and words is equal. The picture superiority effect applies only when people are asked to recall something after more than thirty seconds from the time of exposure. The picture superiority effect is strongest when the pictures represent common, concrete things versus abstract things, such as a picture of a flag versus a picture depicting the concept of freedom, and when pictures are distinct from one another, such as a mix of objects versus objects of a single type.

The picture superiority effect advantage increases further when people are casually exposed to information and the exposure time is limited. For example, an advertisement for a clock repair shop that includes a picture of a clock will be better recalled than the same advertisement without the picture. People not interested in clock repair who see the advertisement with the picture will also be better able to recall the brand if the need for clock repair service arises at a later time. The strength of the picture superiority effect diminishes as the information becomes more complex. For example, people are able to recall events from a story presented as a silent movie as well as events from the same story read as text.[3]

Use the picture superiority effect to improve the recognition and recall of key information. Use pictures and words together, and ensure that they reinforce the same information for optimal effect. Pictures and words that conflict create interference and dramatically inhibit recall. Consider the inclusion of meaningful pictures in advertising campaigns when possible, especially when the goal is to build company and product brand awareness.

See also Advance Organizer, Iconic Representation, Serial Position Effects, and von Restorff Effect.

[1] Also known as *pictorial superiority effect*.

[2] The seminal work on the picture superiority effect is "Why Are Pictures Easier to Recall than Words?" by Allan Paivio, T. B. Rogers, and Padric C. Smythe, *Psychonomic Science*, 1968, vol. 11(4), p. 137–138.

[3] See, for example, "Conditions for a Picture-Superiority Effect on Consumer Memory" by Terry L. Childers and Michael J. Houston, *Journal of Consumer Research*, 1984, vol. 11, p. 643–654.

Advertisements with text and pictures are more likely to be looked at and recalled than advertisements with text only. This superiority of pictures over text is even stronger when the page is quickly scanned rather than read.

Progressive Disclosure

A strategy for managing information complexity in which only necessary or requested information is displayed at any given time.

Progressive disclosure involves separating information into multiple layers and only presenting layers that are necessary or relevant. It is primarily used to prevent information overload, and is employed in computer user interfaces, instructional materials, and the design of physical spaces.[1]

Progressive disclosure keeps displays clean and uncluttered and helps people manage complexity without becoming confused, frustrated, or disoriented. For example, infrequently used controls in software interfaces are often concealed in dialog boxes that are invoked by clicking a *More* button. People who do not need to use the controls never see them. For more advanced users, the options are readily available. In either case, the design is simplified by showing only the most frequently required controls by default, and making additional controls available on request.[2]

Learning efficiency benefits greatly from the use of progressive disclosure. Information presented to a person who is not interested or ready to process it is effectively noise. Information that is gradually and progressively disclosed to a learner as they need or request it is better processed and perceived as more relevant. The number of errors is significantly reduced using this method, and consequently the amount of time and frustration spent recovering from errors is also reduced.[3]

Progressive disclosure is also used in the physical world to manage the perception of complexity and activity. For example, progressive disclosure is found in the design of entry points for modern theme park rides. Exceedingly long lines not only frustrate people in line, but also discourage new people from the ride. Theme park designers progressively disclose discrete segments of the line (sometimes supplemented with entertainment), so that no one, in or out of the line, ever sees the line in its entirety.

Use progressive disclosure to reduce information complexity, especially when people interacting with the design are novices or infrequent users. Hide infrequently used controls or information, but make them readily available through some simple operation, such as pressing a *More* button. Progressive disclosure is also an effective method for leading people though complex procedures, and should be considered when such procedures are a part of a design.

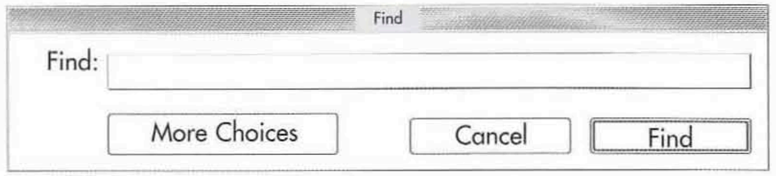
See also Chunking, Errors, Layering, and Performance Load.

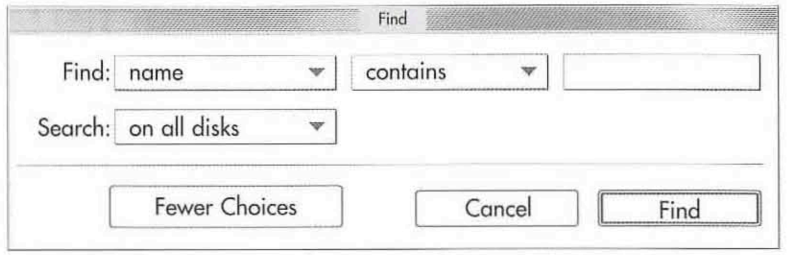
[1] The seminal applied work on progressive disclosure is the user interface for the Xerox Star computer. See "The Xerox 'Star': A Retrospective" by Jeff Johnson and Teresa L. Roberts, William Verplank, David C. Smith, Charles Irby, Marian Beard, Kevin Mackey, in *Human Computer Interaction: Toward the Year 2000* by Ronald M. Baecker, Jonathan Grudin, William A. S. Buxton, Saul Greenberg, Morgan Kaufman Publishers, 1995, p. 53–70.

[2] A common mistake is to present all available information and options at once with the rationale that it reduces kinematic load. Since progressive disclosure affects only infrequently used elements and elements for which a person may not be ready, it will generally have minimal effect on kinematic load. Conversely, presenting everything at once will significantly increase cognitive load.

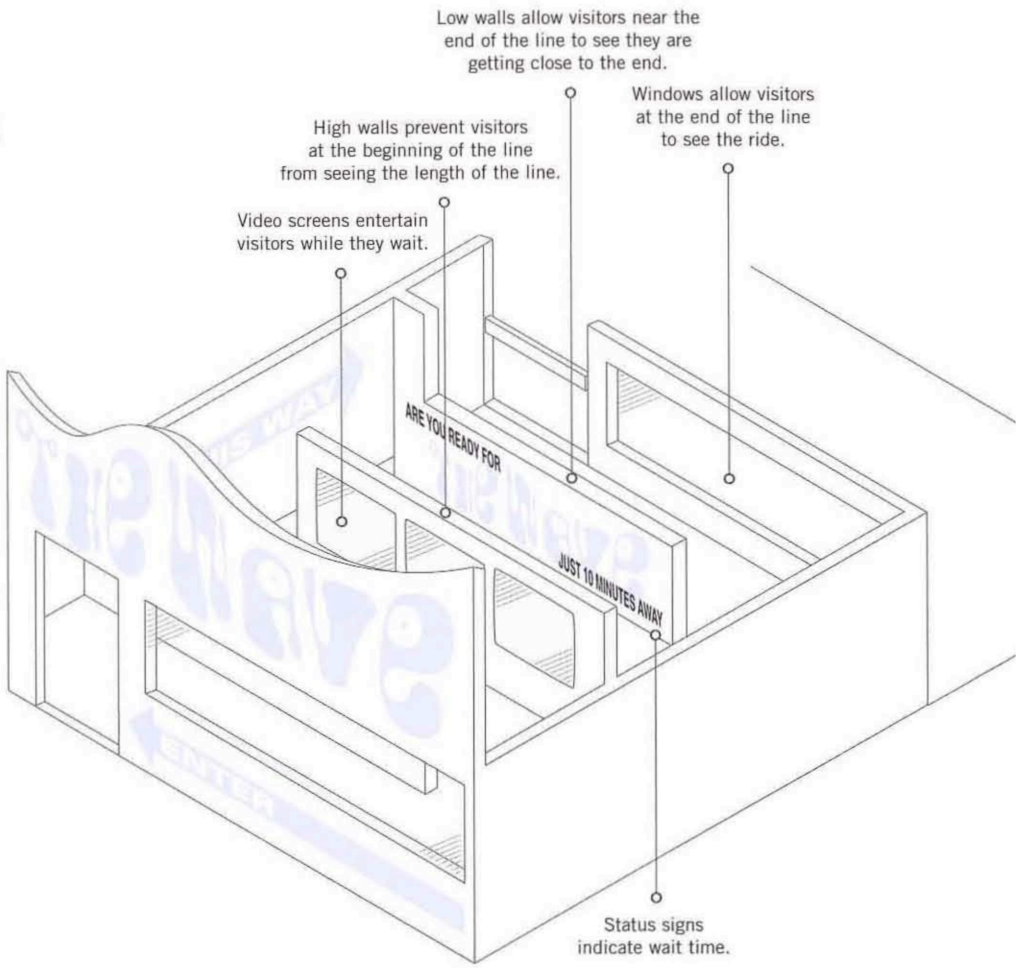
[3] See, for example, "Training Wheels in a User Interface" by John M. Carroll and Caroline Carrithers, *Communications of the ACM*, 1984, vol. 27(8), p. 800–806; and *The Nurnberg Funnel* by John M. Carroll, MIT Press, 1990.

Progressive disclosure is commonly used in software to conceal complexity. In this dialog box, basic search functionality is available by default. However, more complex search functionality is available upon request by clicking *More Choices*.





Theme park rides often have very long lines—so long that seeing the lines in their entirety would scare away many would-be visitors. Therefore, modern theme park rides progressively disclose the length of the line, so that only small segments of the line can be seen from any particular vantage point. Additional distractions are provided in the form of video screens, signage, and partial glimpses of people on the ride.



Recognition Over Recall

Memory for recognizing things is better than memory for recalling things.

People are better at recognizing things they have previously experienced than they are at recalling those things from memory. It is easier to recognize things than recall them because recognition tasks provide memory cues that facilitate searching through memory. For example, it is easier to provide the correct answer for a multiple-choice question than it is for a fill-in-the-blank question because multiple-choice questions provide a list of possible answers; the range of search possibilities is narrowed to just the list of options. Short answer questions provide no such memory cues, so the range of search possibilities is much greater.[1]

Recognition memory is much easier to develop than recall memory. Recognition memory is obtained through exposure, and does not necessarily involve any memory about origin, context, or relevance. It is simply memory that something (sight, sound, smell, touch) has been experienced before. Recall memory is obtained through learning, usually involving some combination of memorization, practice, and application. Recognition memory is also retained for longer periods of time than recall memory. For example, the name of an acquaintance may be difficult to recall, but is easily recognized when heard.

The advantages of recognition over recall are often put to good use in the design of interfaces for complex systems. For example, early computer systems used a command line interface, which required recall memory for hundreds of commands. The effort associated with learning the commands made computers difficult to use. The contemporary graphical user interface, which presents commands in menus, allows users to browse the possible options, and select from them accordingly. This eliminates the need to have the commands in recall memory, and greatly simplifies the usability of computers.

Decision-making is also strongly influenced by recognition. A familiar option is often selected over an unfamiliar option, even when the unfamiliar option may be the best choice. For example, in a consumer study, people participating in a taste-test rated a known brand of peanut butter as superior to two unknown brands, even though one of the unknown brands was *objectively* better, as determined by earlier blind taste tests. Recognition of an option is often a sufficient condition for making a choice.[2]

Minimize the need to recall information from memory whenever possible. Use readily accessible menus, decision aids, and similar devices to make available options clearly visible. Emphasize recognition memory in training programs, and the development of brand awareness in advertising campaigns.

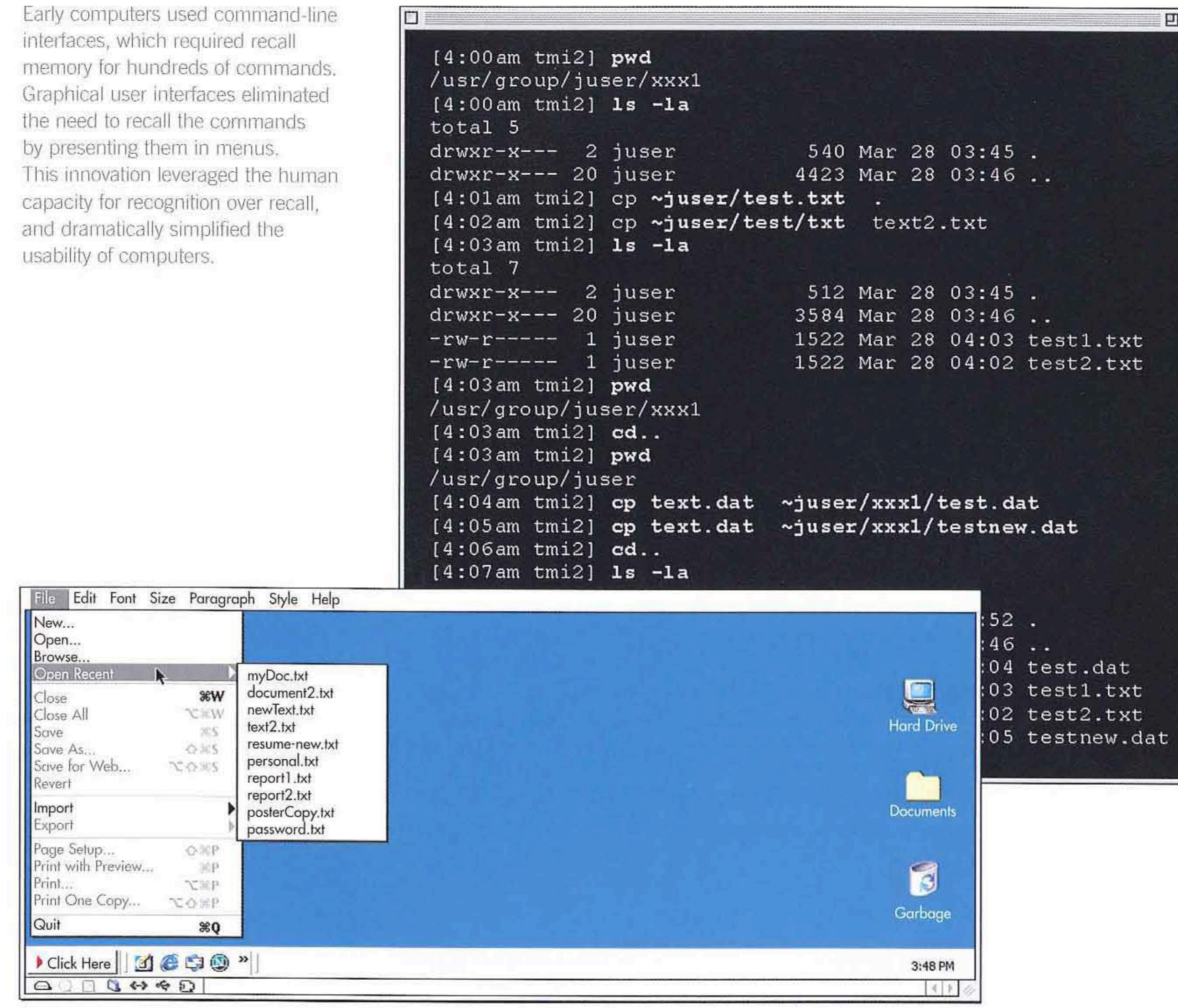
See also Exposure Effect, Serial Position Effects, and Visibility.

[1] The seminal applied work on recognition over recall is the user interface for the Xerox Star computer. See "The Xerox 'Star': A Retrospective" by Jeff Johnson and Teresa L. Roberts, William Verplank, David C. Smith, Charles Irby, Marian Beard, Kevin Mackey, in *Human Computer Interaction: Toward the Year 2000* by Ronald M. Baecker, Jonathan Grudin, William A. S. Buxton, Saul Greenberg, Morgan Kaufman Publishers, 1995, p. 53–70.

[2] Note that none of the participants had previously bought or used the known brand. See "Effects of Brand Awareness on Choice for a Common, Repeat-Purchase Product" by Wayne D. Hoyer and Steven P. Brown, *Journal of Consumer Research*, 1990, vol. 17, p. 141–148.

ex, edit, e
exit
expand, unexp
expr
file
find
finger
fmt, fmt_mail
fold
ftp
gcore
gprof
grep
groups
gzip
gunzip
hashmake,
hashcheck
head
history
imake
indent
install
join
kill
last
ld, ld.so
leave
less
lex
lint
ln
login
look
lookbib
lorder
lp, cancel
lpq
lpr
lprm
ls
mail, Mail
make
man
mesg
mkdir, md
mkstr
more, page
mv
nawk
nice
nm
nroff
od
passwd,
chfn, chsh
paste
pr
printenv
prof
ps
ptx
pwd
quota

Early computers used command-line interfaces, which required recall memory for hundreds of commands. Graphical user interfaces eliminated the need to recall the commands by presenting them in menus. This innovation leveraged the human capacity for recognition over recall, and dramatically simplified the usability of computers.



Visibility

The usability of a system is improved when its status and methods of use are clearly visible.

According to the principle of visibility, systems are more usable when they clearly indicate their status, the possible actions that can be performed, and the consequences of the actions once performed. For example, a red light could be used to indicate whether or not a device is receiving power; illuminated controls could be used to indicate controls that are currently available; and distinct auditory and tactile feedback could be used to acknowledge that actions have been performed and completed. The principle of visibility is based on the fact that people are better at recognizing solutions when selecting from a set of options, than recalling solutions from memory. When it comes to the design of complex systems, the principle of visibility is perhaps the most important and most violated principle of design.[1]

To incorporate visibility into a complex system, one must consider the number of conditions, number of options per condition and number of outcomes—the combinations can be overwhelming. This leads many designers to apply a type of kitchen-sink visibility—i.e., they try to make everything visible all of the time. This approach may seem desirable, but it actually makes the relevant information and controls more difficult to access due to an overload of information.[2]

Hierarchical organization and context sensitivity are good solutions for managing complexity while preserving visibility. Hierarchical organization puts controls and information into logical categories, and then hides them within a parent control, such as a software menu. The category names remain visible, but the controls and information remain concealed until the parent control is activated. Context sensitivity reveals and conceals controls and information based on different system contexts. Relevant controls and information for a particular context are made highly visible, and irrelevant controls (e.g., unavailable functions), are minimized or hidden.

Visible controls and information serve as reminders for what is and is not possible. Design systems that clearly indicate the system status, the possible actions that can be performed, and the consequences of the actions performed. Immediately acknowledge user actions with clear feedback. Avoid kitchen-sink visibility. Make the degree of visibility of controls and information correspond to their relevance. Use hierarchical organization and context sensitivity to minimize complexity and maximize visibility.

See also Affordance, Mapping, Mental Model, Modularity, Progressive Disclosure, and Recognition Over Recall.

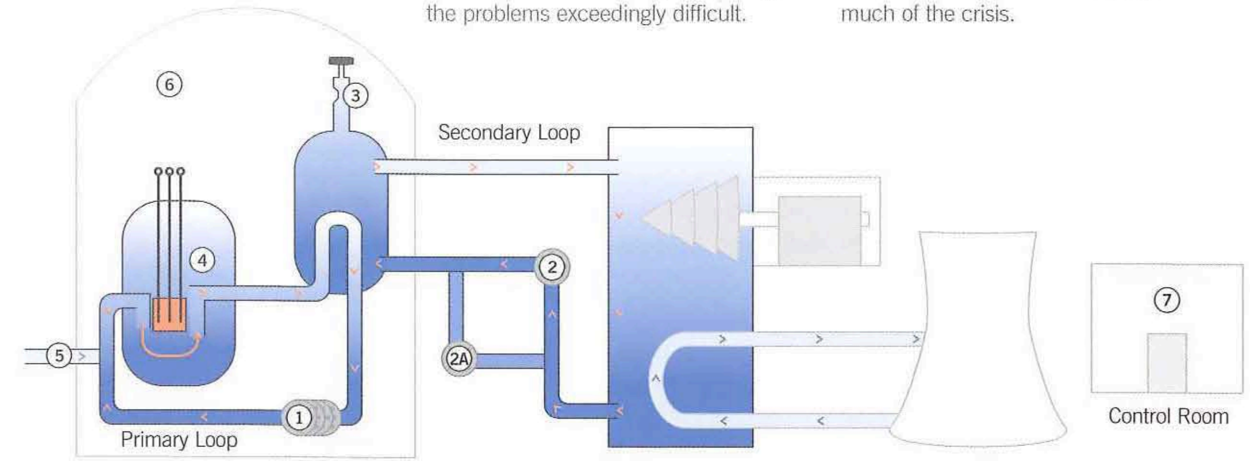
[1] The seminal work on visibility is *The Design of Everyday Things* by Donald Norman, Doubleday, 1990.

[2] The enormity of the number of visibility conditions is why visibility is among the most violated of the design principles—it is, quite simply, difficult to accommodate all of the possibilities of complex systems.

Three Mile Island Unit 2
Harrisburg, Pennsylvania
March 28, 1979, 4:00 A.M.

Visibility of complex systems is essential for problem solving—especially in times of stress. An analysis of key events of the TMI accident reveals a number of blind spots in the system that made understanding and solving the problems exceedingly difficult.

To further complicate matters, alarms were blaring, lights were flashing, and critical system feedback was routed to a printer that could only print 15 lines a minute—status information was more than an hour behind for much of the crisis.



TIME	PLACE	EVENT
00:00:00	2	Coolant pumps in the secondary loop malfunction and shut down.
00:00:02		Temperature and pressure in the primary loop increase.
00:00:03	3	The pressure release valve (PORV) opens automatically to lower the pressure.
00:00:04	2A	Backup pumps automatically turn on.
	2A 7	**Operators do not know that the backup pumps are disconnected.**
00:00:09	4	The control rods are lowered to slow the nuclear chain reaction and reduce the temperature.
00:00:11	3 7	The PORV light goes out in the control room, indicating that the PORV closed.
	3 7	**Operators cannot see that the PORV is stuck open. Steam and water is released through the PORV.**
00:02:00	5	Emergency water is automatically injected into the primary loop to keep the water at a safe level.
00:04:30	5 7	Instruments in the control room indicate that water level in the primary loop is rising. Operators shut down the emergency water injection.
	3 7	**Operators cannot see that the water level in the primary loop is actually dropping. Steam and water continue to be released through the PORV.**
00:08:00	2A 7	An operator notices that the backup pumps are not working. He connects the pumps and they begin operating normally.
01:20:00	1 7	Pumps in the primary loop vibrate violently because of steam in the line. Two of four pumps are shut down.
01:40:00	1	The other two pumps shut down. Temperature and pressure in the primary loop continue to rise.
02:15:00	3 4	The water level drops below the core. Radioactive gas is released through the PORV.
02:20:00	3 7	An operator notices that the temperature at the PORV is high. He stops the leak by shutting a PORV backup valve.
	7	**Operators still cannot see that the water level in the primary loop is actually dropping.**
02:45:00		Radiation alarms sound and a site emergency is declared. The level of radioactivity in the primary loop is over 300 times the normal level.
07:30:00	3 5 7	Operators pump water into the primary loop, but cannot bring the pressure down. They open the backup valve to the PORV to lower pressure.
09:00:00	6	An explosion occurs in the containment structure.
	7	**Operators cannot see that an explosion occurred. They attribute the noise and instrument readings to an electrical malfunction.**
15:00:00	1	The pumps in the primary loop are reactivated. Temperatures decline and the pressure lowers. Disaster is averted—except, of course, for the leaking radiation.