NBS BUILDING SCIENCE SERIES 141

The Development and Evaluation of Effective Symbol Signs

ARTMENT OF COMMERCE • NATIONAL BUREAU OF STANDARDS

NATIONAL BUREAU OF STANDARDS

The National Bureau of Standards[1] was established by an act of Congress on March 3, 1901. The Bureau's overall goal is to strengthen and advance the Nation's science and technology and facilitate their effective application for public benefit. To this end, the Bureau conducts research and provides: (1) a basis for the Nation's physical measurement system, (2) scientific and technological services for industry and government, (3) a technical basis for equity in trade, and (4) technical services to promote public safety. The Bureau's technical work is performed by the National Measurement Laboratory, the National Engineering Laboratory, and the Institute for Computer Sciences and Technology.

THE NATIONAL MEASUREMENT LABORATORY provides the national system of physical and chemical and materials measurement; coordinates the system with measurement systems of other nations and furnishes essential services leading to accurate and uniform physical and chemical measurement throughout the Nation's scientific community, industry, and commerce; conducts materials research leading to improved methods of measurement, standards, and data on the properties of materials needed by industry, commerce, educational institutions, and Government; provides advisory and research services to other Government agencies; develops, produces, and distributes Standard Reference Materials; and provides calibration services. The Laboratory consists of the following centers:

Absolute Physical Quantities[2] — Radiation Research — Chemical Physics — Analytical Chemistry — Materials Science

THE NATIONAL ENGINEERING LABORATORY provides technology and technical services to the public and private sectors to address national needs and to solve national problems; conducts research in engineering and applied science in support of these efforts; builds and maintains competence in the necessary disciplines required to carry out this research and technical service; develops engineering data and measurement capabilities; provides engineering measurement traceability services; develops test methods and proposes engineering standards and code changes; develops and proposes new engineering practices; and develops and improves mechanisms to transfer results of its research to the ultimate user. The Laboratory consists of the following centers:

Applied Mathematics — Electronics and Electrical Engineering[2] — Manufacturing Engineering — Building Technology — Fire Research — Chemical Engineering[2]

THE INSTITUTE FOR COMPUTER SCIENCES AND TECHNOLOGY conducts research and provides scientific and technical services to aid Federal agencies in the selection, acquisition, application, and use of computer technology to improve effectiveness and economy in Government operations in accordance with Public Law 89-306 (40 U.S.C. 759), relevant Executive Orders, and other directives; carries out this mission by managing the Federal Information Processing Standards Program, developing Federal ADP standards guidelines, and managing Federal participation in ADP voluntary standardization activities; provides scientific and technological advisory services and assistance to Federal agencies; and provides the technical foundation for computer-related policies of the Federal Government. The Institute consists of the following centers:

Programming Science and Technology — Computer Systems Engineering.

[1]Headquarters and Laboratories at Gaithersburg, MD, unless otherwise noted;
mailing address Washington, DC 20234.
[2]Some divisions within the center are located at Boulder, CO 80303.

NBS BUILDING SCIENCE SERIES 141

The Development and Evaluation of Effective Symbol Signs

Belinda Lowenhaupt Collins

Center for Building Technology
National Engineering Laboratory
National Bureau of Standards
Washington, DC 20234

U.S. DEPARTMENT OF COMMERCE, Malcolm Baldrige, Secretary
NATIONAL BUREAU OF STANDARDS, Ernest Ambler, Director

Issued May 1982

Library of Congress Catalog Card Number: 81-600192

National Bureau of Standards Building Science Series 141
Nat. Bur. Stand. (U.S.), Bldg. Sci. Ser. 141, 96 pages (May 1982)
CODEN: BSSNBV

U.S. GOVERNMENT PRINTING OFFICE
WASHINGTON: 1982

For sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, DC 20402
Price
(Add 25 percent for other than U.S. mailing)

The Development and Evaluation of
Effective Symbol Signs

by

Belinda Lowenhaupt Collins

ABSTRACT

Graphic symbols have recently been widely adopted for sign systems in the United States. Beginning with traffic sign systems, symbols have become widely used for applications ranging from products to buildings. In this report a brief history of the development of symbols is given, followed by a review of research on experimental evaluation of symbols. Some of the general advantages and limitations of symbols are discussed, along with graphic considerations essential in the development of effective symbols. Research on symbols for five areas of application -- highway, automotive/machinery, public information, product hazard, and safety -- is then discussed.

Finally, issues in the research and development of more effective symbols are reviewed. These include the need for good graphic design, characteristics of the intended user group, use of shape and color to encode information, and general visibility considerations.

Key words: communication; design issues; hazard; pictograms; pictorial; safety; signs; standards; symbols; visual alerting; warning.

ACKNOWLEDGMENTS

While the author of this report takes full responsibility for its contents, I wish to thank all who provided assistance and insight during its many revisions. In particular, I am grateful to Robert Chapman, Roy Clark, Jacqueline Elder, Neil Lerner, Steve Margulis, Brian Pierman, and Mary Reppert for their detailed review and helpful comments.

Cover photo:

A sample of safety symbols.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

SI CONVERSION UNITS

The units and conversion factors given in this table are in agreement with the International System of Units or SI system (Systèm International d'Unites). Because the United States is a signatory to the 11th General Conference on Weights and Measures, which defined and gave official status to the SI system, the following conversion factors are given.

Length

1 in = 0.0254* meter
1 ft = 0.3048* meter

Area

1 in^2 = 6.4516* x 10^{-4} $meter^2$
1 ft^2 = 0.0929 meter

Illumination

1 ft candle = 10.76 lux

---

* Exactly

Facing page:

Symbols designed to provide public information.

1. INTRODUCTION

1.1 OVERVIEW

The development and use of more effective symbols as elements of a communication system requires both good design and thorough evaluation. In the following pages, a brief history of the use of symbols in sign systems is presented as a background to the discussion of graphic and experimental evaluation which follows. Advantages and limitations to the use of symbols in sign systems are noted.

If symbols are to be effective, then explicit attention must be paid to good graphic design. Some of the principles followed by graphic designers are discussed, with three case studies of design evaluation for symbols in sign systems. Design evaluation does not always include scientific sampling of the effectiveness of the symbolic image with the intended user population. Scientific evaluation of symbols is typically done separately, often without feedback to the design process. Similarly, standardization of symbols often proceeds without input from either the graphic design profession or the research community.

Yet, review of the reported research into symbols for various sign systems provides valuable insight into numerous aspects of successful symbol performance and methodologies for evaluation of symbols. Research in five areas of application is discussed: highway, automotive and machinery, public information, product, and building safety. The bulk of this research is presented in terms of the understandability of the symbol, a key factor in the selection and use of symbols for effective communication systems. Research into other psychological processes such as detection, discrimination, recognition, and behavior is also discussed, however.

Finally, the urgent need for further research, particularly in the area of symbols to provide safety and hazard warning information in and around buildings, is discussed. The need to develop a comprehensive model for evaluation of symbol performance in terms of the relevant psychological processing dimensions is noted.

Although the earliest written languages of Egypt, China, and Mesopotamia used pictures to represent ideas, the modern alphabet has evolved far beyond the simple graphic representation of familiar objects (Giedion, 1966). Because the alphabet now bears little resemblance to a series of pictures, the meaning of words constructed with alphabetic characters must be learned. As a result, written representations of objects or ideas are only meaningful to those who can read a particular language. In many instances, however, there is a need to communicate information to all people who use a building, transportation system, highway, machine, automobile, or the like.

The result has been a trend back to the use of pictures, or graphic symbols, to portray information as an integral element of environmental signs and visual communication systems. These symbolic signs are designed to communicate information rapidly and accurately without reliance upon a specific language. Because graphic symbols are intended to communicate visually, they are effective only when both the symbol's creator sends and the user receives the same message.

In the following pages, various concerns which surround the development of effective symbols are discussed. Section 1 defines symbol terminology and discusses the historical development and standardization of symbols. In section 2 the advantages and disadvantages to symbol use are reviewed with suggestions for good symbol design and graphic improvement. Section 3 reviews the experimental evaluation of symbols for five applications: highway, automotive/machinery, public information, product hazard, and building safety. Finally, in section 4 suggestions for further experimental evaluation are presented along with issues in the development and use of more effective symbols.

1.2 DEFINITION OF GRAPHIC SYMBOLS

The term graphic "symbol" has been defined in numerous different ways. Modley (1976) broadly defined a true graphic symbol as a comparatively simple geometric shape which uniquely represents a concept, while Dreyfuss (1972) described it as a written mark or character which is used to represent something. In the following pages the term "symbol" will be used as Modley defined it - a shape which represents a concept or "referent". Although the term "symbol" is sometimes restricted to an "abstract or geometric form which is associated with an idea" (Follis & Hammer, 1979, p. 59) while the term "pictogram" refers more to pictures of readily recognizable objects, no such distinction will be made in the following pages.

There are, nevertheless, several categories of symbols. Modley (1966) and Dreyfuss (1972) defined three categories - pictorial, concept-related, and arbitrary. (See figure 1 which provides examples of pictorial, concept-related, and arbitrary symbols.)

According to both Eliot and Modley, pictorial symbols are those which directly represent an object or feature and which are expected to convey their meaning without further explanation. They can be easily recognized because of their resemblance to a familiar object. Problems with this kind of symbol arise only when the represented object changes or becomes unfamiliar to the audience (Eliot, 1960). Modley (1966, p. 119) claimed "that there are only a few image-related symbols which are unique in meaning, clearly recognizable, and permanent in time ... Many other image-related symbols will have only limited use — geographically and historically. As such, they should be used with caution so as to avoid possible faulty interpretation by 'outsiders'."

Abstract concept-related symbols, however, depend upon an arbitrary code of meaning which is artificially created and accepted. Thus, concept-related abstract symbols refer to perceptual concepts rather than real objects. For example, a directional arrow or a horizontal wavy line can represent the idea of "turn here" or "water" quite effectively. Furthermore, once the basic perceptual concepts have been symbolized, the concept-related symbol is less likely to change, because the underlying referent should remain constant. Modley noted that although a concept-related symbol is not necessarily instantly recognized, it may be easy to learn because the graphic form depicts the perceptual content of the represented concept. As a result, it is also easily remembered.

Arbitrary abstract graphic symbols, on the other hand, "do not resemble real objects nor are they related to the objects or concepts which they represent. Their shape is arbitrarily assigned to them. Because of this lack of reference, these symbols are more difficult to teach, more difficult to learn, and harder

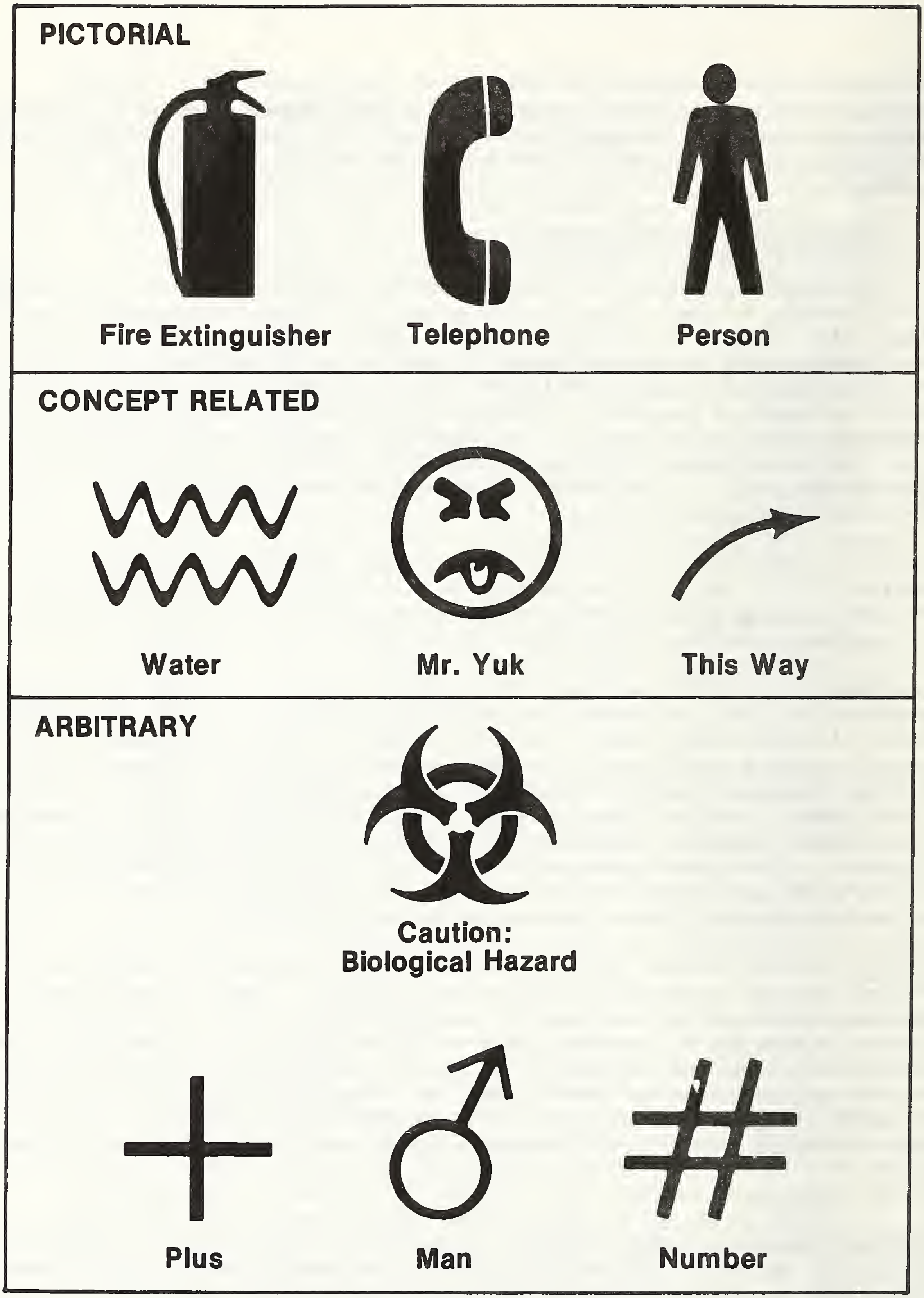


Figure 1. Categories of symbols

to retain. They have to be learned by rote . . . Regardless of their possible origin as image or concept-related symbols, the graphic symbols in current general use - letters, numerals, punctuation marks, and mathematical operators - are all arbitrary symbols. Only in the more recent symbol systems used by professional groups or proposed for public use do we find heavier reliance on image-related or concept-related symbols" (Modley, 1966, p. 120). For the most part, abstract arbitrary symbols are not covered in detail in this report, since the focus is upon effective and meaningful symbols for building signals.

In the following pages the term "symbol" will include the pictorial or abstract graphic representation of objects, concepts, and events. The term "referent" will include the symbol meaning or concept. No differentiation will be made between the terms symbol, pictogram, glyph, pictorial, or image. The discussion will, however, center upon the use of symbols as elements of visual communication and alerting systems.

1.3 SYMBOL HISTORY AND STANDARDIZATION

Modley (1976) suggested that the modern use of symbols such as the barber's pole or the apothecary's mortar and pestle began in medieval times to indicate services and to provide directional information. Through repeated use, these symbols became recognized as the representation of all similar activities rather than the sign for an individual shop or tradesman. In the 20th century, the most prevalent use of symbols has been for highway signs due to the rapid development of the automobile and the growing ease of travel between foreign countries. Eliot (1960, p. 18) commented that "without the motor vehicle, highway signs might well have remained primitive, local, and highly individualistic. The motor vehicle that tremendously expanded the range of travel and brought an era of individual travel for the masses also created new hazards and a need for vastly improved guidance for the strangers who were following new highways into distant places." Eliot suggested that one of the first symbols was the pointing finger, which evolved into an arrow meaning "go this way", and now is one of the most common graphic symbols.

The use of symbol signs has expanded widely since World War II as a result of increased international travel and trade. Symbols now exist for applications ranging from machine tools and automobiles to hazardous materials warnings and buildings. They are used to provide direction, to identify objects and controls, and to warn of hazards.

The rapid expansion of symbols into a multitude of applications has often resulted in the use of different symbols for similar messages. Yet, if graphic symbols are to be effective, their form should be similar, or standardized, for each message for all audiences. One of the first attempts at standardizing symbols for road use occurred at a convention in Paris in 1909 on the International Circulation of Motor Vehicles. These standard symbols were designed to eliminate language barriers and communicate their intended message easily (King, 1971). At this time four symbols were adopted - for curve, bump, intersection, and railroad grade crossing (see figure 2). In 1916 another convention on motor traffic added symbols for linked curves and uneven pavements

ROUGH ROAD

CURVE

LEVEL CROSSING - RR

CROSS ROADS

Figure 2. First standardized symbols. Highway symbols, 1909

while approving a triangle as the shape for warning signs. The League of Nations in 1931 expanded this series to 26 symbols with uniform requirements for color and shape. Finally, the United Nations adopted a protocol on Road Signs and Symbols in 1949, which relied upon the earlier work on symbols, and which is now in widespread use throughout Europe.

In contrast, traffic signs in the United States (U.S.) depended primarily upon English words to convey the intended message until the early 1970's. This practice was standardized in 1925 by the American Association of State Highway Officials. Elliot (1960, p. 22) commented that standardization of word signs was "a byproduct of a program set up to designate and mark an official network of major highways and thus to end an uncontrolled, and often overlapping and competing marking of tourist routes by private promoters." In 1966 the Highway Safety Act standardized signs throughout the county, and by 1971, the U.S. practice was revised to include European style symbols in addition to English words based upon the U.N. protocol (see figure 3). Even so, highway sign practices are not fully standardized throughout the world (or even throughout Europe and North America). Nevertheless, there is the beginning of a viable system of highway information signs using a mixture of symbols and words for motorists. Furthermore, the success, however limited, of implementing a system of symbolic messages for highways has sparked the development of symbol systems for other applications.

Margaret Mead, Rudolf Modley, and Henry Dreyfuss were pioneers in bringing the need for symbols in all applications to widespread attention. Mead and Modley (1968) noted that in 1968 people spoke at least 2800 languages and could travel to any part of the earth in no more than 36 hours, a time which today's supersonic transport has most assuredly shortened. The speed of modern travel demands unambiguous signs understood by those who speak any language from any culture. Mead termed such symbol signs "glyphs." In her view, glyphs would form the basis of an international communication system. "What is needed internationally is a set of glyphs which does not refer to any single phonological system or to any specific cultural system of images but will, instead, form a system of visual signs with universally recognized referents" (Modley, 1966, p. 113).

Dreyfuss (1972, p. 18) envisioned a set of universally understood symbols as a "supplement to <u>all</u> languages to help create a better and faster understanding in specific areas. Symbols have already evolved to the point of universal acceptance in such areas as music, mathematics, and many branches of science." Beyond these specialized symbols, however, the only symbols which currently have a claim to international standardization are those for highways and hazardous materials transport. Yet, because of the immense popularity of graphic symbols, they are currently used in numerous applications from hazard warnings to Olympic games. New symbols are often developed without consideration for existing symbols or standards. Modley and Myers (1976) and Dreyfuss (1972), for example, published extensive compilations of numerous symbols for particular referents. Modley and Myers present examples of 19 symbols for first aid, 24 symbols for telephone, and 17 symbols for no entry. The compilation by Dreyfuss

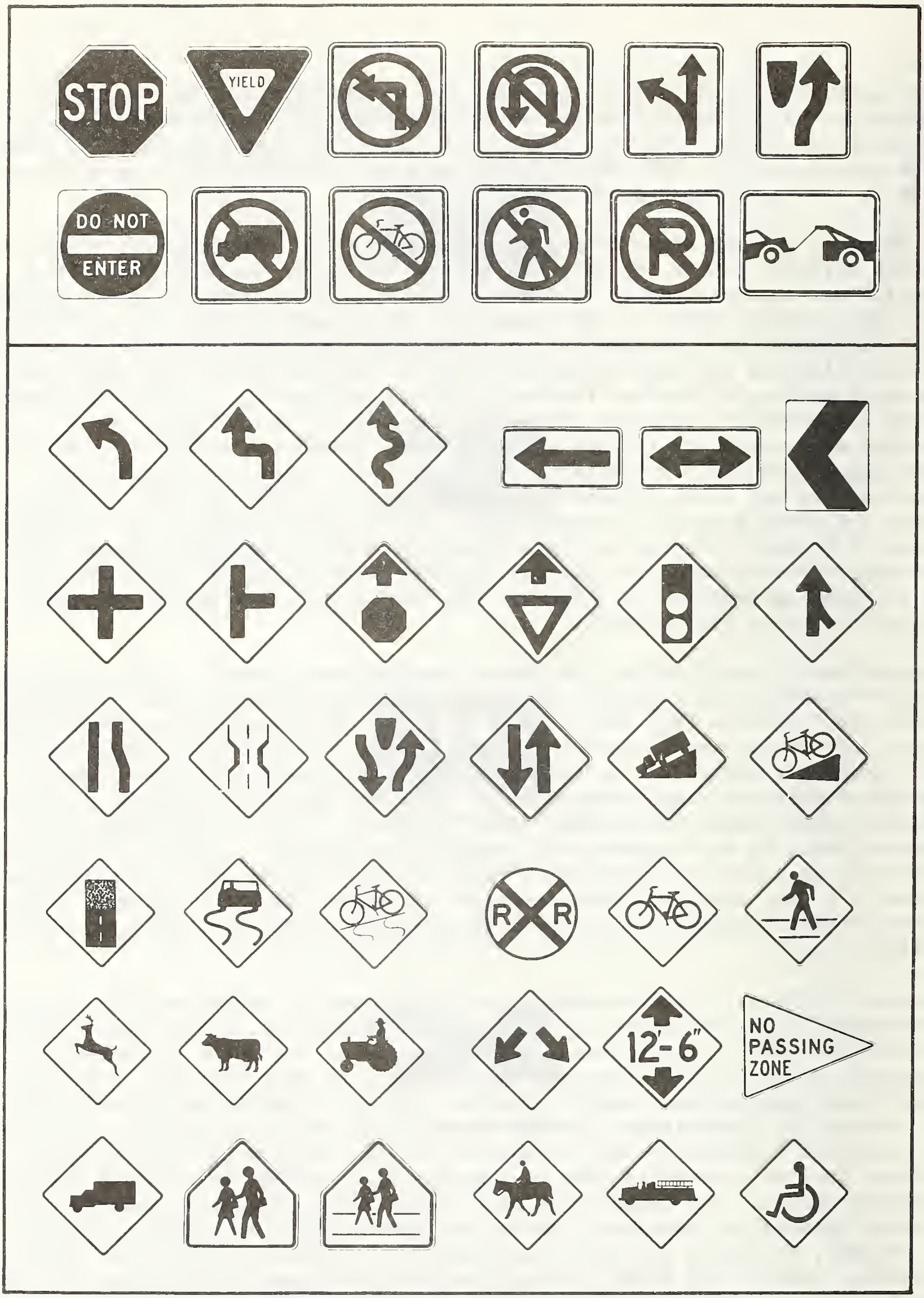


Figure 3. Standard U.S. highway symbols

includes symbols representing geography to safety, and adds form and color as well. These compilations present only a fraction of the numerous symbols now used for many referents. The result of the proliferation of symbols and the lack of standardization "has been much experimentation, some progress, and a great deal of confusion" (Mead & Modley, 1968 p. 58). As an example, figure 4 depicts the confusing variety of symbols proposed and used for ear protection. Reduction of the confusion surrounding symbols, and the concomitant development of effective symbol signs, requires systematic research and eventual standardization.

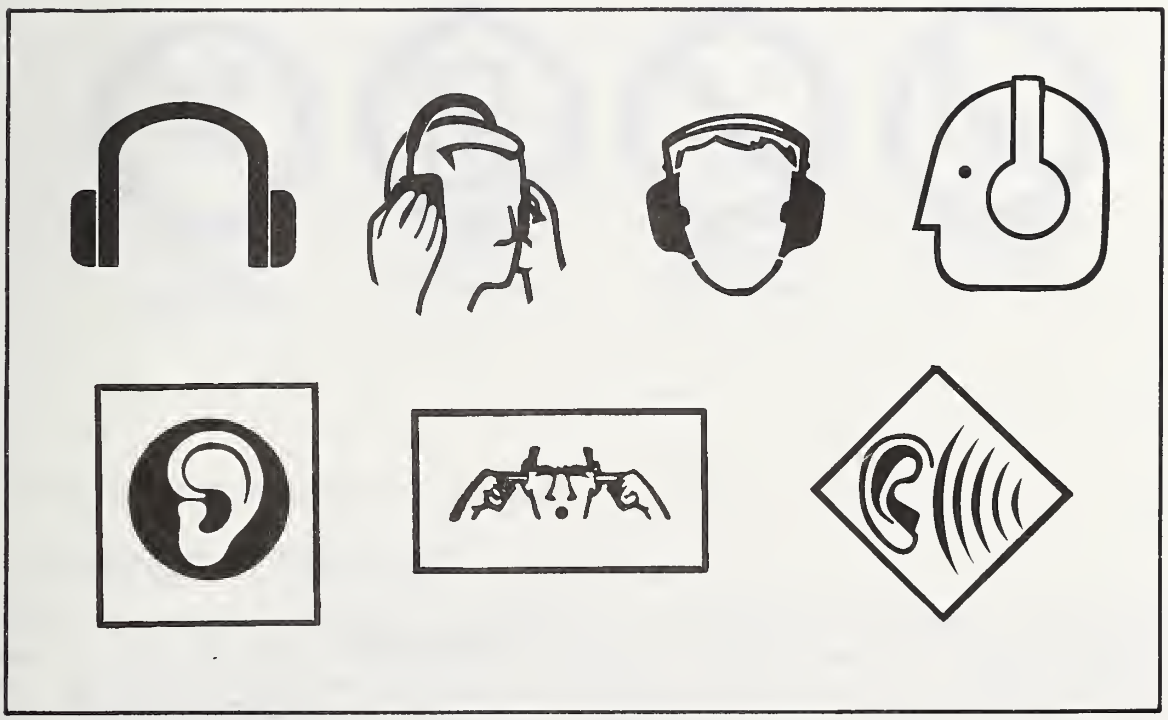


Figure 4. A sample of the variety of symbols proposed for ear protection

Facing page:

Symbols intended for fire safety information


Fire Extinguisher (White on Red)


Hose and Reel (White on Red)


Fire Ladder (White on Red)


Fire Bucket (White on Red)


Fire Fighter's Equipment (White on Red)


Direction to Equipment (White on Red)


Break Glass for Access (White on Green)


Slide Door to Right (White on Green)


Do Not Use Water To Extinguish (Black on White, Red Circle & Slash)


Do Not Lock (Black on White, Red Circle & Slash)


No Smoking (Black on White, Red Circle & Slash)


No Open Flame (Black on White, Red Circle & Slash)

2. DEVELOPMENT OF EFFECTIVE SYMBOLS

2.1 ADVANTAGES OF SYMBOL USE

Estimates of adult illiteracy in the U.S. vary widely from about 2 million to about 40 million (Kirsh & Guthrie, 1977-78). Still another estimate suggests that "as many as half the adults in the United States may be semiliterate and unable to understand driving manuals, newspapers, and job applications, and possibly direction signing. But even normally literate persons have been found to more readily recognize and retain shorter, more familiar terms and phrases" (Cantilli & Fruin, 1972, p. 236). Anxiety may also reduce capability for visual recognition, recall, and retention. In addition, English is not the native language for about 8 million Americans. As a result, the need for sign systems which are not totally language dependent is great, particularly for safety information and hazard warnings. Symbols can, when chosen appropriately, meet this need.

In addition to communicating apart from a specific language, symbols stimulate visual thinking. Modley (1966) claimed that a symbol portraying a right curve can be more effective than a sign presenting the words 'Right Curve.' Because this symbol inspires visual thinking, a driver sees the curve ahead and will adjust to it automatically. With a word sign, he must first interpret its meaning before he can act upon it (Modley, 1966). Modley (1976, p. 60) claimed that "uniqueness and clarity of meaning, independence from language and cultural differences, and visual directness are the ideal objectives which should underlie the creation of any new symbol system."

Mead and Modley (1968) suggested that simplicity, unambiguity, and understandability are the major advantages of symbols or glyphs. Furthermore, they "create a direct and immediate impact and this permits immediate response. This applies as well to those who know a language as to those who do not. This immediacy of response saves thinking" (Mead & Modley, 1968, p. 57). Mead even anticipated that carefully chosen glyphs could supersede the need for written language.

Another advantage of symbols is the possible use of a portion of an object to represent the whole object (Kolers, 1969). This critical characteristic of successful pictorial representation has the advantage that it can occupy less space on a sign or label. "For example, a drawing of a telephone hand set can convey information about the location of telephone facilities more compactly than words can, and a drawing of a curve on a road-sign can signify the nature of the road more compactly than the word 'curve.' The spatial compactness is achieved, however, through greater reliance on the reader's memory and interpretive powers" (Kolers, 1969, p. 355). The interpretability of pictorial symbols can thus be enhanced by portraying them so that they draw upon the user's previous experience.

2.2 LIMITATIONS TO SYMBOLS

Although Mead firmly believed that glyphs could be the cornerstone of future international communication, others, such as Kolers (1969) and Modley (1976), disagreed. These authors claimed that although written language began as a series of pictures, most languages have developed alphabets in which characters represent sounds rather than ideas. Pictorial writing is successful only when the audience is familiar with the object depicted, and when that object or idea can be portrayed as a picture. Furthermore, written languages tend to shorten repeated items and reduce the amount of detail - a process that could wreak havoc with a pictographic communication system (Kolers, 1969). Although Mead claimed that pictograms are directly and immediately understood, Kolers pointed out that "Recognizing even a realistic pictorial representation of an object requires a great deal of perceptual learning, abstracting ability, and intelligence" (1969, p. 355). Furthermore, Modley (1976) noted that abstract symbols impose particularly heavy demands, requiring people to learn and remember unfamiliar pictures.

In addition to the audience's need to learn unfamiliar and abstract symbols, as well as to be familiar with the object depicted, the effectiveness of symbols is limited by the lack of syntactical rules for combining ideas pictographically. Although words and sentences are formed from combinations of alphabetic elements, it is nearly impossible to follow a similar process with symbols. There are no rules for constructing combinations of pictograms from individual elements, nor is there necessarily any clue to the meaning of the final compound (Kolers, 1969; Drake, Moblin & Shaw, 1979). Nevertheless, combined symbols have been made successfully by using slashes for prohibition and selecting background shapes such as triangles, circles, and squares to encode particular meanings (see figure 5). Unlike true compounds, however, these combinations do not have meanings unique to the combined elements.

Drake, Moblin, and Shaw (1979) contend for example, that much of the current uncertainty about symbols involves the development of syntax and grammar for symbols. "Most symbol systems seem to balance uneasily between trying to be a collection of emotive pictograms that are immediately recognized and understood, and accepting the necessity of a language logic that has to be learned" (Drake, et al., 1979, p. 58). Once isolated pictures (symbols) are developed, the next challenge is to develop the rules and logic of a grammar for relating or combining such symbols (Drake, et al., 1979).

An example of the difficulties which arise with compound symbols is the depiction of prohibition (Modley, 1966). Typically, it is depicted with a combined symbol - one showing prohibition, the other showing the forbidden item or action. Yet, the forbidden item is often overlooked or misunderstood. Modley (1966) suggested that a designer would be better advised to devise a positive symbol which could be simpler visually. The positive symbol would inform the audience of the correct action instead of telling it what <u>not</u> to do without providing guidance for the correct action to substitute. The FMC Corporation (1978), for example, uses symbols to depict the result of the hazard, such as losing fingers or limbs in gears or falling into machinery, to avoid the problem of misunderstood prohibitory symbols (see figure 6). For some referents, however, such as "no smoking," it is difficult to avoid the need for a prohibition symbol.

In addition to presenting syntactical problems, symbols are also not well-suited to commenting upon or evaluating information. These limitations also imply that, "contrary to the assertion of some of its proponents (Mead, 1965; Mead & Modley, 1968), pictograms can never fully replace alphabetic languages nor are they a major road to international communication. Restricted to identifying and locating objects and to conveying instructions, they can fill a most useful role, however" (Kolers, 1969, p. 360).

	ISO/EEC	CANADIAN	U.S.
Red Shape	Prohibition 	Prohibition Danger 	• Danger • Stop • Fire & Emergency
Orange	— —	— —	• Dangerous Machine • Energized Equipment • Warning - Proposed
Yellow	Warning 	Caution 	• Caution - Storage for Flammables - Containers for Explosives, or Unstable Materials -Radiation
Green	Information 	Emergency Information 	• Safety Information • First Aid & Safety Equip.
Blue	Mandatory Action 	Miscellaneous	• Information • Bulletin Boards • Railroad
Black		Mandatory Action 	

Figure 5. Use of shape and color to encode meaning

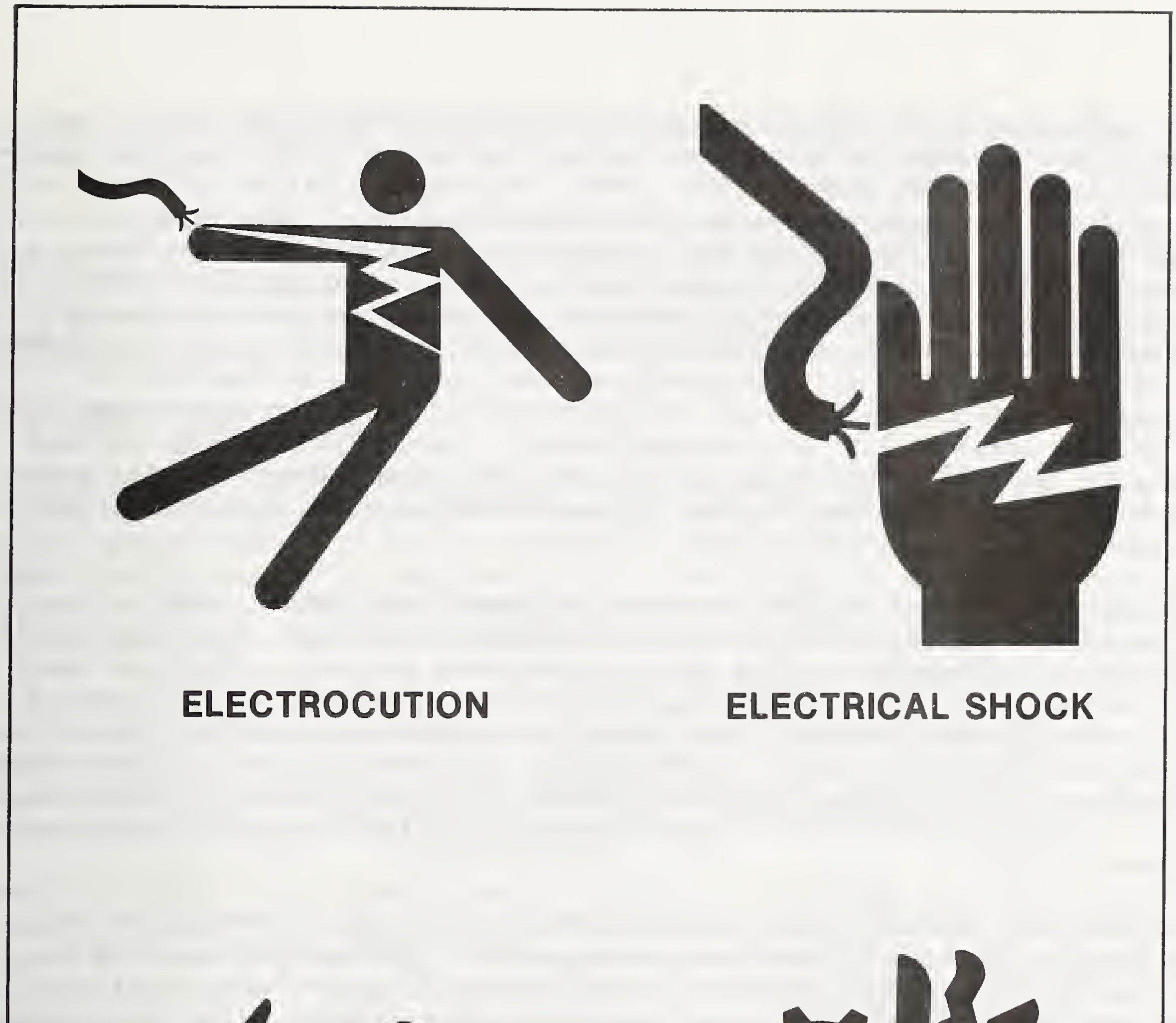


Figure 6. Symbols which depict result of hazard

Symbols are also limited in their ability to convey processes or activities such as ticket purchase or hotel reservations, according to the American Institute of Graphic Artists (AIGA) (1974, 1979). Furthermore, it is difficult to use symbols to represent temporary or changing situations. Thus parking restrictions which change by the day, such as alternate-side-of-the-street parking, are difficult to convey symbolically. Still other problems arise because of graphic limitations upon symbols. A major concern is the use of two-dimensional symbols to represent three-dimensional objects. Modley (1966) noted that it is difficult for a two-dimensional road sign to represent a horizontal direction. As a result, an arrow pointing straight up has come to mean "proceed horizontally" or "straight ahead." Yet, this same sign is used to indicate "up" and "down" on elevators and escalators. These changing meanings can cause real problems for the intended audience which must adjust its response according to the situation.

Few symbols can express complete messages by themselves, (Eliot, 1960). The required additional information is typically added by context, previous experience, words, or redundant cueing such as sign shape or color. In fact, context is often critical in interpreting the meaning of a symbol. Thus, even a directly pictographic rendition of a cow will not make sense to the viewers unless it is placed by a country roadway where it presumably means cattle crossing (Modley, 1976). Machine operating symbols typically need to be placed on or near a control before they will accurately convey the intended information (Dreyfuss, 1966).

The audience may introduce other limitations such as lack of familiarity with the underlying concept. Those who are not familiar with machine operations may not understand certain operating symbols which do communicate effectively to the intended, skilled audience. In addition, the audience directly addressed by the symbol may be only a small portion of the total population who actually views the symbol (Modley, 1960). Traffic symbols, for example, only address drivers; yet, they are placed in full view of pedestrians so that a red light means "stop" for a driver, but "walk" for a pedestrian. Thus, people must learn when a symbol is applicable to them and when it is not.

Finally, the symbol may lose meaningfulness due to technological advances, cultural differences or local associations. Thus a steam engine may not represent the modern diesel locomotive, nor will a man in evening dress represent pedestrian crossing to all people (Modley, 1966; Eliot, 1960). A pictogram of an elevator may mean nothing to someone who has had no experience with elevators, while Mead noted that not everyone will associate a skull and cross bones with poison or danger (Follis & Hammer, 1979). Furthermore, local associations may alter the effectiveness of symbols. Eliot mentioned that although the upraised hand of a policeman was suggested as a symbol to represent "STOP," "it seems, however, that in Paris, among taxicab drivers, the same symbol, personally performed, conveys an insult somewhat equivalent to a thumbing of the nose. A 'stop' sign with such an implication might command slight respect in Paris, regardless of its effectiveness elsewhere" (Eliot, 1960, p. 20). (See figure 7, which depicts symbols which have become outmoded or regionally unacceptable.)

Even when symbol use is restricted to more limited applications, such as providing directional information or hazard warnings, there are limitations to the kinds of information that symbols can convey effectively. Consideration of some general constraints can facilitate the design of better symbols, however, and improve visual communication despite the limitations of symbolic imagery.

2.3 DESIGN CONSIDERATIONS

Although all of the advantages and limitations mentioned in 2.1 and 2.2 should be addressed during the design of a symbol, they do not alter the ability of well-designed symbols to convey information rapidly and effectively. Many of these limitations can be overcome through careful attention to the symbol design itself and through thoughtful assessment of the message content, beginning at the design phase. While any evaluation of symbol effectiveness should include a final assessment with the target population, initial evaluation at the design stage will permit the development of the most effective symbols. Such thorough evaluation of symbols is needed for those symbols designed to replace or supplement word signs intended for mass consumption. The use of symbols as corporate identification, wall graphics, or product embellishment does not, however, demand such complete evaluation (Modley, 1976).

Ideally, symbol effectiveness should be evaluated from the twin viewpoints of graphic integrity and conceptual meaningfulness. Both kinds of evaluation should proceed hand-in-hand, with feedback from each used to improve the rendition of the symbol. Typically, however, the two types of evaluation occur at separate stages by different investigators, with little feedback or interaction between them.

Some of the suggested considerations for design review will be addressed in this section. A detailed review of the research on the evaluation of meaningfulness for different applications will be presented in the next section.

2.3.1 Graphic Concerns

Modley (1976) suggested that the most common graphic reasons for symbol failure are: conceptual failure (poor expression of the object or idea); poor design or draftsmanship; conflicting meanings (several symbols for one idea or one symbol for several ideas); poor use of color; inappropriate or excessive use of the symbol; and failure to use background shape in a consistent fashion. Avoidance of such failures can be achieved through design evaluation, beginning with a subjective review of each symbol's graphic content.

Doblin (1979) and Whitney (1979) discussed the evaluation of the graphic effectiveness of visual messages as an integral part of the design process. To aid the designer, Doblin (1979) developed a twelve-stage model for determining the effectiveness of a graphic message, either before or after implementation.

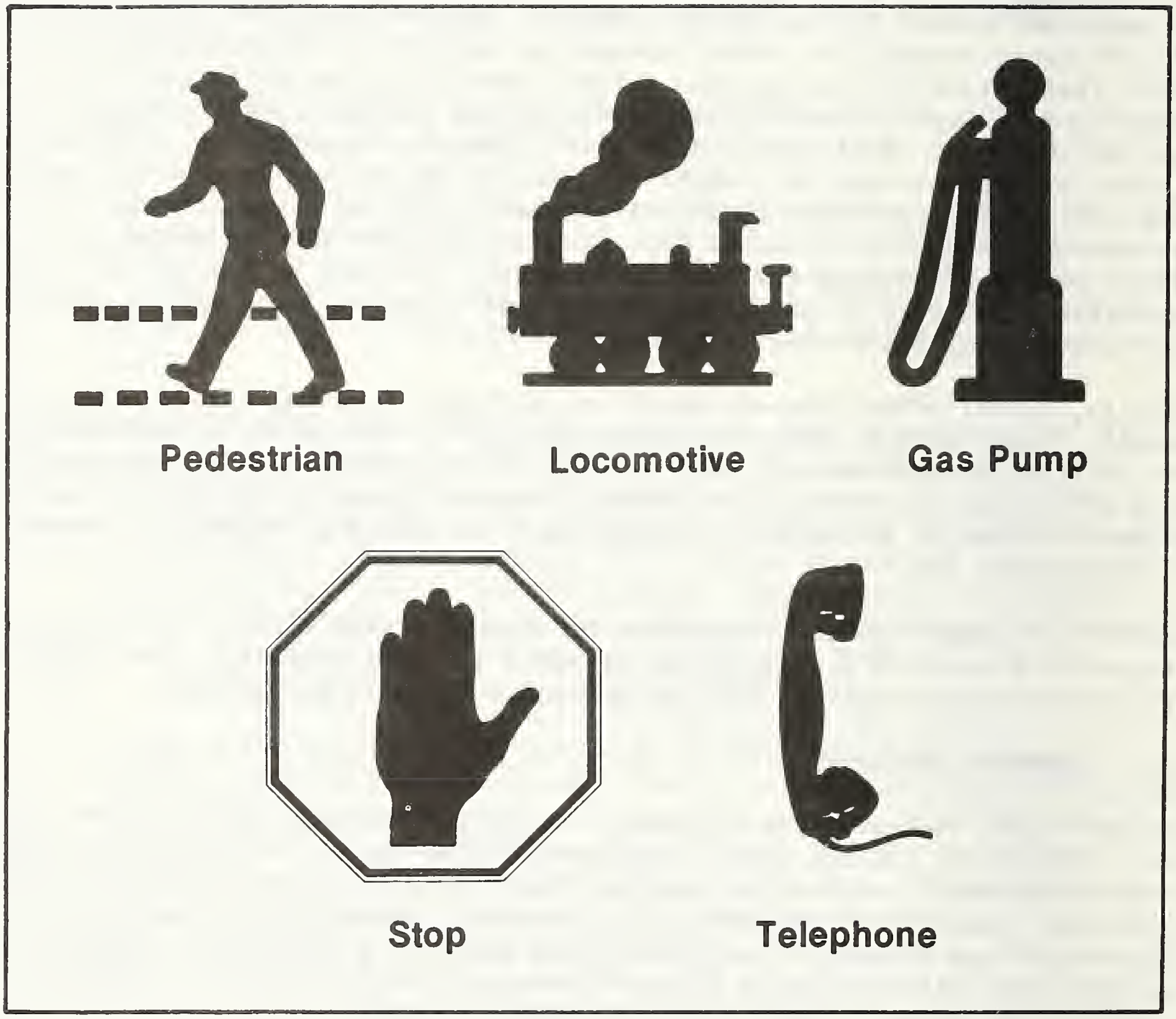


Figure 7. Symbols which have become outmoded or regionally unacceptable

To be effective, the message must reach its audience, be powerful enough to gain attention, and unique enough to compete successfully with other messages. The time required to decode the message should be carefully selected to correspond to the audience's needs. Lengthy messages can become so demanding or boring that people will not take the time needed to understand the meaning. "This is why 'stop' is used to signal motorists, rather than a sign saying, 'Chicago Statute 3413R requires all motorists to stop at designated intersections. This is one such intersection. Violators are subject to a $15 fine. Repeated offenders can lose their licenses'" (Doblin, 1979, p. 42). The message should be easily seen, in a language that is understood, and its form should be congruent with its content. The message should prepare the audience so that it will understand the meaning of any symbols used. It must also be sufficiently exciting that its audience is not bored. A version of Doblin's model is presented in Table 1.

Table 1 - Qualities of Effective Graphic Messages from Doblin (1979)

1. Message must reach appropriate user
2. Message must be powerful - must compete successfully with other messages
3. Message must observe time restrictions
4. Message must be perceptible
5. Message must be decodable (understandable)
6. Message must have proper hierarchy - emphasize essential elements
7. Message must prepare user
8. Message must have integrity
9. Message must be properly crafted - of good quality
10. Message must be credible - believable
11. Message must have congruity
12. Message must have excitement

Although Doblin favored evaluation of graphic designs, he noted that designers may be particularly threatened by criticism from outsiders. Furthermore, it can be difficult for outsiders to place a value on the visual design in and of itself. Whitney emphasized, however, that design evaluation is a tool for determining if the finished product is successful and for providing feedback for any necessary improvements. He suggested that a particular design can be evaluated in two ways: mock-ups and similar methods, to predict the probable success of several alternative designs; and post-design evaluation, to provide information about the reasons for a particular design's success or failure. The goal is to ensure that the design successfully communicates the desired information.

Geismar, Chairman of the AIGA, commenting upon the many difficulties facing those who design symbols for the public, noted that "For the designer, the development of a successful pictorial symbol system is an extremely difficult task. Unlike the eighteenth-century craftsman, who had only to be concerned with a single representation, today's designer of a symbol system is concerned with giving a unified graphic style to a diverse group of images" ... including both pictorial and abstract symbols. "But when faced with the problem of dealing with a mixture of pictograms and abstractions, the designer must struggle to maintain clear communication and, at the same time, establish some visual consistency" (in Herdeg, 1978, p. 18).

2.3.2 Application of Perceptual Principles to Symbol Design

Other tools for improving graphic design include knowledge of the psychological principles of perception. For example, Easterby (1967, 1970) discussed the application of the Gestalt principles of perception to symbol design. He noted that while size, brightness, and contrast are physical parameters which affect the visibility of a display, complete perception (and subsequent understanding) of a symbol rests heavily upon its discriminability and meaningfulness. Easterby defined discriminability as the ability to tell one symbol from another, and meaningfulness as the ability to define the concept symbolized. Application of the Gestalt principles of perception to the design of a symbol can improve its discriminability and even its meaningfulness (Easterby, 1970).

A major Gestalt principle is the relationship between the figure and its background (Easterby, 1967, 1970; Follis & Hammer, 1979). The figure is determined by a contour which may be a line outline or a contrast boundary. (With a contrast boundary, there is a sharp change in contrast between the figure and the ground). Easterby (1970) claimed that for effective symbol design the contrast boundary is superior to the line boundary — perhaps because there is an inside and an outside figure for a line form, while there is only an inside figure for a contrast boundary.

Whichever boundary is used, its sharpness and clarity will strongly affect the ease of perception and recognition of the resulting figure (Follis & Hammer, 1979). The stability of the relationship between figure and ground is critical. The tendency for an observer to complete a broken contour (closure,) and to group like elements, should not be overlooked, either (Morgan, 1961). Other design characteristics which affect perception include: continuity, contour, simplicity, symmetry, unity, shape, form, and size. Finally, the visual integration or overall coherence of a symbol will influence its perception (Easterby, 1970).

Eliot (1960) noted that a crucial requirement for good symbol design is the limitation of the amount of fine detail. Ideally, a good design should have a bold outline, a minimum of fine detail, and be distinctly different from other symbols within a set. The symbol should be simple with a smooth continuous outline. If asymmetry or discontinuity is used, it should be limited to those cases where it is needed to add meaning. Finally, symbols within a set should be unified through consistent use of the same sizes and proportions of individual elements (Geismar, 1978). Application of these perceptual principles to symbol design can improve symbol effectiveness.

2.4 CASE STUDIES OF DESIGN EVALUATION

Several assessments of symbol sets performed during the design phase will be discussed in this section. In such an assessment the designer reviews the graphic content of a symbol and may make a preliminary estimate of its probable meaningfulness or understandability for the intended user. Generally, however, the designer does not conduct a large-scale evaluation of the symbol's effectiveness with the actual target audience.

Dreyfuss (1966) and Easterby (1967) each reported the development of sets of symbols for industrial equipment and machinery. The American Institute of Graphic Arts (AIGA) (1974, 1979) developed two sets of public information symbols for use in transportation facilities. In each graphic assessment, a listing of essential messages (or referents) to be symbolized was among the first steps.

Following a compilation of existing symbols for each essential message, Dreyfuss (1966) reviewed the graphic content of existing symbols for machine operation. He also drew a preliminary design for each idea and reviewed existing symbols for their graphic content and relationship to his design and to the intended message. During this review, Dreyfuss decided which existing symbols were acceptable, and which ones needed to be redesigned. He commented that future new symbols should be reviewed to determine that they do not have erroneous or undesirable associations and that they are not so similar to existing symbols that they will conflict with them. Dreyfuss then refined the set of symbols by considering the consistency among symbols for objective and nonobjective elements, the operator's probable background and experience, the expected ease of identification, and the frequency of use.

A major concern was the degree of abstraction of the symbol as a function of the machine control to be symbolized. Dreyfuss claimed that abstract symbols, traditionally more difficult to understand, could be used for frequently used controls because these symbols would be learned through constant exposure. Less frequently used controls require more pictographic symbols for which less learning is needed. If the symbol is located on or near the control, however, Dreyfuss suggested that it could be more abstract, since the control itself will provide important information.

In the final step of the graphic review, Dreyfuss evaluated the graphic form of each symbol developed. He commented that "the graphic form of a symbol is almost as important as its content. This form is determined by the graphic design elements, techniques of reproduction, and conditions of application. Graphic considerations include: simplicity and clarity of design; consistent weight and relation of line; use of light and dark areas; general proportions; and legibility at reduced sizes" (Dreyfuss, 1966, p. 133). Dreyfuss concluded that evaluation of the graphic content of symbols is only the beginning. Each symbol must then be tested in field applications.

In a similar effort, Easterby (1967) discussed the development of a set of symbols for machine tool operations. He claimed that there are three sets of principles of language construction which are relevant to display or sign design. The first, pragmatic rules, pertains to the relationship between the display and the particular content. The second, semantic rules, refers to the need to agree upon the referent, object, action or idea for each sign. The third, syntactic rules, refers to the need to determine relationships for combining signs into a coherent language. Easterby applied these principles to the design of a set of machine tool symbols for which he developed a semantic structure with a basic classification of elements, directions, and functions. Other concepts for combining symbols syntactically included direction, rate, and maneuver. In Easterby's system, the various operations that a machine or an operator might perform and that could require symbolism were identified - and then classified as symbolic adjectives (directions), symbolic verbs (functions), or symbolic nouns (elements). Easterby noted that consideration of each of these structural items is important in determining the meaningfulness of a display to the eventual user.

For a very different application, namely the use of symbols for public information, the AIGA (1974; 1979) formulated a number of explicit considerations for determining the graphic effectiveness of symbols. Their prime consideration was to use symbols only for truly essential messages. They stated that "we are convinced that is is more harmful to oversign than to undersign. To mix messages about relatively insignificant activities and concessions with essential public messages weakens the communication. While there may be some messages beyond this basic group that require symbols, only those messages that are truly essential should be considered" (AIGA, 1974, p. 6). The AIGA asserted that symbols should be used to represent objects and concrete entities, rather than processes or activities. For example, it is much easier to develop a successful symbol to depict bar or coffee shop than to develop a symbol for ticket purchase.

During the development of symbols for public information, the AIGA defined four categories of message areas: public services, concessions, processing activities, and regulations. Symbols from 24 sources were grouped into each of these areas depending upon the concept symbolized. Then, committee members rated each symbol for a concept using syntactical, semantic, and pragmatic dimensions. Although the three dimensions are similar to those discussed by Easterby (1967), the AIGA defined each one in greater detail. Their discussion follows:

By the "semantic" dimension the AIGA meant the relationship of a visual image to a particular meaning. They suggested the following questions:

> "How well does this symbol represent the message?
> Do people fail to understand the message that the symbol denotes?
> Do people from various cultures misunderstand this symbol?
> Do people of various ages fail to understand this symbol?
> Is it difficult to learn this symbol?
> Has this symbol already been widely accepted?
> Does this symbol contain elements that are unrelated to the message?"

By the syntactic dimension, the AIGA referred to the relationship between two visual images. They raised the following questions.

> "How does this symbol look?
> How well do the parts of this symbol relate to each other?
> How well does this symbol relate to other symbols?
> Is the construction of this symbol consistent in its use of figure/ground, solid/outline, overlapping, transparency, orientation, format, scale, color and texture?
> Does this symbol use a hierarchy of recognition?
> Are the most important elements recognized first?
> Does this symbol seriously contradict existing standards or conventions?
> Is this symbol, and its elements, capable of systematic application for a variety of interrelated concepts?"

By the pragmatic dimension, the AIGA meant the relationship between the user and the visual image.

> "Can a person see the sign?"
> Is this symbol seriously affected by poor lighting conditions, oblique viewing angles, and other 'visual noise'?
> Does this symbol remain visible throughout the range of typical viewing distances?
> Is this symbol especially vulnerable to vandalism?
> Is this symbol difficult to reproduce? Can this symbol be enlarged and reduced successfully?"

Five members of the AIGA rated each of the proposed symbols on a five point rating scale for each dimension. Based upon these ratings, a single image was developed for each message. Once the symbol images had been selected, the AIGA used the following considerations for establishing a uniform graphic vocabulary. These included: simplification, (in which unimportant details were eliminated to make the symbol bold and direct,) consistent line elements, rounded curves to create distinctive visual elements, symmetry, and uniform background. The AIGA used these considerations in the development of an initial set of 34 images (1974) followed by a subsequent set of 16 images (1979). See figure 8 for a presentation of both sets of DoT public information symbols.

The case studies discussed in this section demonstrate the kinds of criteria that are used in evaluating the graphic content and effectiveness of symbols. Yet each study stops short of evaluating the meaningfulness for the intended audience of the proposed symbols, leaving that task to others. In the case of some of the AIGA public information symbols, however, the Franklin Research Institute has evaluated their effectiveness at communicating the desired messages to the general public. This research is discussed in section 3.4.

Figure 8. DoT public information symbols

2.5 NEED FOR EVALUATION OF EFFECTIVENESS

Although Easterby (1970) advocated the careful application of psychological principles to symbol design, he disagreed with the idea that all symbols should be understandable at first glance. He claimed further that accuracy and speed of response after training are more suitable indices of symbol effectiveness. In testing symbols, the size of the set itself can affect recognition, with large sets of 100 or more symbols being much more difficult than small sets of 10 or so. Nevertheless, Easterby asserted that laboratory studies can improve symbol effectiveness by assessing the discriminability of a symbol through comparison with other similar symbols, and can determine whether a particular shape will adequately convey an intended meaning. Easterby (1970, p. 151) concluded that "An intuitive design approach guesses at a good symbol while an experimental approach, which we advocate here, attempts to define a symbol stereotype and then, further, to validate this stereotype by experiment. To achieve the optimum form of that particular shape we may then proceed to manipulate this shape using perceptual notions from the theory of pattern perception and discrimination."

Although each of the studies reviewed in section 2.4 discussed the effectiveness of various symbol designs, in no case were these designs assessed for their specific effectiveness with the potential user group. Consequently, no feedback from the user could be considered for improving or changing a particular symbol. Careful graphic review is essential for creating good, clear design, but it cannot answer the question of whether the user will understand the final product. Only the systematic evaluation of proposed designs with the eventual user can adequately answer this question.

Although meaningfulness is not the sole criterion for judging a symbol's effectiveness, it is the dimension upon which failure will have the most serious consequences for the user. Despite Easterby's (1970) assertion that initial understandability may not be the best measure of effectiveness, the efficacy of training programs is doubtful, particularly for building safety symbols. Since no permit or license is required to use the built environment, no natural opportunity exists for training or testing. As a result, a building symbol's effectiveness will be heavily dependent upon its initial understandability.

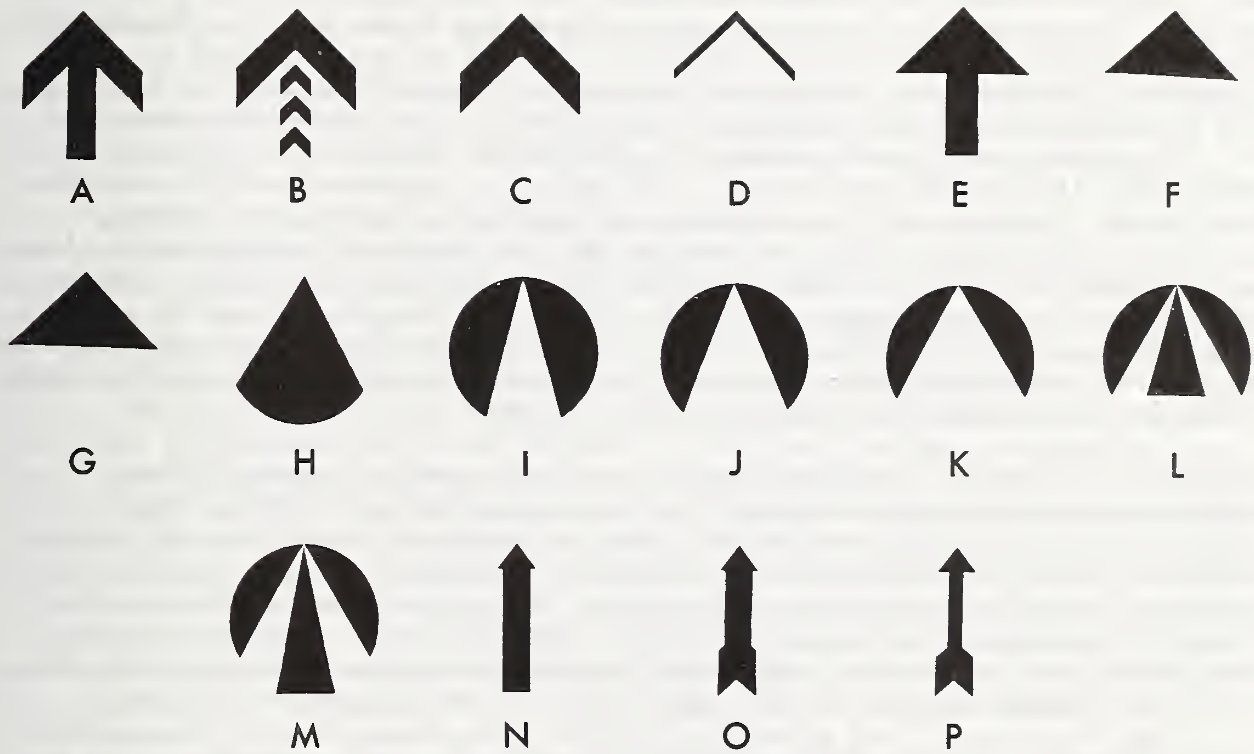
Symbols, which are intended to replace or supplement written signs, are useless--or dangerous--if they do not communicate accurately. For example, in a recent experiment, Collins and Pierman (1979) determined the understandability of a set of fire safety symbols proposed for international standardization. They found that a number of the symbols were poorly understood, with less than 25 percent of the subjects responding correctly to them. Symbols which performed poorly included such concepts as "exit" and "no exit." The "no exit" symbol particularly represented a serious symbol failure. Not only did the overwhelming majority of the subjects fail to understand this symbol, a large number interpreted the symbol to mean "this way to safety". An instance in which a symbol is given a meaning opposite to that intended represents a potentially dangerous situation.

Because of the need to determine the effectiveness of a set of symbols experimentally, the International Organization for Standardization (ISO) has issued a number of guidelines for evaluating symbols. Any set of symbols which involves public safety must be evaluated, and the final set of proposed symbols must achieve a criterion level of understandability (ISO, 1980). ISO has also proposed procedures for evaluating symbols which rely heavily upon the assessment of initial understandability. Such assessment is essential for predicting the eventual effectiveness of a symbol set after implementation in buildings. Attention to graphic principles can improve a symbol graphically and perceptually, but cannot determine if it is meaningful to the eventual user groups. Yet, the need for testing symbols does not negate the need for graphic evaluation and review. Rather, it expands the task of developing good symbols to include the user as well as the designer. As Modley (1966, p. 125) put it, "Neither engineers nor designers alone should be permitted to have the final say on symbols. The task requires the combined efforts of psychologists, linguists, educators, anthropologists, sociologists, lawyers, engineers, designers and many others. The need is so important that we cannot afford to take hasty and inadequate measures, but we should not delay in undertaking this essential task."

Ideally the use of sound evaluative techniques to design and select symbols can be used to combat the current proliferation of symbols. As Dreyfuss (1972, p. 19) put it, "Symbols have multiplied to an alarming degree, along much the same lines of divergence as languages. Today it is this very diversity and multiplicity of symbols in our international life that is a matter of such immediate concern." Without careful design, comprehensive evaluation, and subsequent standardization, symbols will fail completely at the task for which they are intended - communication.

Facing page:

Arrow symbols evaluated experimentally at NBS



3. EXPERIMENTAL EVALUATION OF SYMBOL EFFECTIVENESS

3.1 REQUIREMENTS FOR COMMUNICATION

In section 3.0, the results of experimental assessments of symbols for specific applications are discussed. This review concentrates upon the methodological procedures and experimental findings, and suggests implications for the development of effective symbols.

Before graphic symbols are implemented as part of a visual communication system, they must be evaluated in terms of their effectiveness in transmitting information. There are a number of psychophysical and psychological testing methods that have been used to evaluate symbols, with varying degrees of success. These various methods will be discussed for specific sets of symbols in five major applications: highway and road symbols (section 3.2); vehicle and machine symbols (section 3.3); public information and directional symbols (section 3.4); product hazard warning symbols (section 3.5); and safety symbols (section 3.6).

The effectiveness of a symbol is determined by its success during several stages of information processing. A symbol must be detectable or visible. It must be discriminable or easily differentiated from all other symbols in the display or set. A symbol should be recognizable when it is seen again even in a different context. (Assessing the recognizability of a symbol often requires determining its understandability, since the observer is typically provided with a set of linguistic labels and is required to "recognize" the correct one). A symbol must be understandable and convey the same meaning to all those who encounter it. Information must be processed at each of these stages if a symbol is to communicate its message successfully. In addition, symbols should be evaluated in terms of attention-getting, salience, or ability to "grab" the viewers. Finally, a symbol should also be assessed for its actual effectiveness in altering behavior, although this last response has rarely been examined in traditional research. For example, a symbol such as "no right turn" may be well understood, but if vehicles continue to turn right, this symbol is not effective. The various research methods which have been used to answer these questions are discussed in the context of the individual studies.

In summary, six major questions should be addressed when determining the effectiveness of symbols as elements in the communication process. These include: Can the symbol be detected? Can it be discriminated from all other symbols? Can it be recognized when seen in a different context? Does it consistently communicate the intended meaning? Does it get the user's attention? And, does it alter behavior appropriately? Although these questions are not particularly design oriented, successful symbol design requires answers to them so that the symbol message may be communicated to the intended user group.

As the following experiments are reviewed, it will become rapidly apparent that not all researchers have addressed all six questions. The majority of studies have concentrated upon determining the meaningfulness or detectability of a symbol set, using a variety of research approaches. Nevertheless, the questions of discrimination, recognition, and behavior have been studied in selected instances and should be addressed in any complete evaluation of symbol effectiveness.

The evaluation of symbol effectiveness will be discussed for each of five applications in the following pages. Then, general conclusions will be drawn about the kinds of research that have been done, as well as the need for more extensive investigations, particularly in the area of safety symbols.

3.2 HIGHWAY SYMBOL RESEARCH

The first application to be discussed in detail is highway symbology. Much symbol research has concentrated upon the use of graphic symbols for vehicular traffic information. Because of the speed of movement on highways and extreme variations in lighting conditions, issues such as speed of detection, reaction time, and comprehension have been researched along with issues such as overall meaningfulness and discriminability.

Many of the experiments on highway symbol signs either measured reaction time directly, or imposed response time constraints upon their subjects. In most instances, the correctness of the response was also assessed. Reaction time experiments will be discussed first.

3.2.1 Reaction Time to Word and Symbol Signs

In an early experiment, Janda and Volk (1934) measured reaction time, or speed of recognition, as well as the accuracy of response to various highway sign types. The speed of learning the meaning of each sign was also assessed.

Twenty signs were used, of which eight were words with a surrounding shape, six were graphic symbols, two were sign shapes, and four were words without a surrounding shape. In this experiment, subjects pushed a lever in the direction indicated by the symbols and signs. Reaction time and response accuracy were recorded. Each trial was repeated five times. Comparison of the reaction times of the 160 drivers studied revealed significant differences between the various signs. The reaction time was shortest for the symbols and greatest for words alone, for a difference of about 200 msec. Reaction time was shortest for the arrow used to indicate turn or curve, and greatest for the standard "stop" sign. The use of a specific sign shape appeared to increase rather than decrease reaction time, as well as to increase the number of incorrect responses. Because reaction time decreased after five successive presentations of a particular sign shape, Janda and Volk suggested that the initially poor responses were due to ignorance and that training should improve performance. Consequently, dual coding using both a message and specific shape might prove to be most effective after training. The beneficial effects of training were also demonstrated by an improvement in performance over trials for symbols for "cross road" and "side road." Finally, the authors observed that the subjects' reaction time was shorter for several initially unfamiliar symbols than for familiar words after training. They concluded that word signs are a relatively poor way of conveying information.

Walker, Nicolay, and Stearns (1965) compared the understandability of symbol and word road signs when each was presented briefly. The authors measured the ease of learning and accuracy of response (or understandability) for both symbols and words. Walker et al. noted that two conditions should have predisposed the word signs to be recognized more easily: the word signs were part of conventional U.S. sign practice at that time, and all stimuli were printed in black and white to eliminate potentially helpful color cues. (In a subsequent stage, they tested colored symbols.)

In the first stage, the accuracy of recognition for seven black and white signs and symbols was determined for 70 undergraduates. Each stimulus was presented tachistoscopically for .06 sec. after five minutes of familiarization. Subjects identified the sign or symbol, using a written code. This procedure was repeated using red slashes and bars on the symbols for 11 subjects. Finally, 24 hours after phase 1, 27 of the original subjects defined the meaning of the three international symbols in a test of retention.

Analysis of the data indicated that the international symbols were identified significantly more accurately on both trials by both sexes than were the word signs, regardless of color or delay before re-testing. Furthermore, the 27 subjects who participated in the retention test demonstrated perfect recall. Thus, these results indicate for this very limited sample that "symbols can be recognized significantly more accurately than word signs... A plausible explanation for the superiority of the symbol sign would be the greater perceptual simplicity of the symbol. The symbol is more visually integrated, whereas the letters of the word signs are more fragmented" (Walker, et al., 1965, p. 325).

The authors noted that symbols performed better than words even when the subjects were familiar with the word signs, and color cues were eliminated. Furthermore, the symbols chosen were those that Brainard, Campbell, and Elkin (1961) had demonstrated to be difficult to learn. Walker et al. also suggested that the major criteria for evaluating symbol effectiveness should be accuracy, speed of recognition, and meaningfulness of the symbols. Finally, they concluded that because symbols are easier to recognize and interpret, they should lead to increased safety, not only on highways, but also in industrial settings.

Dewar, Ells, and Mundy (1976) reported reaction time data from three studies. They compared the reaction time to symbolic messages and word signs for tasks of increasing complexity designed to approximate an actual driving situation. Then they compared the laboratory reaction time measures with and without visual distraction with measures of sign perception in actual driving.

In the first study, 16 subjects were shown 26 signs in three categories: regulatory, warning, and information. Subjects verbally classified a subset of these signs as either warning or regulatory and also defined (identified) them. Subjects were instructed not to respond to the six information signs. Half of the 26 signs contained words and half contained symbols. Two visual angles were used: 0.57° and 1.42°. Each subject's reaction time for both the classification and identification tasks was measured from the onset of the slide to the onset of the verbal response. A four-way analysis of variance indicated that reaction time was shorter for the classification task than for the identification tasks, as well as for larger signs, word messages, and warning signs. The superiority of word messages over symbolic messages, however, decreased for smaller visual angles.

In experiment 2, the use of the classification and identification procedures was repeated. In addition, a secondary loading task was added to this procedure to divide the subject's attention. This task required subjects to attend to a list of randomly presented numbers, from 1 to 99, and press a response key when a number between 50 and 59 appeared. The results were similar to those found in experiment 1, although the use of a loading task lengthened reaction time. Again, word signs performed better than symbol signs. Both task and sign type affected reaction time significantly.

In the third experiment, a motion picture of a rural highway was used as a sign background to simulate visual distraction during the sign identification/classification task. In addition to classifying and identifying the signs, subjects were also required to move a knob to maintain a constant speedometer reading of 80 km/h (while the experimenter varied the readings on the speedometer) to simulate an actual driving task. In this experiment, the superiority of word messages over symbols disappeared.

In all three experiments, reaction time was shorter for the classification task than for the identification task, and shorter for warning than for regulatory signs. In the first two experiments reaction time was faster for words than for symbolic messages. The authors suggest that this occurred because the verbal response given by the subjects was the same as the verbal messages on the signs. For the symbols, however, subjects had to develop their own verbal response -- not merely read it off the sign. The advantages of the word signs over the symbol signs disappeared when the signs became less legible. The authors suggest (1976, p. 387) that "it is possible that any reduction in legibility will detrimentally affect verbal messages more than symbolic messages." Symbolic signs were also found to perform better than legibility measures obtained in an earlier, on-the-road experiment, perhaps because of the lack of a verbal response bias mentioned earlier or because the symbolic signs are actually legible at a greater distance. (Relative size measurements for the two sign types were not reported). The authors determined that (1976, p. 390) "for symbolic messages, the verbal reaction time was a valid predictor of legibility distance only under conditions that involved attention demands and visual distractions of the type experienced under normal driving conditions (experiment 3)." Dewar et al. concluded, however, that reaction time is a valid predictor of traffic sign perception, particularly when combined with the loading and distraction tasks.

In a subsequent experiment, Ells and Dewar (1979) re-evaluated the effectiveness of reaction time in comparing traffic signs and symbols. Because previous research (Dewar & Ells, 1974; Dewar, et al., 1976) had determined that reaction time was shorter for verbal messages when the response was to verbalize the sign message, Ells and Dewar attempted to develop a less verbally biased measure. Ells and Dewar (1979, p. 162) note that "the smaller latencies for verbal messages may have been peculiar to the verbal identification task rather than resultant from more efficient processing of the meaning of verbal messages." As a result, after the experimenter read a traffic sign message aloud, the subject viewed a slide of the traffic sign and responded "yes" or "no" depending upon whether the visual sign and the spoken message were the same or different. The time to initiate the subject's verbal response was recorded, along with its accuracy.

In the first experiment, "yes-no" reaction time data were collected from 12 subjects for eight warning signs and eight regulatory signs. Half of the messages for each sign type were verbal (printed words) and half were symbolic. These data were correlated with previously obtained legibility data for the same signs (Dewar & Ells, 1974). (Legibility distance was defined as the maximum distance at which a sign could be correctly identified for a particular vehicular velocity.) The sign stimuli were shown at a visual angle which corresponded to that of a traffic sign viewed at 59 m and which was similar to the visual angle used in the legibility distance trials. Examination of the data between the two experiments revealed a moderate, but significant, correlation between reaction time data and legibility data.

In the second experiment, a direct measurement of comprehension time was made for 14 verbal and symbolic sign messages under both normal and degraded visibility conditions for 24 subjects. To eliminate extraneous information from the shape and color of the sign, half the subjects viewed only regulatory signs (7) and half viewed only warning signs (7). Again the reaction time to verbalize a "yes" or "no" response was recorded.

Analysis of the results indicated faster reaction times with the normal visual condition and the symbolic stimuli. Visual degradation resulted in a greater decline in performance for words than for symbols. Although symbolic messages typically performed better, some symbolic messages such as "bump" and "hill" were responded to more slowly than their verbal counterparts. Ells and Dewar suggested that this may be due either to the simplicity of these word signs or to the difficulty in symbolizing these concepts.

The authors suggested that symbolic messages are usually understood more rapidly than verbal messages, if the correct response measure is chosen. The shorter response time under degraded visual conditions for symbolic messages "again stresses the general superiority of symbolic signs, and also suggests that their effectiveness may extend over a broad range of environmental conditions (e.g., daylight, darkness, glare, fog)" (Ells & Dewar, 1979, p. 167).

Finally, Ells and Dewar commented that although there were significant correlations between the data on reaction time and legibility distance, these two measures do not necessarily assess the same qualities of traffic signs. "For example, legibility distance may be related to the visibility of individual features comprising a traffic sign message, whereas reaction time is more closely related to the speed of comprehension of the message. These considerations lead to the recommendation that both legibility distance and reaction time measures should be taken whenever a complete evaluation of sign adequacy is undertaken" (Ells & Dewar, 1979, p. 168).

Hoffman and MacDonald (1980) compared the effectiveness of symbolic and verbal turn restriction signs in terms of retention in short-term memory. In two separate experiments, they examined the effects of verbal interference (counting backwards) and visual interference (tracing a maze) upon the retention of sign information. In both experiments, verbal and symbolic signs containing either permissive or prohibited information were assessed. In the first experiment, 64 subjects were shown a sign, given a delay of 15 s with or without interference, and asked to select that sign from an array of 12 similar signs. The authors found that the accuracy of sign perception was very high even at a 100 msec exposure with no interference. A differential effect of interference type was found in which verbal interference had a greater effect upon verbal signs than did visual interference. Furthermore, there was a significant interaction between sign type and interference type for permissive, but not prohibition, signs. This suggested that subjects used visual coding for permissive but not prohibition signs.

In the second experiment, the response task was modified to include a "yes/no" decision indicating a subject's comprehension of the sign message. Again, two types of interference were studied, verbal and visual, with 40 subjects. Number of errors and response time were recorded. The response (recognition) time was found to be significantly lower for symbolic signs than for verbal signs. Verbal interference resulted in a higher error rate for verbal signs than for symbolic signs for the recognition task, while for the "yes/no" response task, symbolic signs with visual interference had a significantly higher error rate. The authors suggested (p. 249) that "information from both types of sign was retained in some non-verbal form for the yes/no response condition. In fact, several subjects reported recoding each sign as soon as it was presented in terms of the specific motor task which would subsequently be required; evidently this was the most commonly used strategy."

Because of the interactions among sign type, interference type, and response task, and because the real world is likely to contain both visual and verbal interference, the authors suggested (p. 250) that "in real use it appears unlikely that there would be significant differences between verbal, symbolic, permissive and prohibitive types of sign in the accuracy with which drivers retained them in short term memory." Nevertheless, Hoffmann and MacDonald's data do suggest that symbolic signs may be perceived more rapidly than verbal signs, even though the detrimental effects of verbal and visual interference upon sign perception are not clear-cut.

3.2.2 Glance Legibility Assessments of Understandability

One technique for assessing symbol effectiveness which sets time constraints upon subjects, although it does not measure reaction time directly, is termed "glance legibility".

King (1971, 1975) defined "glance legibility" as the percentage of correct matches between a symbol and an answer chosen from an array of symbols. The symbol is typically presented at several different viewing times. In addition, the array can contain either word or symbol signs, so that the viewers must match a symbol to a word or vice versa. In two separate experiments, King (1971) and (1975) assessed the "glance legibility" and meaningfulness of both symbol and word signs.

In the first experiment, King (1971) presented 208 subjects with 30 black and white signs. Ten of these were U.S. word road signs; ten were Canadian Pan American symbols (series 1); and ten were Quebec-United Nation symbols (series 2). First the subjects defined the various symbols, then they matched briefly presented (.05 to .3 s) symbols with one of the nine symbols shown in an answer array. In this way, meaningfulness was first measured, followed by glance legibility (which appears to be a combination of detectability and discriminability). Finally, subjects indicated which set of symbols they preferred.

The results indicated that the response to the symbols of series 1 was significantly more accurate than that to series 2. For series 1, the percentage of correct and generally correct responses ranged from 62.5 percent to 91.4 percent, while for series 2, the range was from 26 percent to 56 percent.

The glance legibility experiments indicated that at shorter viewing times, there were fewer correct matches for the word signs. This decrease did not occur for either of the two symbol series, nor was there a significant difference between these two sets. There was a significant difference, however, between the glance legibility scores for the word and symbol signs. Finally, 65 percent of the subjects found the symbol signs easier to match than the word signs.

King (1971) noted that the prohibition-type signs of Series 2 were particularly difficult for subjects to interpret, with a large number of opposite responses to these signs. The glance legibility data, however, indicated that even these "difficult" symbols were more accurate than words in transmitting a given message. These data also hinted that words may require more time for recognition (and matching). King (1971) cautioned that these results were obtained in a laboratory experiment and may or may not hold true in an actual driving situation where many distractions exist. Yet he suggested that symbols can be more effective than words, particularly after people have been familiarized with their meaning.

King (1975) used the glance legibility approach to study the effects of both delayed response and interference. His intent was to simulate an actual driving condition, in which there is typically a time interval between observing a highway sign and acting upon its information. During this time, the driver also continues to perform driving-related tasks, which could interfere with the comprehension of sign information. As a result, King (1975) repeated his earlier experiments with glance legibility but delayed the subject's response for 5 or 10 seconds after symbol presentation. In another phase, he added interference during the 10-second delay period. The symbols and signs were the same used during the 1971 experiment; namely, word signs, warning symbols (Series 1), and prohibitory symbols (Series 2). Again, they were presented tachistoscopically (for .33 and .05 s) in black and white to eliminate color cues. The first time represents the time that an automobile driver typically spends looking at highway signs, while the second is a more extreme time limitation. Ten subjects were tested at .33 s exposure; sixteen were tested at .05 s. All 26 subjects made 90 observations -- 30 with a 5 s delay, 30 with 10 s delay, and 30 with 10 s interference and delay period. Each subject matched the single sign/symbol stimulus appearing on the projection screen with an identical sign from a 10-sign answer array. During the interference period, subjects read aloud a series of individual letters for 10 s and then made the match.

Very few errors occurred at the longer viewing conditions for any of the three test conditions. For the shorter condition, the percentage of correct responses decreased as the delay period increased from 5 to 10 s, and decreased further when interference was introduced. Symbols received significantly more correct responses at the .05 s viewing condition, although there was no significant difference between the two symbol series. The use of interference degraded the response to words significantly more than that to symbols. Thus, even under more complex conditions, King (1975) concluded that symbols are recognized more accurately than words for short viewing times.

In another assessment of glance legibility, Plummer, Minarch, and King (1974) compared the reaction time and accuracy of response to symbol and word signs for 20 subjects. Using a tachistoscope, they presented black and white slides of 10 symbol and 10 word messages for 200 s or more. Half the subjects received special training in the meaning of symbolic traffic signs, while the other half did not. Subjects matched a word sign with one of 3 symbol signs contained in a subsequently shown answer array in the first portion of the experiment. In the second portion, a symbol was matched to one of 3 word messages. Subjects made a total of 60 observations, viewing each message/answer combination three times during the experimental test. Both the reaction time and accuracy of response were measured.

Analysis of the results indicated that the response to symbol signs was more accurate than that to word signs, but that the reaction time was faster for word signs. Training resulted in increased speed but had no effect upon accuracy. There were significant differences in the accuracy of responses and the reaction times observed for particular symbols. For example, the symbol which caused the most problem was the "school bus ahead" sign. It was suggested that this symbol should be re-designed. The authors concluded that highway symbols can be more effective than word signs if accuracy of response is used as the criterion.

In a different use of the glance legibility technique, Dewar (1976) compared the legibility of 15 symbols in each of four prohibitory modes. These modes included slash superimposed above the symbol, slash below the symbol, partial slash, and a circular red surround only. Dewar noted that the convention in European signs has been to use a red circle with a red slash over the symbol, to indicate prohibited acts, while in Canada the trend has been to provide a permissive symbol, surrounded by a green circle, which indicates only those acts which are legal. The choice between the two approaches is difficult, as there may be a number of permissive actions at an intersection but only one prohibited action. Yet "the prohibitive message must convey two essential pieces of information -- the specific action (e.g. riding bicycles) and the fact that this action is not allowed, thus making it more complex cognitively than the permissive message" (Dewar, 1976, p. 253). As a result, subjects often give the prohibition symbol a meaning which is opposite to the intended meaning.

In addition, the prohibitory slash may obscure the symbol. In this experiment, Dewar (1976) used symbols which contained the following characteristics -- two items, a vertical element, a horizontal element, or visual complexity -- to determine if various versions of prohibition would differentially affect the meaning of the symbols.

In the first portion of the experiment, Dewar presented 60 slides containing 15 symbols under the four different prohibition modes. These symbols were presented tachistoscopically for 40 msec to 34 subjects in groups of two to four. Using the glance legibility procedure, subjects matched each stimulus slide to a set of 16 symbols provided on an answer sheet. The entire set of stimuli was presented twice.

Analysis of the results from the first glance legibility experiment indicated that performance was best for symbols with no prohibitory slash for about 66 percent of the symbols. It was poorest for those symbols with the slash underneath.

A second experiment assessed glance legibility under visually degraded circumstances. Subjects sat further from the screen, visual angle was smaller, background visual noise (a street scene) was provided, and the stimulus was slightly out-of-focus. Exposure duration was lengthened to 100 msec, however. Subjects again matched the stimulus symbol to an answer array of symbols. As before, the presence of the slash reduced the legibility of the symbol. A comparison of the results for specific symbols from the two experiments using a rank-order correlation, however, revealed a significant correlation only for the "no slash" condition. Dewar (1976, p. 257) suggested that the lack of "correlations between legibility scores under the two conditions may indicate that at the greater distance (and lower clarity) subjects are depending upon different cues. At the closer distance, for example, it may have been possible to distinguish details on the symbols regardless of the slash conditions, while at the greater distance, subjects may have used cues such as sign and general shape of some symbols." Exposure duration may also have influenced these results. Dewar concluded, however, that his data do indicate that prohibition slashes can interfere with the perception of symbols. Yet the use of a red (or green) circle only to indicate prohibition (or permission) may be completely ineffective for those who are color defective. Hence, Dewar suggested the use of a partial slash for prohibition.

The experiments reviewed in Section 3.2.1 and 3.2.2 indicate clearly that symbolic signs can be perceived more rapidly and accurately than verbal signs, particularly under very short viewing conditions. These advantages are enhanced if the response measures and viewing conditions are chosen appropriately.

Because the use of a verbal labeling response appears to bias the reaction time data toward word signs, Ells and Dewar's (1979) use of a verbal matching procedure or King's glance legibility technique are useful experimental approach alternatives. Nevertheless, the reaction time and glance legibility measures provide an indication of a symbol's eventual effectiveness in an actual driving situation where the speed and accuracy of response to a highway sign can be critical.

3.2.3 Direct Assessment of Meaningfulness

The previously described research has dealt primarily with time-dependent measures such as reaction time and glance legibility. Although accuracy of the response has been recorded, symbol meaningfulness has not been the major factor. In other experiments, however, researchers have selected meaningfulness as the most salient characteristic of symbols for evaluation. In these experiments, subjects are asked to provide a definition of the symbol's meaning, rather than match a symbol against a word. Time constraints are typically not imposed.

In a five-phase assessment of meaningfulness, Brainard, Campbell, and Elkin (1961) evaluated 30 European traffic symbol signs with subjects in the United States. In the first phase, 29 subjects defined the meaning of each symbol. In the second phase, 33 new subjects completed a matching test in which the correct meaning was selected from a large list for each symbol. In the third phase, these subjects then received a brief training period, in which the symbols were presented while the experimenter defined them orally. After this training, the subjects again defined all the symbols. In the fourth phase, after 16 symbol definitions were read aloud, 31 new subjects rapidly sketched an appropriate symbol for each definition. Finally, 29 new subjects gave definitions of 10 new symbols constructed structed from sketches generated in the fourth phase.

Brainard et al. determined that the answers (definitions) of phase one were significantly correlated with those (matching answers) to the same symbols of phase 2, although there were fewer correct answers in phase 1 (54 percent compared with 74 percent). After training, the correct answers in phase 3 approached 100 percent. In phase 4, the authors found that common stereotypes emerged from the drawings for at least 9 of the 16 definitions. In fact, for three definitions all of the drawings were essentially the same. Even in the remainder of the cases, common elements emerged for the majority of the symbols. After testing new symbols based upon these stereotypes, Brainard et al. commented "that the interpretability of the signs based on the stereotypes is superior to that of the corresponding European signs in all instances" (1961, p. 131).

The authors also found that those concepts which were generally interpreted correctly in phases 1 and 2 also elicited more images that were similar in phase 4 (sketch) - suggesting that there is a "common" stereotype for particular concepts. The signs with low scores or lack of agreement in symbolism were those which tended to be more abstract (using a prohibitory slash, for example).

Those with scores above 85 percent correct tended to be a direct pictorial representation of the concept or had a counterpart in existing U.S. symbol systems. (The symbols that were most accurately interpreted tended to be those of a unidimensional character and relied upon more common stereotypes.) Brainard et al. claimed that the presence of stereotypical images increases the interpretability (or understandability) of symbols, as evidenced by the fact that the European symbols had an average interpretability score of 45 percent, compared with a score of 75 percent for the stereotype-based symbols. Nevertheless, at least four of the stereotypical symbols were interpreted only moderately well with scores ranging from 76 percent to 38 percent. Brainard et al. also found that the meaning of prohibitory signs was often reversed, although brief training improved performance on all symbols to near 100 percent. Although the use of circle and slash to symbolize prohibition was not widespread in the early sixties when this experiment was performed, it should be noted that the current extensive use of such symbols for highway signs should have improved their meaningfulness.

Finally, the authors indicated that even very brief training increased the interpretability of all the symbols to near 100 percent accuracy. Brainard et al. (1961, p. 136) concluded that: "A small number of the European road signs could be efficaciously used in the United States, without necessitating prior instruction as to their meaning. The majority of the signs, however, could not be used without a minimal degree of familiarization." This experiment is relatively unique in that the authors evaluated the initial understandability of the symbols, the effectiveness of training, the existence of symbol stereotypes, and provided suggestions for improving the meaningfulness of specific symbols.

In a more recent experiment, Griffith and Actkinson (1977, 1978) evaluated a very large set of traffic control symbols currently used in Europe. Because they were concerned about the high failure rate of Army personnel on the road sign section of the European driver's license test, Griffith and Actkinson assessed both initial interpretability and training effectiveness for the symbol set. They measured interpretability in terms of the percentage of people who responded incorrectly to the signs -- both before and after training. The effects of training method were assessed by determining interpretability after one of three different instructional techniques had been administered.

In the first training condition, "sign only", individual signs were presented as slides, while their meanings were read orally. In the "sign elaboration" condition, mnemonic cues were added to the oral presentation. These cues were intended to enhance the recognition of the sign by providing a memory aid. The third instructional condition was the standard Army instructional technique which did not use slides. In the first two conditions, training was followed by a testing period followed by feedback on the meaning of each sign. During all training conditions, subjects were shown a total of 128 slides.

After one week, subjects were tested on their ability to recognize 50 of the 128 signs by writing the number of the slide next to the correct answer on an answer sheet which contained the 50 alternative signs. Each sign slide was projected initially for 30 s, and then for 15 s in a second iteration. Thirty-eight subjects were tested in the "sign only" condition, 45 in the "sign elaboration" condition, and 117 in the standard or Army condition. Two measures were used to assess the effectiveness of training -- the number of signs recognized correctly during training, and the number recognized correctly one week later.

Analysis of the results indicated that there were no differences between the "sign only," "sign elaboration," and the standard Army instructional conditions. All three techniques appeared relatively inefficient in decreasing the number of errors. Nevertheless, the percentage of errors did decrease as a function of the innovative two hour training procedure, although not significantly. As a result, the authors suggested that improved training techniques could result in better performance if a minimum of four hours could be devoted to training. "Although it remains our contention that effective memory training techniques can be devised for road sign instruction, the mnemonic enhancement of road signs has yet to be demonstrated. Regardless of the method of instruction, it appears that two hours of training is insufficient" (Griffith & Actkinson, 1977, p. 394).

Griffith and Actkinson also questioned the ready interpretability of the international road signs. They contended that performance was still below 100 percent even after two exposures. There was a wide range of variation in the percentage of incorrect answers -- from 0 to 86 percent during training and 0 to 32 percent during testing. Furthermore, during the testing phase, 10 of the 128 signs were missed by more than 50 percent of the subjects in at least one condition. The authors comment that "these data support the contentions of Cahill (1975) and Kolers (1969) regarding the difficulty of developing a truly universal symbology and the need for empirical research to develop a symbology based on psychologically realistic principles" (1977, p. 394).

As Brainard et al. (1961) had noted, the best understood signs were those with direct counterparts in U.S. signage or those which were direct pictorial representations of the hazards. "Given a lack of familarity with the signs, the general rule appears to be that the more abstract the pictorial representation, the lower the interpretability" (Griffith & Actkinson, 1977, p. 394). The authors also concluded that there does not appear to be a good technique for teaching people the meaning of symbols quickly and effectively. Griffith and Actkinson (1978) suggested that memory for pictures can be superior to that for words. To improve symbol understandability, however, the problem becomes one of linking the visual image, which may or may not have an intrinsic meaning, to some semantic representation. Hence, the authors decided to use a mnemonic technique to add semantic meaning to the visual images. Using this technique, they elaborated upon the pictorial representation of the sign so that the subject would have a verbal description of the picture as well as the visual picture itself. Difficulties with this procedure may have arisen because of the short training time. In addition, because the study was developed and executed on very short notice, the best mnemonic cues may not have been developed. The authors concluded that the results might be improved if the subjects were first taught how to make mnemonic cues in general, and then allowed to generate their own idiosyncratic cues for specific signs. Griffith and Actkinson also noted that subjects could miss a particular sign, which they really understood, because of confusions with similar alternatives on the answer sheet. These authors also considered the test situation to be artificial because no list of alternative meanings is ever provided for a given symbol in an actual driving condition.

Although Griffith and Actkinson did not discuss it, problems may have arisen because they presented their subjects with a very large number of symbols to learn. It is possible that the task of learning these 128 signs was extremely taxing and may well account for some of the poor performance. (Easterby (1970) commented that recognition is poorer with symbol sets containing more than 100 symbols). In addition, a quick glance at the symbol set reveals that many are unfamiliar and, in some cases, contradictory to those used in current U.S. practice. (See figure 9.) The prohibition signs are particularly confusing in that a red circle is used to indicate prohibition while a slash -- a double negative, in fact -- is used to remove this prohibition. As a result, the correct answer may have been difficult to remember. Nevertheless, the data collected by Griffith and Actkinson reinforce the need to evaluate the understandability of symbols for public use. These data also indicate instances where the symbols could have profitably been redesigned, at least for U.S. observers.

Using another response method, the semantic differential, Dewar and Ells (1977) assessed the meaningfulness of traffic sign messages. The semantic differential measures the meaningfulness of an idea or object by having a subject rate the idea on a series of paired adjectival scales. These adjective pairs are designed to group attributes into four factors -- evaluation, activity, understandability, and potency. Dewar and Ells (1977) commented that although the application of the semantic differential to traffic sign evaluation is unusual, previous research has indicated that this measure can provide a valid and reliable indication of subjective meaning. Thus, the evaluation factor should measure attitudes, while the understandability factor should reflect the subject's knowledge of the concept being rated. Both understandability and subjective evaluation can be important factors in the comprehension of traffic signs.

Dewar and Ells tested 31 undergraduates, using 20 color slides of traffic symbols, to see if psychological meaningfulness as defined by the semantic differential were in fact related to the understandability of these traffic sign symbols. Subjects were asked to give the meaning of all the signs first, and then rate them on the semantic differential. Correlation of the scores for actual and rated meaningfulness was found to be quite high.

In a second experiment, Dewar and Ells examined the relationship between glance legibility and rated meaningfulness. Twenty subjects were tested on 20 sign messages, composed of both symbols and words. The subjects, who were tested individually, matched colored slides of each sign, projected for 40 msec., with one of 20 signs located on an answer sheet. After the legibility experiment, the subjects rated each sign on the semantic differential. The glance legibility scores correlated with the evaluation and understandability factors for the verbal signs only.

Dewar and Ells commented that the semantic differential provides a good way of determining the meaningfulness of a sign because it does not ask for preference

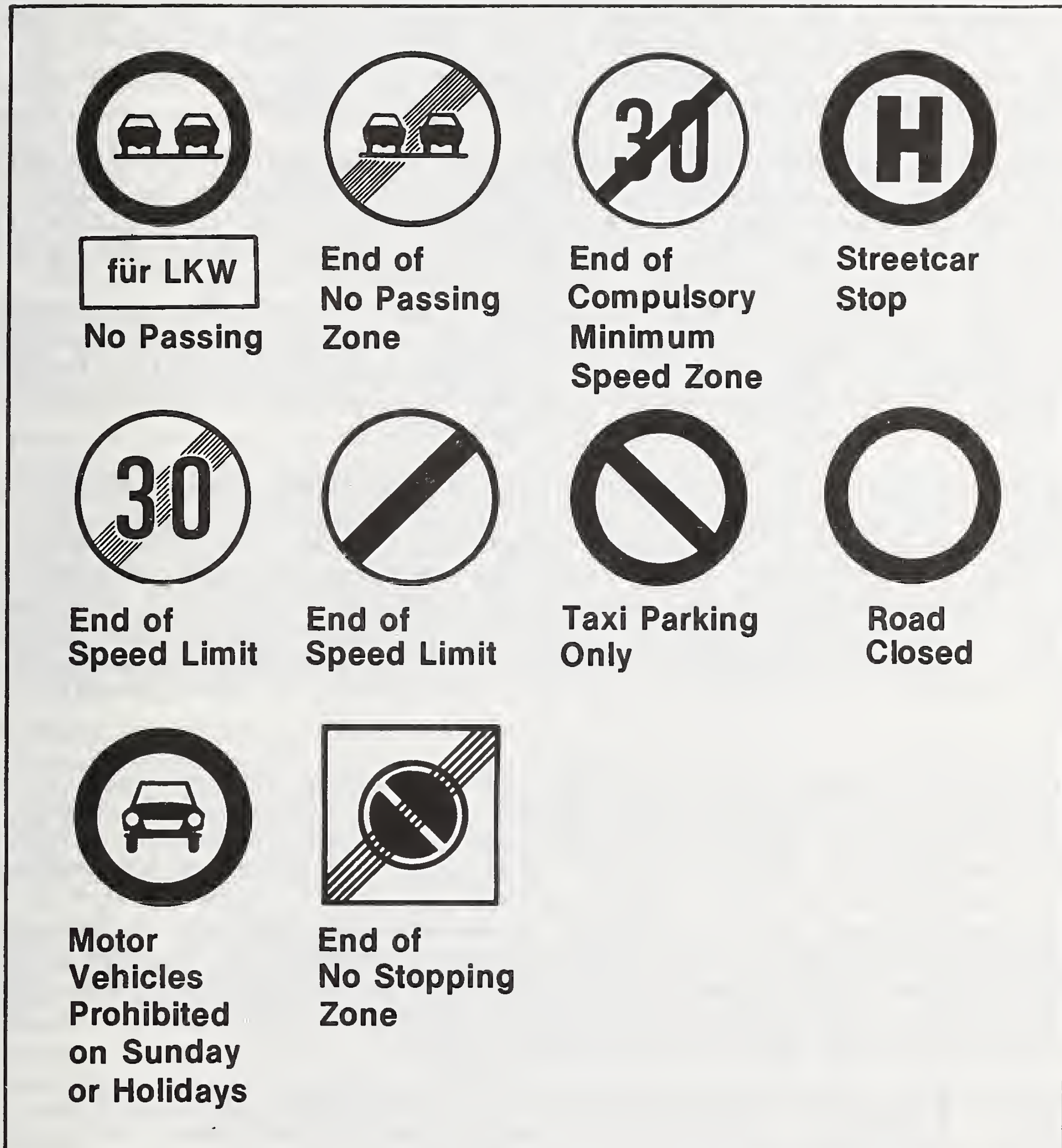


Figure 9. Highway symbols contradictory to U.S. practice

or familiarity judgments, and because it requires ratings on several adjective pairs for each concept. Yet, because it only rates meaningfulness, the semantic differential should be used along with measures of perception, such as glance legibility and legibility distance, in a complete assessment of symbol effectiveness. The authors did not discuss the lack of correlation between the semantic differential scores and the glance legibility scores for symbolic signs. Use of the semantic differential does, however, provide some indication of how meaningful a particular symbol is for a given concept. This technique could be used to discriminate between several different symbols proposed for a single concept. Use of the definition technique could then provide some data on confusions and misinterpretations within a set of symbols.

Hulbert, Beers, and Fowler (1979) evaluated the understandability of 16 traffic control devices, including symbols, with 3164 drivers throughout the United States. The devices studied included nine symbol signs, six signal lights, and eight pavement markings. The authors developed a 15-minute film depicting the various control devices which they showed to subjects in various civic and other groups. The film contained a question on each device, which was answered by the subjects before the next device appeared. Subjects completed a multiple-choice questionnaire which was chemically treated to provide immediate feedback on the correctness of each response. The authors also obtained demographic data in terms of sex, age, driving experience, etc.

The understandability of the traffic control devices ranged from 18 percent for the school zone symbol signs to 97 percent for the U-turn symbol sign. No device was understood by all drivers. Generally, however, the symbols were better understood than the signals or pavement markings, although the school zone and school crossing symbols were not well understood (and were often confused). Although there were wide differences in understanding by individual motorists, there were no regional trends. There was a tendency for males to score slightly better than females and for older drivers to score generally lower.

The authors conclude that "Even the best understood signs were not well understood by 5 to 10 percent of the drivers tested. Ninety percent understanding may seem all right, but if there are 10 uncertain drivers of every 100 using an intersection having 5,000 ADT, that means a potential of 500 confused drivers who may make an unsafe and/or unlawful maneuver or suddenly slow to a stop to figure things out... It cannot be assumed that motorists' failure to understand the meanings of these traffic control situations is due solely to weaknesses of perceptual powers of our nation's drivers" (Hulbert et al., 1979, p. 47). In fact, the device itself may not adequately convey its meaning, or the driving situation may be inordinantly complex.

Some of the poor response may have been due to complexities of the testing situation. Inspection of the multiple choice answers reveals some rather similar and confusing alternatives. For school zone and school crossing area, the answers provided were: "school children in area", "school crossing located here", "school zone", and "pedestrian crossing". Discriminating among these choices appears to be a rather difficult task. Furthermore, a driver's correct response to any of those ideas would be to slow down whether or not he/she fully comprehended the four gradations in meaning presented by the various choices. Generating good "wrong" answers to multiple choice questions is one of the most difficult aspects of this type of procedure. Nevertheless, the use of immediate feedback on the questionnaire administered by Hulbert, et al. provides a unique opportunity for both testing and simultaneous training.

3.2.4 Comparison of Laboratory and Behavioral Data

Another measure of symbol meaningfulness is the direct assessment of changes in behavior due to the symbol. Thus, Forbes, Gervais, and Allen (1963) compared the understandability of a symbol in laboratory experiments with the actual reaction to the same symbol on the highway. Concerned with the use of symbols for lane control in emergency situations, they devised two control symbols, a green arrow and a red "x". They assessed the effectiveness of each in three separate laboratory experiments followed by a field experiment.

Forbes, Gervais, and Allen (1963) conducted laboratory studies in which the amount of information given to the subjects about the meaning of the signals varied. Colored slides which depicted six symbols in place on a bridge were shown to a total of 253 students. In the first study, subjects, who were given a series of six possible driver actions such as "slow in lane", indicated the driving action that they would perform in response to a series of symbols. In the second study, subjects who received no information about possible driving actions indicated the appropriate action for each symbol in all traffic lanes. In the third study, subjects who received the driving information responded only for certain traffic lanes (rather than all three). The results indicated that the lack of information in the second study resulted in confusions and poor performance. Nevertheless, for all three groups, one symbol, the red "x", performed significantly better than the other five symbols tested.

The field experiment assessed the effectiveness of the red "x", used in the laboratory experiments, for an actual driving situation. The researchers placed a lightweight barrier in one lane, and then measured the point at which a car would swerve to avoid it both with and without the red "x" lane control symbol. Analysis of the results indicated that the starting point of the swerve was earlier when the red "x" was present.

Forbes et al. (1963) selected an effective symbol for lane control based upon performance in laboratory testing. This symbol was judged to be more discriminable and understandable than the other symbols tested. Direct observation of highway behavior confirmed that this symbol did alter behavior in the direction desired. Although Forbes et al. did not directly compare their symbol with word signs, they pointed out that a comparable word sign would take up much more space, be more expensive, and be more difficult to use as a temporary lane-change measure.

Dewar and Swanson (1972) also compared the effectiveness of words and symbols in both a laboratory and field experiment. As a rationale, they commented that "The use of a symbol assumes that the viewer knows the meaning of it. It is generally assumed that certain symbols or pictures will be understood on the basis of some intrinsic meaning that is obvious to all; however, cultural differences do exist, and some symbols may be inappropriate for certain countries" (Dewar & Swanson, 1977, p. 16). Hence, there is a need to test the recognizability of symbols for specific populations and applications.

Dewar and Swanson (1972) claimed that symbols can be visible at a greater distance than a written message for a sign of equal size. Furthermore, because different written messages vary in length, they require different sign shapes and sizes as well as varying letter types and sizes. As a result, uniformity is impossible, particularly for regulatory signs. These problems can be avoided through the use of symbols, although certain messages such as "Keep right except to pass" or "Slow down" remain difficult to symbolize.

Dewar and Swanson also discussed certain inconsistencies in current symbol use. A symbol may reflect the nature of a hazard, the result of a hazard, or provide positive or negative instructions. The authors concluded that the real concern of researchers should be with the action taken in response to a sign rather than the "recognizability" of the sign. For example, they noted that numerous inappropriate responses have been reported for the prohibitory-type signs.

As a result, Dewar and Swanson conducted both laboratory and field studies of traffic signs used in Canada. In a laboratory experiment, they examined the response to these signs under short exposure conditions. Before the experiment began, subjects were shown all the signs for 30 s each. Then they identified the meaning of color slides of 23 signs projected for .04 s tachistoscopically. In the second part, subjects identified signs located in a photograph of an actual intersection. Part 1 was always presented before Part 2. Three groups of subjects were used. The first, 148 volunteers from the city of Calgary, was tested in sets of six. The second, a group of 216 drivers' education students, was tested in sets of 42 to 68. The third group, 51 different drivers' education students, was tested after the end of training to assess any training effect. (No difference was found for distance from the projection screen).

The results indicated that all subjects performed better for symbols than for words (in Part 1). In Part 2, symbols such as the positive turn-restriction symbols performed better than words alone or the negative turn-restriction symbol. Only the words "No U turn" did better than the comparable symbol sign. "In general when a sign was compared with a similar sign containing additional information, such as time or words or both, the simpler version was more easily recognized. Adding something to a sign appears to increase confusion and make the symbol more difficult to recognize" (Dewar & Swanson, 1972, p. 19). The authors found some limited sex and age differences. Generally, the differences favored younger subjects.

In the second experiment, a field study, the relative effectiveness of a negative and positive version of a "no left turn" symbol was determined by counting the number of cars making illegal left turns during a prohibited time.

The results indicated a range of violations from 1.1. to 9.3 per day for the two symbols. Unfortunately, no baseline field data had been obtained for the positive symbol so it was not possible to compare the two symbols directly. Furthermore, the violations were much greater for westbound cars than for eastbound cars. Nevertheless, both the field and the lab study suggest that the positive symbol performed better than the negative. "This may be due to the positive signs being intrinsically more meaningful or possibly to the fact that they have been in more common use in Alberta and are simply more familiar. Word messages were generally more poorly recognized than symbols" (Dewar & Swanson, 1972, p. 22). Indeed, the addition of words or numbers was often found to reduce the comprehension of a symbol.

Dewar and Swanson concluded that future symbol research should concentrate on the relative effectiveness of positive versus negative symbols. In addition, the recognizability (or understandability) of traffic control symbols should be evaluated, particularly for abstract symbols that have no intrinsic meaning of their own.

3.2.5 Visibility of Highway Symbols

Johnston, Cole, Jacobs and Gibson (1976) reported legibility data obtained for 16 pairs of symbolic and alphanumeric highway signs. Subjects identified the signs both when their vision was blurred deliberately and when it was clear. The legibility distances for the symbols were typically about twice those for the alphanumeric signs, although these distances varied for different sign pairs. Of interest is the fact that certain alphanumeric signs performed almost as well as symbols because they had "characteristic word layouts which afforded a unique shape cue to the message. This shape cue enabled the signs to be identified as 'symbols' at sizes much smaller than those necessary to resolve the words. Examples of these signs are 'slippery when wet,' 'road narrows,' 'no entry' and 'divided road ends." If the legibility of such signs depended simply on the visual resolution of the letter limbs then these signs should be legible at distances when the only cue evident was the layout of the words" (Johnston et al., 1976, p. 603).

On the other hand, the rate of information transfer was greater for alphanumeric signs than symbolic signs under clear vision conditions. This may have been due to the "higher compatibility between the verbal identification of the legend display than that of the symbolic display" (Johnston et al., 1976, p. 603). This problem was also raised by Dewar et al. (1976). Whatever the reason, the superiority of the word signs decreased under conditions of blur so that the information transfer rate was about the same for both sign types when vision was defocused.

In the final study of highway symbols reviewed here, Smith and Weir (1978) evaluated the effectiveness of different directional symbols under poor visibility conditions. Convinced that symbolic road signs were superior to word signs, they wanted to improve them by finding the best image for each message. For this task, Smith and Weir used eight variations of a directional symbol (arrow) to assess the most visible symbol under laboratory conditions of blur and low contrast. Blur was chosen to simulate the effects of different visual acuities upon visibility (particularly for nighttime conditions). Contrast was chosen to represent a glare factor affecting visibility during bright sunlight. Smith and Weir also performed a subjective assessment, since previous research had suggested that a symbol which performed well under conditions of blur and contrast might not be liked because of its unconventional shape.

Eight symbols were rear projected for 0.5 s. In the blur experiment subjects judged the direction in which the symbol pointed. Eight levels of blur and eight directions were used for each of the eight symbols making a total of 512 presentations for fourteen subjects. The results indicated that there were three distinct groupings of symbols in terms of response to blur. In the second phase, the percentage of correct direction determinations was measured for each of the eight symbols as a function of contrast level. Eight levels of contrast were used. A second projector provided a uniform veiling glare, and varied overall luminosity. Twelve subjects participated (nine from the first experiment). Analysis of the results indicated that there were two distinct groupings of symbols in terms of response to contrast. In the third phase, subjects ranked the relative acceptability of the eight symbols. Seventy-one subjects ordered photographs of symbols in terms of their apparent effectiveness as a directional indicator. Because different performance criteria can lead to differences in rank ordering, the authors suggested that these criteria be as relevant as possible to the final task. The results from the three phases showed that the data on visibility in the laboratory agreed with the data on recognition distance and reaction time from previous field experiments. Although two symbols tested well in terms of visibility at low contrast and resistance to blur, one of these was ranked as the least acceptable symbol. As a result, Smith and Weir commented that "unconventional and unfamiliar symbols may be significantly more visible than the more conventional and familiar symbols. However, any advantage offered by completely new symbols may be offset by lack of understanding of their meaning" (Smith & Weir, 1978, p. 251).

3.3 AUTOMOTIVE AND MACHINERY SYMBOLS

Other investigations of symbol understandability have centered on the application of pictograms to mark controls and provide operating information and instructions in cars, trucks, and machinery. The impetus for this application derives from the international sale of machinery and equipment and the need to convey operating instructions accurately without the use of written language. Also, because so many controls are quite small, symbols are preferred to words on operating equipment. Finally, because of product liability factors, warning information and hazard indications must be applicable to all users -- literate or not. Consequently, symbols are increasingly used on farm and industrial machinery, cars and trucks.

Easterby (1970) claimed that the impetus for symbol use is often based upon commercial considerations and the need to expand foreign trade markets rather than any real consideration of the symbol's effectiveness in communicating information. He commented that "It is of course paradoxical that man has used pictographic and iconic recording (which was followed by the development of alphabets and language) and now, with increasingly complex technology, he must revert to symbolic representation. There may well be unexplored ideas in Egyptian hieroglyphics which we could use to define more adequately the principles of symbol design" (Easterby, 1970, p. 149). Nevertheless, the sets of automotive/machinery symbols in current use have received considerable research attention.

3.3.1 Machine Symbol Evaluation

Cahill (1975, 1976) assessed the interpretability of many of the symbols that Henry Dreyfuss (1966) had designed for use on farm vehicles and industrial machinery. (See Section 2.3.2). Cahill noted that although Dreyfuss and his associates had expended considerable time and energy in developing the symbols, they did not test their effectiveness with the intended users.

Cahill assessed the effects of both context and previous experience upon the understandability of 10 of Dreyfuss' symbols for 30 male subjects. Subjects were divided into two groups: context and no context. The context subjects were given a drawing of a typical cab (as in a bulldozer) with the correct location for each symbol indicated numerically. Both the context and the no-context groups wrote down a meaning for each symbol. All symbols were presented to the subjects as slides for 30 s apiece, with a 5 s response interval between slides.

Each response was independently scored as correct or incorrect by three judges. The judges also rated the subjects' previous experience with farm or industrial equipment based upon their answers to a brief questionnaire. Subjects were considered "experienced" if they had operated or designed heavy industrial or farm equipment or worked as a mechanic around it. Seventeen of the 30 subjects were classified as experienced and were about evenly divided between the context groups.

The results indicated that both context and experience facilitated performance, but that there were also wide differences in the understandability of various symbols. Nevertheless, the ordering of symbols in terms of correct identification was similar for both context and no context groups.

When Cahill (1976) examined the performance of individual symbols in greater detail, she determined that some symbols were affected more by context than others. A significant effect of context was found for both the "transmission" and "choke" symbols, while only the "turn signal" symbol was significantly affected by experience. In addition, the graphic quality of the symbol affected response accuracy. For example "fuel" and "horn", fairly graphic representations of a gasoline pump and a trumpet horn, were recognized by all the subjects, while "hand brake" and "turn signals" were also well understood. All of these are fairly direct pictorial representations of commonly encountered objects. "Engage" and "choke", on the other hand, were understood by very few of the subjects, with neither receiving a correct response from the "no context" group. "Engage" may have caused problems because it is a verb, not a noun, and because it is a conceptual, rather than pictographic, representation.

Cahill (1976) commented that symbols should always be provided in a context which can aid interpretability. Because symbols are not used in isolation, the use of a drawing to provide context is a conservative test of a symbol's understandability. Nevertheless, she noted that (p. 650) "we may now hold it to be an experimentally verified truth that not all symbols are created equal. Moreover, preexisting differences in difficulty level are not obliterated by context or user experience. While a few symbols may benefit more than others from the presence of either of these factors, the evidence suggests that in no sense can context or experience be relied upon to compensate for whatever aspects of certain symbols make them consistently difficult to interpret." She noted further that the most successful symbols should not require a great deal of specific operational familiarity. Symbols for specific concepts, such as machine-control actions or technological objects, cause problems because they are unfamiliar and occur only in particular situations. Cahill noted that there does not appear to be a relationship between the actual location of a symbol on the machinery and the accuracy of response. Furthermore, whether the symbol represents a control knob, instrument, or dial does not appear to affect understandability. Symbol representation or design is the primary influence on the understandability of a particular symbol, even when it is presented in context to a knowledgeable audience. As a result, Cahill concluded that symbols must be tested before application to determine if their design is effective.

3.3.2 Automotive Symbol Evaluation

Although Cahill's investigation is one of the few evaluations of the effectiveness of machine symbols, there have been several investigations of automobile symbols. Unlike the highway symbol research, the effectiveness of words and symbols for automobiles has rarely been compared. The use of symbols to indicate automotive controls has become accepted due to the advent of the "world car", designed to be marketed on a world-wide basis (Lord, 1980.) As a result, investigations of automotive symbols have concentrated on determining the understandability to a wide audience of one or more symbols for a given referent.

In a study of international automotive symbols, Heard (1974) compared the effectiveness of three different symbols for each of 24 referents for a very large (2593) number of licensed drivers. These symbols were designed to identify controls and indicators in automobiles. Since the study was commissioned by the International Organization for Standardization (ISO), Heard studied drivers in the United States, the United Kingdom, France, and Germany. Three age groups were studied -- 16-25, 26-55, and over 55 -- for both males and females.

Fifty-four symbols were studied in total. (See figure 10 for examples.) Three variations of each of 15 proposed referents were studied, along with one version of nine already selected for standardization by ISO. These symbols were tested in the appropriate location in an actual automobile or automobile mock-up.

Subjects were read a simulated test drive which involved the use of each of the symbols. As each command was read, subjects touched the appropriate control using the symbol for identification. The time to find and touch the symbol was recorded along with the accuracy of response. (All the symbols were obscured by cardboard before and after the actual test.) In addition, the subjects completed a questionnaire which provided data on their age, sex, driving experience, nationality and personal car. Most of the subjects were from the U.S. (1097) with 771 from the U.K., 572 from Germany, and 153 from France.

An analysis of variance indicated that there were significant differences between countries and symbol sets. Age and gender were ultimately ignored because of the difficulty of finding enough female drivers, particularly in the older age groups. In addition, the response time measures varied tremendously from country to country, so that the final data were only approximate. Problems also arose in that not all cars contained the appropriate controls, so that not all symbols were tested with all subjects. As a result, the percentage of correct response to a symbol is based upon the number of subjects responding, rather than the number tested. Finally, poor wording of the simulated test drive caused one of the "brake" symbols to be preferred over the others. Nevertheless, the data did permit a clear differentiation between the symbols in each of the proposed sets.

Heard (1974) recommended 12 symbols for the 15 new referents for use in automobiles on the basis of these data. For recommendation, the symbol had to be recognized correctly by 75 percent or more of the subjects, and be confused with other symbols no more than 5 percent of the time. Data for all countries, ages, and genders were combined to provide an indication of effective symbols.

Heard's approach is interesting in that she used a behavioral context to study the recognition and understandability of symbols. The subject's ability to locate and touch the appropriate control provides some indication of the potential effectiveness of the symbols in an actual driving situation. Nevertheless, Heard commented that although her approach allows the effectiveness of different symbols to be compared, it does not eliminate the need for an additional educational program to familiarize drivers with all symbols and their meaning.

In another evaluation of automotive symbols, Wiegand and Glumm (1979) evaluated the effectiveness of various pictographic controls and displays used on military vehicles. At the time of their experiment, at least five different symbol systems were used or proposed for use in both military and civilian vehicles. For at least one of these sets, the majority of symbols had been accurately identified by less than 50 percent of the German drivers tested. As a result, Wiegand and Glumm selected 25 pictographic symbols proposed by ISO in 1976 for

Figure 10. Examples of automotive control symbols

testing. In the experiment, Wiegand and Glumm asked 125 U.S. citizens to match pictures of the 25 symbols with the correct definition drawn from a list of 35 referents.

Analysis of the results indicates that the males performed slightly better than the females, although this difference occurred for only four symbols -- lower beam, windshield washer, head lamp cleaner, and choke. For the most part, the percentage of correct recognition was quite high -- above 80 percent for 20 of the 25 symbols. The five symbols which performed poorly included lighter, choke, master lighting switch, rear fog light, and front fog light. Wrong answers to the fog light symbols were due to confusion between "front" and "rear". At least 85 percent of the subjects knew that the symbol in question referred to fog lights.

The authors conclude that "the data obtained in this investigation demonstrate a high recognition value of 20 of the current 25 ISO symbols which, in general, is not influenced by sex or additional military driving experience" (Wiegand & Glumm, 1979, p. 14). Only two symbols -- "choke" and "master lighting switch" have recognition problems serious enough to suggest that they should be redesigned. As a result of this study, the authors proposed that the ISO symbol set should be used for automotive controls and displays, except for those individual ISO symbols which did not test well.

In a similar experiment, Green and Pew (1978) examined the effectiveness of 19 pictographic symbols used in automotive displays. They noted that previous comparisons of highway signs and symbols demonstrated that symbols could be recognized more accurately and rapidly. Similarly, symbols could be advantageously used for automotive displays and controls to reduce the time required to select a specific control and minimize confusion among alternative choices. In a test of this hypothesis, Green and Pew studied the effectiveness of automotive control symbols for a variety of responses. Fifty subjects, all licensed drivers aged 17-25, were employed in a series of five tasks. Nineteen pictographic symbols, intended for use in automotive displays, were selected based on performance on these tasks.

The first task was to determine the subject's familiarity with the symbols. Subjects were given copies of the symbols and asked to circle those that they were "reasonably sure" that they had seen before. Secondly, the authors determined "association norms" for each symbol. In this task, subjects were given copies of the symbols, read a driving scenario similar to that used by Heard (1974), and asked to point to the symbol appropriate for the scenario. The accuracy of the response was recorded.

In the third task, subjects made magnitude estimations of the communicativeness of each symbol. The authors defined "communicativeness" as "how well a symbol suggested its designated norm". All subjects rated each symbol twice. In the fourth task, the relative difficulty of learning each symbol was assessed (using a paired associate task). In this task, subjects were shown cards with symbols printed on them and asked to provide a label for each. They were given the correct label immediately after each response. The task was repeated until the subject could label each symbol correctly. In the last task, reaction time was assessed. The experimenter read a label and then showed subjects a slide of a symbol. The time for subjects to respond "same" or "different" to the slide was recorded. All symbols were shown twice -- once with a correct (or "same") label and once with an incorrect (or "different") label.

Two groups of subjects were studied -- 30 participated for only one day, while 20 participated for three days. On the first day, all 50 subjects participated in the familiarity, paired associate, and rating tasks. On the second day, 20 subjects participated in the paired-associate and reactiontime tasks, while on the third day, these two tasks were repeated along with a final rating task.

The analysis of the results indicated that education (technical or nontechnical), road experience, and specific vehicle experience all affected a subject's symbol knowledge. Analysis of task 1, familiarity, indicated that most symbols were not familiar to the subjects. The mean number of symbols recognized was 2.6 (out of 19). On task 2, only 6 of the 19 symbols tested met the acceptance criteria of minimum 75 percent recognition and maximum 5 percent confusion, as set by Heard (1974). A large number of symbols were confused with each other. The third task, rated communicativeness, suggested that a number of symbols did not communicate effectively at all. During a final, informal interview, subjects provided a number of suggestions for modifying the symbols.

The second and third day's tasks showed that subjects were able to do the paired-associate learning task with increasing accuracy. There was an initially large number of errors (mean of 7.1 per subject) for the set of 19 symbols, although this was partly due to difficulty in learning the predetermined labels. Thus "coolant temperature" was termed "engine temperature" or "radiator temperature". By the second test, subjects made an average of only 1.3 errors and almost none after that. Finally, the reaction time results indicated that there was a pronounced learning effect and variation in a subject's ability to do the task rapidly.

An examination of correlations between the tasks indicated that familiarity was not correlated with the other measures, even though Cahill (1975) had suggested that it might be an important factor in determining performance. In addition, correlations of associative strength with reaction time were only barely significant. Green and Pew (1978, p. 112) commented that "to date, it has been tacitly assumed that association norms were always highly correlated with reaction time, the measure of interest while driving; and that it was only necessary to measure associative strength. These data do not support this assumption." The one measure that emerged as useful was that of "communicativeness" which was highly correlated with association norms and reaction time. The authors claimed that rated communicativeness could be effectively used as a measure of the utility of a symbol.

Green and Pew noted that sex and technical ability affect the ability to recognize a symbol at first glance, but do not appear to be related to the performance of other tasks. Although reaction time decreases with learning, it is affected by the discriminability of the symbol. The authors also pointed out that association norms for and confusions within a set of symbols may not follow the intuitive feelings of the experimenter. "It is therefore very important that the experimenter interview a sample of the user population to ascertain why symbols are mistaken and confused. Subjects usually have logical reasons for their responses" (Green & Pew, 1978, p. 113). Finally, the authors concluded that these data underline the urgent need to evaluate symbol alternatives empirically before symbols are adopted.

3.3.3 Automotive Symbol Production

Because previous research had shown variability in the understandability of symbols, Green (1979) explored ways to develop better symbols for automotive controls and displays. His concern was to develop symbols which would be meaningful to all people who drive cars. Green (1979, p. 4) suggested that "Among other things, a good pictographic symbol is meaningful, is discriminable from other members of the set, and is easily distinguished from its background, even when degraded by wear, dirt, or reduction in size. The most important characteristic, in general, is meaningfulness. Meaningfulness is a function both of being able to associate a previously unseen symbol with its intended message (initial identification) and of the ease with which that is done after having been told what the symbol represents (later recall)." As a result, Green used the production method in which a group of people sketched symbols for each of seven referents. Then, another group of subjects rated the meaningfulness of six or seven most frequently drawn symbols for each referent. (Brainard, et al, (1961), had used a similar technique for highway symbols.)

In the first phase, 43 subjects, aged 20-64, drew pictures for seven referents -- heater, air conditioner, fresh air vent, radio volume, radio tuning, tire pressure, and lamp failure. Three judges scored the drawings by giving them descriptive labels such as "fire" or "snowflake". The drawings were then grouped by label and the most frequently suggested drawings were used as stimuli for the second phase. Subjects also provided definitions for a proposed symbol for "diesel starting aid," and indicated the best verbal labels for "battery" and "malfunctioning electrical system." Green found that only one of the 43 subjects had any idea of the meaning of the proposed "diesel starting aid" symbol, and preferred the term "battery" over the alternatives suggested.

In the second phase, Green (1979) had 62 subjects, aged 17-64, estimate the informativeness of the newly drawn symbols. These subjects were given seven sheets on which the desired label (referent) appeared in the center surrounded by four to ten symbols produced during phase one. Subjects made magnitude estimations of the informativeness of each suggested symbol for each label. Although Green found that some subjects did not fully understand the process of magnitude estimation, the data analysis revealed that subjects were generally able to agree on one or more symbols for a given referent, and that these symbols often differed from those in current use. One symbol for each of "radiotuning", "heater", "air conditioning", and "airflow" was readily selected, while several symbols for "exterior lamp failure", "tire pressure", and "radio volume" were chosen as equally informative.

Green (1979, p. 7) concluded that "It is very important that a production study be the first data collection effort in a symbol program. No matter how well conceived and executed the research plan, it is doomed to failure if the candidates are poor." Secondly, the use of a magnitude estimation task of symbol informativeness can lead to the selection of a "best" symbol for a given referent in many cases. Green cautioned, however, that his findings are best used as a guide for selecting effective symbols. Green (1979, p. 7) concluded that "these experiments demonstrate the need to include the users in all phases of symbol development and evaluation. Opinions of the symbol designer often disagree with those of the users."

Green and Burgess (1980) expanded the production method to include the development of combined concepts. As in Green's 1979 study, the concern was with symbols for various automotive systems and functions. A total of 28 subjects drew pictures of seven automotive systems such as air, brake, or fuel; four system properties such as fluid level, pressure or temperature; and 20 of a possible 28 system-property combinations, such as oil pressure. Analysis of the drawings indicated some differences between those drawn by engineering and by non-engineering students. Engineering students tended to draw the operating mechanisms of the machinery while other students tended to take an "external" view, depicting either a driving action or the consequences of failure to act. It is not clear whether the non-engineering students fully understood the meaning of all the referents. If the verbal referent is not clear to the users, the symbol produced is unlikely to be very clear either.

Green and Burgess (1980) then had 26 students make magnitude estimations of the informativeness of 5 to 12 candidate symbols for each of 26 functions. These candidates included the student drawings along with manufacturer's suggestions and international standards. In general, these latter symbols were not rated as informative as the students' drawings. (The same set of subjects was used in both experiments.) As a result, the authors concluded that some of the ISO symbols for fuel, brake failure, engine oil, and coolant temperature should be reconsidered. They also found wide differences in the informativeness of some of the symbols--particularly the combined symbols. Their results underline the need to develop a syntax for combining symbolic messages to create coherent, unique meanings. Green and Burgess concluded, though, that the strength of the production technique is that it elicits innovative drawings from a driver's rather than a manufacturer's perspective.

3.3.4 Orientation of Symbols

Another issue, explored by Green and Davis (1976), is the effect of variations in the orientation of symbols for automobile controls. Previous research, such as that conducted by Heard (1974), evaluated the recognizability of automotive symbols placed in an upright position only. Yet symbols placed upon automotive controls are often rotated away from upright, and, consequently, may not be rapidly or accurately identified. As a result, Green and Davis' study concentrated on the time required to recognize a rotated symbol.

Ten subjects studied three different symbols which had been rotated into different positions. These symbols were: parking light, windshield wiper, and lower beam. The subjects were presented with numerous pairs of symbols, of which one symbol varied in orientation. In addition, half of the varied symbols were reversed (mirror image). Subjects had to rotate (mentally) the left-hand symbol of the pair to agree with the right-hand symbol and then decide if it were the same or different (mirror image). They responded by writing S (for same) or D (for different) next to each pair. Each subject judged 864 symbol pairs in blocks of 24 pairs for a given symbol. The time to complete each block was recorded.

Analysis of the results indicated that increasing the rotation of the symbol away from upright significantly affected response time. Green and Davis claimed that their data suggest that mental rotation of symbols occurs at a constant rate which averages about 130 degrees per second for this experiment. There were no significant differences between symbols. Green and Davis (1976, p. 183) concluded that the data obtained for rotated symbols indicate the potential for a serious hazard. "If a symbol were to appear in a random orientation, a reasonable assumption for a rotatable control, then it would be on the average 90 degrees from upright. At freeway speeds, one would travel roughly three carlengths in the time it would take a driver to perform just the rotational transformation." Because a driver could have difficulty responding to an emergency during this time, Green and Davis suggest that symbols used for controls always be mounted in an upright position.

3.4 PUBLIC INFORMATION SYMBOLS FOR BUILDINGS

The third area of application to be covered in this report is that of public information symbols. Typically, investigations in this area have centered on the understandability of one or more sets of symbols designed to convey information to travelers. As with automotive symbols, no comparisons have been made of the relative effectiveness of word and symbol signs, probably because the intended audience is comprised of both English and non-English speaking persons. Because questions such as reaction time or speed of detection appear much less relevant for public information symbols, accuracy of understanding has been the primary variable studied.

Easterby and Zwaga (1976) compared the understandability of several different symbols for each of six referents in a cross-cultural evaluation. In a three-part experiment, they evaluated public information symbols with subjects from seven countries under the sponsorship of ISO.

In the first portion of the experiment, groups of subjects in two countries rank-ordered numerous symbols for six referents in terms of their appropriateness for a given referent. These referents indicated the location of drinking water, information, stairs, taxi, toilets, and waiting room. From these rankings, three symbols for each of the six referents were selected for further study. Figure 11 presents these symbols.

In the second phase, 25 subjects in two age groups provided definitions for each of the three sets of six test symbols. Context was provided by an additional 18 symbols. Analysis of the results indicated that the wrong answers fell into two categories: incorrect definitions and "don't know" answers. Inspection of the incorrect definitions provided valuable insight into some of the design elements that affect understanding. As a result, Easterby and Zwaga claimed that more representational symbols were more easily recognized than highly abstract ones.

In the third phase, subjects selected one symbol from a much larger set to match against a given referent. In all, three versions for the six referents were tested, although each subject matched only one symbol. A total of 900 subjects in six countries participated. Although the authors were able to obtain percentages of correct responses for each symbol, they commented that these percentages can be interpreted only in terms of the symbol set tested. In some cases, all responses to a given referent were incorrect. While the matching data do not consequently indicate the best symbol for a referent, they do provide information on the confusions generated between symbols within a set. As a result, "the confusion data generated gave valuable information, both on the image content of the symbols under test and of the other symbols present in the test material. It would be unwise then to base any decisions on the image content of the test symbols on the data from the matching test alone, since percentage correct responses are misleading. The correct response data must be interpreted both in relation to the confusions generated, and the features of those other symbols used in the test material and any final decisions on image content are as likely to be based on the recognition test results as the matching test results" (Easterby & Zwaga, 1976, p. 45).

Because the matching test reveals little about a symbol in absolute terms, and because the recognition test indicates specific confusions, Easterby and Zwaga recommended that the recognition test be conducted as the critical phase to determine the general image content of a symbol. Furthermore, ... "it is now recommended that the recognition and matching test become successive rather than parallel parts of the program, with the matching test being used to evaluate a complete set of symbols which have been integrally designed as a set. The particular features of the matching test can then be used to evaluate confusions between symbols which have been designed with an image content determined by the results of previously conducted recognition tests" (Easterby & Zwaga, 1976, p. 49). As a result of this research, Easterby and Zwaga recommended that the matching test follow the recognition test to evaluate the interdependence between symbols in a complete set.

The other major study of public information signs was commissioned by the U.S. Department of Transportation (DoT). DoT first sponsored the design of 34 public information symbols by the American Institute of Graphic Artists (AIGA). The AIGA (1974) compiled a list of existing symbols for each referent and then selected or modified the best symbol for each idea based upon this compilation.

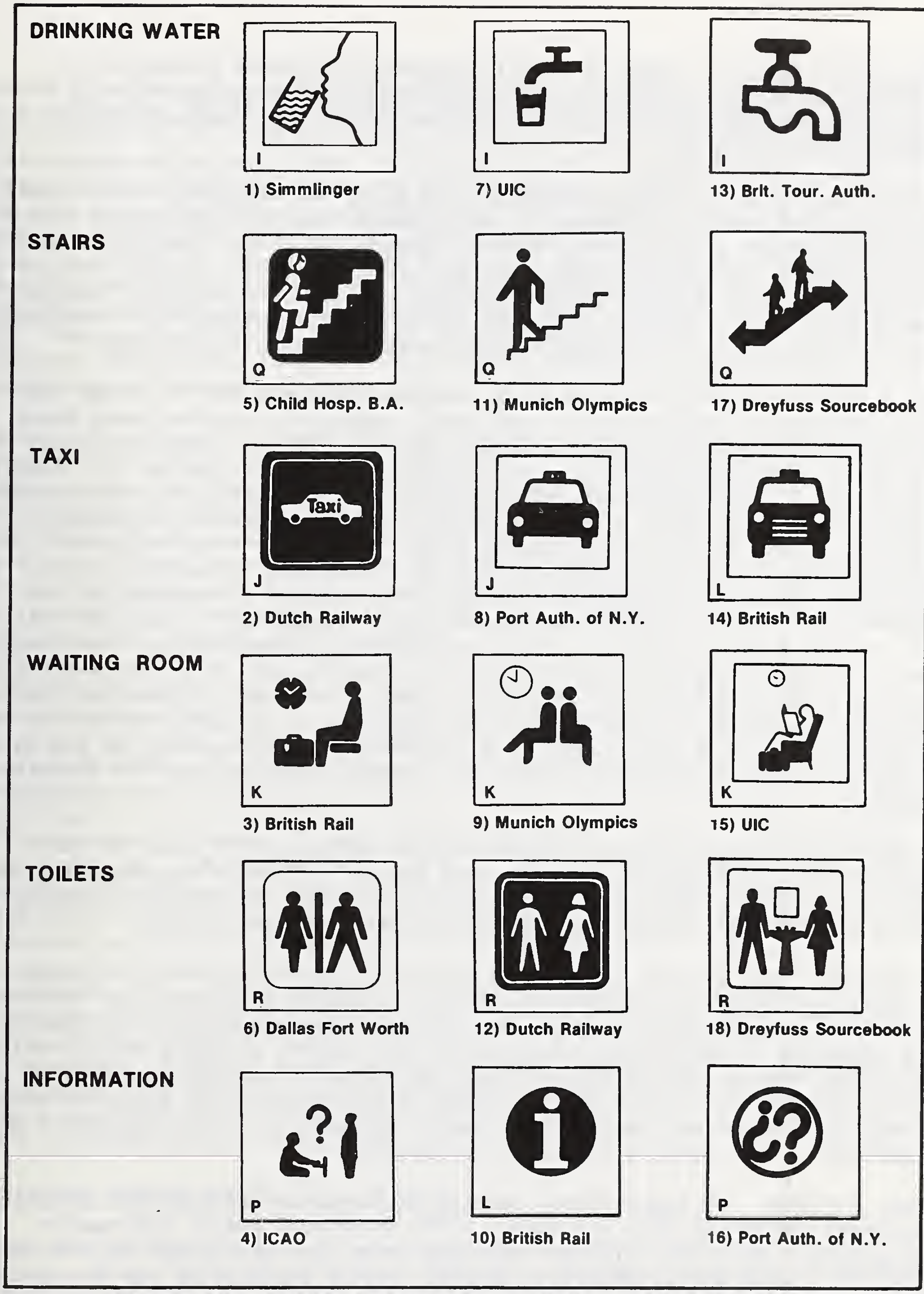


Figure 11. Public information symbols studied for ISO

(See Section 2.3.2). Next the Franklin Research Institute conducted an extensive evaluation project to determine the understandability of each of the initial 34 symbols. Preliminary reports of their procedures will be described briefly here.

In the first interim report, Freedman, Berkowitz, and Gallagher (1976) used both paper-and-pencil and behavioral performance tests to assesss the 34 symbols. They began with a series of tests intended to assess the recognizability of the symbols, elicit confusions, and provide an indicator of the relative difficulty of the symbols. These tests provided input to the major testing phase and indicated, as well, that the symbols varied somewhat in recognizability. Eleven symbols, for example, were understood by all subjects.

Both a multiple choice and a matching test were designed for the large scale testing phase. In the multiple choice test, subjects were shown each symbol and asked to choose one of four possible meanings for it. These "meanings" were based upon correct and incorrect answers elicited during the preliminary testing. In the matching test, subjects selected the most appropriate symbol for a given referent from a larger display of symbols, ranked the semantic strength of each symbol, and indicated their familiarity with the symbol. The final portion of the large-scale testing phase included a "walking rally." In this rally test, the various destinations were depicted by exhibits in the Science Museum of the Franklin Institute. All conventional directional signs were removed and the new destinations indicated by placard signs. The time to reach each destination was recorded with a stopwatch held by a test monitor who accompanied each subject. Correctness of the destination was monitored along with time to arrive. All 34 symbols were assessed in two split halves for the paper-and-pencil test while only 17 of the symbols were assessed for the walking rally. Field tests at an airport and a subway station used the matching procedure.

In the second interim report, Freedman and Berkowitz (1977) reported that analysis of the procedures indicated that the multiple-item matching test maximized reliability while the performance or rally test ensured test validity. Scores on both tests were correlated and provided similar results.

Preliminary comparison of the lab and field data indicated few differences in the subjects' responses (Freedman, 1978). The author concluded that the characteristic responses of both groups were similar. Because a number of symbols were missed by a large number of subjects, the authors proposed several criteria for effectiveness. They suggested that symbols which are recognized by 60 percent or fewer of the subjects are clearly unacceptable, while those recognized by more than 85 percent are equally clearly acceptable. Those recognized by 60-85 percent of the subjects need some improvement.

Caron, Jamieson, and Dewar (1980) applied the semantic differential technique (discussed earlier by Ells and Dewar, 1979) to several public information pictographs. Sixty-two volunteer subjects rated each pictograph on a series of 35 bipolar, seven-point, adjective scales. Factor analysis of the pictographs determined that they represented the dimensions of evaluation, activity, and potency. In a second experiment, 96 students made ratings of the 19 adjective scales defined as the relevant dimensions in the previous experiment. One group of subjects rated the six pictographs, while the other group rated a set of six corresponding word signs. Comparison of the ratings for each component/pictograph revealed that four of the six ideas were similarly related. Two which were dissimilar were "restaurant" and "women's room." The authors theorize that because the "women's room" pictograph does not depict the function actually intended, it is only successful in practice due to the context of juxtaposed male/female signs. The "knife and fork" symbol for restaurant "does not communicate well, even using recognition measures" (Caron et al., 1980, p. 142). The authors note that semantic differential responses to pictographic and verbal signs can be obtained from different groups of subjects at different times/places and subsequently compared.

Experiment 3 expanded the technique of experiment 2 by comparing responses to two components of each of the six pictographs with the responses given previously to the lettered signs. Two groups of students (N = 96) evaluated the pictograph component, using the 19 adjective scales. In the most part, evaluations of the pictograph components were quite dissimilar to those for the verbal signs. In most cases the component alone was less successful than the complete pictograph in communicating a particular message. In one case, however, the "telephone receiver" component was more successful in conveying its message than the entire "telephone" pictograph.

The authors concluded that their results demonstrate the utility of the semantic differential tool in evaluating pictographs or symbols with different subjects at different times and places. "More importantly for sign evaluation, the classification analysis technique allows rapid evaluation of meaning matches between pictographs and lettered signs. The fact that this analysis uses <u>distinct</u> groups of subjects to evaluate pictographs and lettered signs is particularly important in practice; once classification equations have been developed for a target set of lettered signs, evaluation of corresponding pictographs may proceed at any future time with new samples of observers and pictographs. This result allows continuous evaluation of new sign design and convergence on optimal designs in the most parsimonious fashion" (Caron et al., 1980, p. 145).

The semantic differential evaluation may be most effective when combined with a direct assessment of meaningfulness. The semantic differential technique is most useful for measuring how meaningful a stimulus is; it does not provide data on whether a subject can accurately define a symbol. Once a baseline of understandability has been determined, however, semantic ratings and subsequent factor analysis appear to be a most useful tool for predicting and comparing symbol performance along a continuum.

3.5 HAZARD WARNING SYMBOLS FOR PRODUCTS

Another application of symbols covered in this review is that of product labeling. While there are several standards in this area proposed by the Common Market, the British, and the Canadians, there has been little if any research on the effectiveness of symbols for product warnings. In the U.S. there are a number of instances in which controversial product labeling symbols have been produced. These include "Mr. Yuk", (see figure 1) produced by the Pittsburgh Poison Control Center, to be used instead of the skull-and-crossbones to warn children of poisonous substances, and the lawnmower and CB antenna symbols developed by the Consumer Product Safety Commission (CPSC) to warn consumers. While these have caused much discussion, only "Mr. Yuk" has received any evaluation at all, and that only on a limited basis.

In England, however, Easterby and Hakiel (1977a, 1977b, 1977c) conducted a large-scale evaluation of product hazard symbols to be used in the United Kingdom (U.K.). They (1977b) pointed out that safety signs for product labels should improve the processes of discriminating between products, identifying hazardous products, attending to a conspicuous product, searching for and locating a particular product, classifying a product, and understanding the meaning of a product sign. Noting that formal coding systems are typically used to provide additional information, they raised the following questions (1977b, p. 4) -- "What are the psychological consequences of these formally developed shape and color codes? Do observers perceive any structure in such formal codes? Do these formal coding features affect the comprehension of the sign? Are the codes dealt with separately or is the sign perceived integrally -- as a whole sign? Are there stereotypes which uniquely associate specific shapes and colors with specific types of messages?" These authors believed (1977b, p. 4) that" some stereotyping does exist in relation to the shape and color coding of image, background, enclosure and surround elements of safety signs." If such stereotyping does exist, then when subjects generate their own signs, certain elements should appear more frequently than others. Easterby and Hakiel proceeded to test this hypothesis in a series of laboratory and field studies.

In the first study, Easterby and Hakiel (1977a, 1977b) asked subjects to design signs to represent fire, poison, and caustic hazard information. Subjects were asked to construct signs from a selection of image forms and colors, background colors and shapes, enclosure shapes and colors, surround shapes and colors, and supporting field colors. Three alternative images were provided for each of the three hazards. One described the nature of the hazard, one provided information for avoiding the hazard (prescriptive), and the third depicted a prohibited action associated with the hazard. Four shapes were provided as background, enclosure, and surround -- square, circle, rectangle, and triangle. The seven colors provided were red, orange, yellow, green, blue, black, and white. After receiving these materials, a subject designed a sign to be useful in indicating a hazard on a typical household product. One week later, the procedure was repeated with the constraint that signs had to be produced for each of the three symbol modes -- descriptive, prescriptive, and prohibitory.

Easterby and Hakiel (1977b) repeated these procedures in three experiments -- a pilot study, a field study, and a laboratory study. The results from the field study demonstrated that shape and color preference were about the same as for the pilot study. Problems with background color did arise in the pilot study, which led to some modifications in the stimulus presentation in the field study. These changes led to a slight reduction in the use of white backgrounds and an increase in yellow and red backgrounds in the field study. The general result from the three studies demonstrated that red was the preferred color for fire signs. For poison, black was preferred to red, while for caustic, red and black were about equally used. (There was also a preference for a triangle as the primary enclosure shape. Background color seemed to be chosen to maximize contrast, while enclosure color tended to correspond to image color). The authors commented (p. 35) that "It is evident, therefore, that not only is there hazard specificity for the preferred value of the image colour -- the colour stereotype -- but the strength of the stereotype is also hazard dependent." Stereotypes of triangles and circles accounted for 60-75 percent of the background shapes. In all of this, fire showed the strongest set of stereotypes and caustic the weakest.

The authors noted that subjects did tend to choose specific patterns of colors for images and enclosures. Easterby and Hakiel concluded (1977b, p. 42) "that the coding of signs is perceived in terms of color combinations, with the comparatively independent addition of shape. The colour combinations are used to specify the identity of the hazard, reinforcing the function of the image." They commented that the use of color should probably be confined to indicating the nature of the hazard. The stereotypes identified in these experiments were used to construct signs for a subsequent large-scale study of the understandability of specific warning signs.

In the final study of this series, Easterby and Hakiel (1977c) conducted both a pilot study and a large-scale survey of the effectiveness of the shape and color stereotypes developed in the previous study, along with a variety of proposed symbol images. In the pilot study they asked 38 students to rank order a set of symbols in terms of how well each conveyed a particular message. This procedure was used to reduce the large number of existing symbols to a more manageable number. From these results, a set of four symbols for each of three hazards -- Fire, Poison, Caustic -- was selected. Subjects preferred descriptive symbols to either prescriptive or proscriptive symbols. When there were several versions of a somewhat similar image, subjects preferred more visually complex images to graphically simplified images.

The symbols selected for study in the pilot test were then used in a nationwide (U.K.) survey of 4000 respondents. In the recognition test, subjects provided a meaning for each of 17 signs. The set of 17 signs contained five test signs and 12 contextual signs which might be found on consumer goods or in public environments. The five test signs included one example of each of the three referents identified earlier, as well as one example each of electrical and general hazard signs. Subjects were given a booklet which contained five context signs, followed by the five hazard signs, with each variant of each hazard sign being presented to 500 people. The interviewer recorded each subject's responses and supplied a rating of his/her confidence in each response. In addition various demographic data such as age, sex, occupation, and family composition, etc., were collected. During the data analysis, frequencies of response were tabulated for each symbol and criterion categories of ranked effectiveness were developed.

Easterby and Hakiel concluded that attributes of the sign, in terms of image, color-coding and shape coding, and characteristics of the respondents, such as age, sex, household composition, and experience with signs, influenced the recognizability of the signs.

The two characteristics of the image which most influenced recognition performance were communicative mode and graphic quality. For the most part, descriptive imagery was effective along with the use of more visually complex graphic renderings. Graphic simplification or abstraction seemed to lessen the perceived appropriateness of the symbols in the ranking study, although the authors suggested that "actual recognition performance will only deteriorate if simplification is more extreme than the examples used in this study" (Easterby & Hakiel, 1977c, p. 95).

The effects of color were related to the meaningfulness of the sign. "It was found that for those signs which performed best, for a given hazard, colour had no influence on performance -- the differentially colour-coded variants of the best performing fire, poison and caustic signs gave statistically indiscriminable recognition performances. When overall recognition performance was poor, however, differences between different coloured variants of otherwise equivalent signs became apparent . . . when the image component of the sign is not readily recognizable to the respondent, more attention is given to those other sign attributes which may enable a better interpretation of the sign" (Easterby & Hakiel, 1977c, p. 95).

The characteristics of the subject sample did not affect the relative recognition performance for a sign, but did alter the absolute levels of recognition. Previous familiarity with the symbol improved performance. Sex had a differential effect, with housewives failing to recognize the electrical symbol, except that those with young children at home showed better recognition of the overall symbol set. The most consistent effect, however, was the poor recognition performance of those over 55. These data suggest that those who are older or who do not have young children could profit from a more intensive educational program. Easterby and Hakiel (1977c, p. 100) concluded that "the shape and form of the images, independently of colour, is the factor which primarily influences recognition performance of the sign. We conclude, therefore, that so long as the sign satisfies some minimum requirements of legibility in term of figure-ground contrast, its recognizability depends on the image used, independently of the colouring of the components."

3.6 SAFETY SYMBOLS FOR BUILDINGS

Although there do not appear to be many studies which have assessed the effectiveness of safety symbols for buildings, at least three studies have examined some aspects of safety signs. Collins and Pierman (1979) and Lerner and Collins (1980) evaluated the meaningfulness of fire safety symbols. Laner and Sell (1960) determined the behavioral effectiveness of safety posters. Although Laner and Sell did not assess the use of individual symbols, their work is of interest because it measured the effectiveness of safety signs directly in terms of changes in unsafe behaviors. A review of their methodology can provide some guidance for evaluating the effectiveness of safety symbols using behavioral measures. Collins and Pierman (1979) reported an experiment in which they determined the understandability of 22 fire safety symbols. They asked 143 subjects to provide a short definition for each symbol. Three judges rated each answer as "correct", "incorrect", or "no response". In addition a tally was kept of the number and kind of incorrect answers. The percentage of subjects responding in each of the three ways was calculated for each symbol.

The authors found that some symbols such as "telephone", "no smoking" and the conventional U.S. "exit" sign were understood by almost all the subjects tested. Yet other symbols such as "blind alley", "do not block", and "break glass" were understood by only 20-25 percent of the subjects. Not only were some symbols not understood, several symbols were given a meaning opposite to that which was intended. Thus, "no exit" or "blind alley" was interpreted as "exit" or "safe area" by almost all subjects who gave a definition for this symbol. Altogether, over 95 percent of the subjects either misidentified or did not respond to this particular symbol.

Collins and Pierman (1979) commented that an instance in which a symbol is given a meaning opposite to that which is intended is potentially very dangerous. They recommended that before symbols are adopted, particularly those which communicate emergency information, their effectiveness must be evaluated. A safety symbol must be understandable before it can begin to alter behavior and prevent accidents. Figure 12 presents selected symbols which demonstrate the extremes of understandability found in this study.

Lerner and Collins (1980) continued the assessment of fire safety symbols in an evaluation of various testing methods. Using the same set of symbols, with the addition of three new exit symbols, they compared the use of multiple-choice versus definition-style answers, as well as the use of slides, placards, or booklets as stimulus material.

A set of 91 subjects in three age groups participated in the experiment. Analysis of variance of the data indicated that there was no significant difference for stimulus presentation method. While the analysis indicated that there were slight differences between response methods depending upon how strictly the definition answers were scored, the ordering of symbols according to understandability remained about the same for the two methods. The multiple-choice method tended to overestimate the understandability of poorly understood symbols. As a check on the multiple choice method, the authors had subjects make confidence ratings of each of the four multiple-choice answers. The confidence ratings provided insight into the extent of guessing and confusions. In some instances a wrong answer was given a higher confidence rating than the correct answer. Use of multiple choice answers appears to

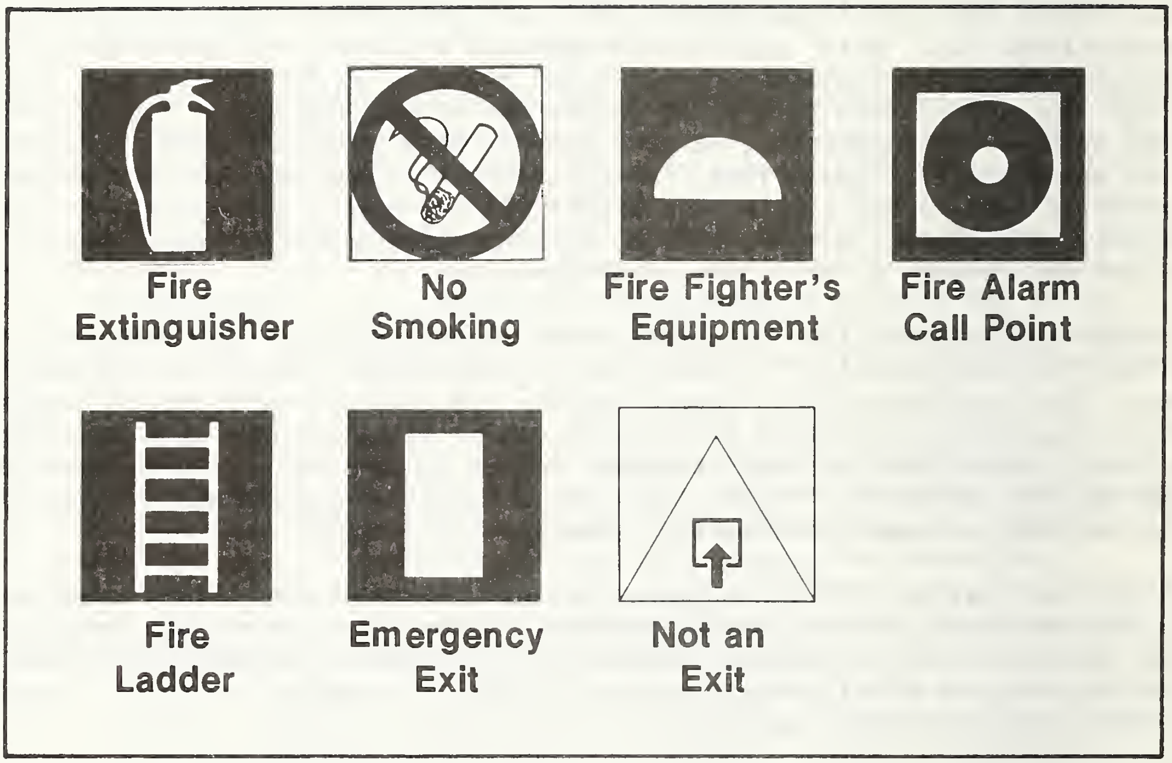


Figure 12. Examples of fire safety symbols

be a viable option if confidence ratings are included as a check in guessing, and if "wrong" answers are selected from those given in a previously administered definition experiment.

Lerner and Collins also obtained production data (drawings) for fifteen safety referents. In seven cases, the majority of the drawings were essentially the same as the ISO image. In eight cases the images were different, with no consistent majority image emerging in three cases--"exit," "blind alley," and "general hazard."

As in the Collins and Pierman study, certain critical life-safety symbols were poorly understood. The newly proposed ISO "exit" and "no exit" symbols were better understood than the former ISO symbols, but still did not perform nearly as well as symbols for "no smoking" or "fire extinguisher." In addition, the production data revealed that in many instances subjects were not able to agree upon symbolic representations for the poorly understood concepts. In particular, despite instructions not to use words, most subjects insisted upon the word "exit" in drawings for this concept. Lerner and Collins (1980) concluded that the use of both understandability and production methods provides good insight into the effectiveness of a given symbol concept.

Bresnahan and Bryk (1975) reported a study in which the hazard association values of standard ANSI word safety signs were measured. Using 64 industrial subjects, they measured the degree of hazard associated with the signal words - danger, caution, think, and notice - and the colors - red, yellow, green, and blue. They found that "danger" and "caution" were considered representative of greater hazard values than "think" or "notice". Red and yellow colors also represented a higher level of hazard. While the use of color and signal word together elicited the most extreme values, it is interesting to note that the use of color alone (with no signal word) elicited values that were almost as strong. This study did not assess the effects of symbolic imagery and color upon perceived hazard, although it suggests that sign color can play a role in communicating the level of hazard. Whether this is due to learned associations with familiar ANSI (American National Standards Institute) signs, experimental artifact, or similar influence is not known.

Laner and Sell (1960) examined the effectiveness of safety messages in altering unsafe behavior. They hypothesized that the temporary creation of an unsafe situation -- a common cause of accidents in industry -- can often be avoided by altering human behavior. Although safety posters with various sorts of warning messages are widely used in an attempt to stop unsafe acts, their effectiveness in actually modifying these behaviors has rarely been assessed. Effectiveness could be measured by studying accident reduction directly, except that the time between accidents is so great that the experiment would be inordinately long. Laner and Sell also rejected the idea of measuring poster effectiveness in terms of the extent to which a poster could be recognized, remembered, or liked, because these measures do not necessarily predict behavior.

Consequently, Laner and Sell (1960) selected a behavioral measure which would involve an operation that was potentially dangerous, frequently carried out, and readily measurable -- namely, the hooking back of chain slings onto a crane hook when not in use. Seven steelworks participated in the experiment in which posters depicting safe steelworking practices were developed and displayed. First a baseline of behavior was established over 5 weeks, followed by a lapse of 7 weeks without measurement, concluding in 2 additional weeks of measurement.

Laner and Sell found the posters had a positive effect in the six test steelworks, with a substantial effect in four of the six works and no effect in the seventh, or control steelwork. Furthermore, they noted that the behavior affected by the posters was at least maintained, if not improved, following the 7 week period in which behavior was not measured. The authors suggested that the posters were effective because they acted as perpetual reminders or because they established and reinforced working habits which were selfmaintained. They also found that the greatest increase in safe behavior occurred where the greatest hazard to personnel existed. They concluded that posters may be more effective if the message they carry can be seen to be directly relevant to the situation. This study confirms that the use of a behavioral measure -- reduction of unsafe acts -- is one of the most potent measures of a sign or symbol's true effectiveness. The research approach suggests a means of assessing the most critical question in safety communication; namely. "Given that written and symbolic messages can both be understood, which are more effective at inducing people to act in accordance with the warnings?" (Dorris & Purswell, 1977, p. 345). Behavioral assessment is the most accurate and probably the most difficult means of answering this question.

3.7 CATEGORIZATION OF PREVIOUS RESEARCH

The preceding review of the research literature has indicated that symbols can be more effective than words under many circumstances. Symbols can be understood more rapidly and accurately than words. Because the effectiveness of symbols depends heavily upon their meaningfulness, most researchers have concentrated upon the assessment of meaningfulness. Yet, other psychological processes have been investigated as well. Table 2 categorizes the symbol research reviewed in section 3.0 according to the psychological process assessed and the symbol application addressed. The six aspects of psychological information processing activities identified in 3.1 are presented: detection, discrimination, recognition, understanding, attention, and behavior. Although there is a logical sequence or progression to the psychological processes on paper, there are many interactions and feedback loops among them which do not appear in table 2. Five applications of graphic symbols are included: highway, automotive, public information, product labeling, and general safety.

Table 2 also includes the response methods which have been used to assess symbol performance for each process. These include the use of reaction time, understandability measures, discrimination measures, production techniques, and behavioral indices. Although a complete evaluation of the effectiveness of a symbol should include an assessment of its performance for each process, table 2 demonstrates

Table 2. Symbol Research: Psychological Process and Application

PROCESS	HIGHWAY	AUTOMOTIVE/MACHINERY CONTROLS	PUBLIC INFORMATION	PRODUCT LABELING	SAFETY
	APPLICATION				
DETECTION					
glance legibility	King, 1971, 1975 Dewar, 1976				
DISCRIMINATION					
direction identification	Smith & Weir, 1978	Green & Davis, 1976			
RECOGNITION					
(match symbols to referent)	Plummer et al., 1974 Ells & Dewar, 1979 Griffith & Actkinson, 1977, 1978 Brainard et al., 1961	Wiegand & Glumm, 1979 Green & Pew, 1978	Easterby & Zwaga, 1976 Freedman et al., 1976		
UNDERSTANDABILITY					
definition	Brainard et al., 1961 Walker et al., 1965 King, 1975 Dewar & Swanson, 1972 Dewar & Ells, 1976	Cahill, 1975, 1976		Easterby & Hakiel, 1977	Collins & Pierman, 1979
magnitude estimation	Dewar & Ells, 1977	Green & Pew, 1978 Green, 1979			
rank order	Smith & Weir, 1978		Easterby & Zwaga, 1976	Easterby & Hakiel, 1977	
multiple choice					Lerner & Collins, 1980
ATTENTION					
contrast	Smith & Weir, 1978				
BEHAVIOR					
sketch	Brainard et al., 1961	Green, 1979			
touch, push	Janda & Volk, 1934	Heard, 1974			
movement	Forbes et al., 1963		Freedman et al., 1976		

that no one researcher to date has evaluated a symbol's performance for all processes for any application. Table 2 also demonstrates clearly that the bulk of the research has concentrated upon assessing the understandability of symbols for highway and vehicle applications.

3.8 CONCERNS FOR FUTURE RESEARCH

The concentration of research on understandability, noted in table 2, reflects a certain sense of priority, in that determination of understandability is a necessary first step in symbol evaluation. Once a symbol has been found to be understandable, then its effectiveness at other stages of processing can be determined. If a symbol is not understandable, it makes little sense to evaluate its detectability, discriminability, recognizability or the like. Improving its performance on these dimensions is not likely to improve its understandability. Yet, if a symbol is not detectable, discriminable, or the like, it will also not be understood. Consequently, all aspects of information processing and behavior must be assessed to determine a symbol's total effectiveness. The importance of these processes may vary with the application, however.

Although the processes of detection and discrimination may appear to be most applicable to highway situations where speed is a factor, they also enter into building applications. Both smoke and dirt or dust may obscure a building symbol, causing it to be less detectable than necessary. Thus, issues such as color, shape, legibility, and figure-ground contrast must be assessed for symbols whose detection is critical during a building or personal emergency. Determining the effect of symbol characteristics such as figure-ground relationship, complexity, and abstraction, upon symbol detectability is second only to determining a symbol's meaningfulness.

Determining discriminability, or the ease with which one symbol in a set can be distinguished from another, is another critical research concern. Although discriminability usually refers to the process of differentiating between similar visual stimuli, it is also related to a cognitive discrimination process which occurs when similar images are given different meanings in different applications. For example, the ISO "break glass for access" symbol resembles the DoT explosion hazard symbol (see figure 13). The user is forced to decide whether the two very similar symbols are intended to convey the same idea, or have very different meanings. Although context can often provide valuable clues, these confusions point out the need to consider all existing symbol sets when selecting a symbol for a particular referent.

A related issue concerns those situations in which a user must decide whether symbols which look different in fact convey the same meaning. (This issue is compounded by the current proliferation of numerous symbols for a particular referent.) For example, there are currently a large number of symbols used for the concept of "no smoking." The user must decide if these differences are meaningful -- if he/she should discriminate between "no cigarette smoking" and "no cigarette, cigar or pipe smoking" or simply not smoke anything and furthermore not light a match either. Finally, the same symbol may have different

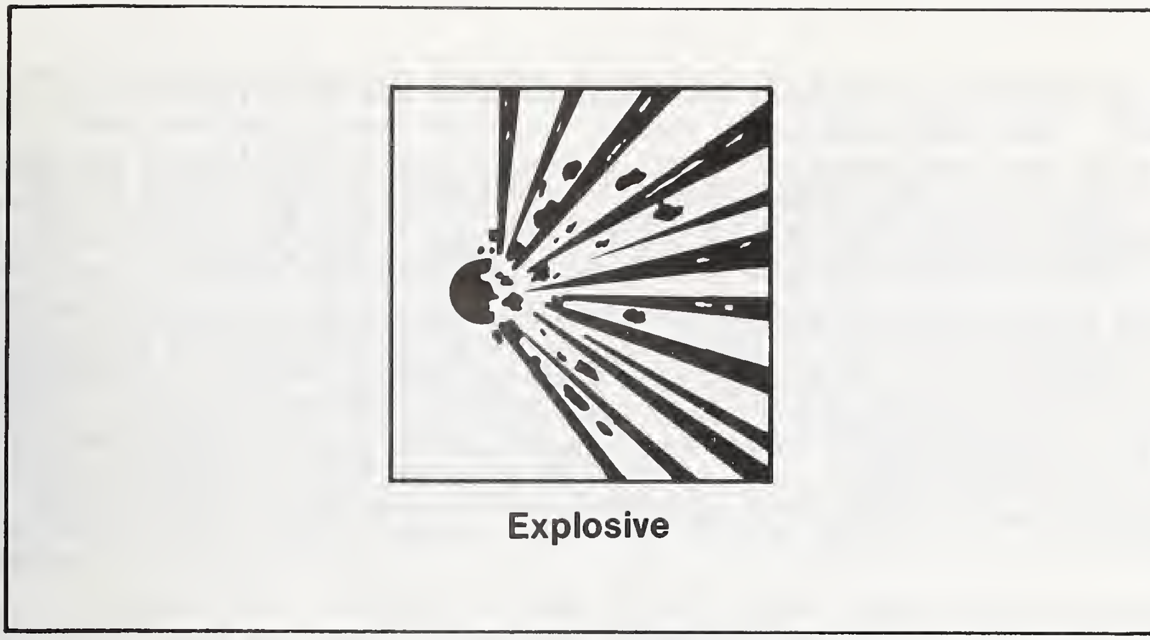


Figure 13. Examples of potentially confusing symbols.

meanings in different contexts with location being the primary cue to its meaning. Thus, the "handicapped access" symbol is used to indicate both access for the handicapped and denial of use for the non-handicapped for parking places but not restrooms. These concerns reach far beyond those of discrimination and detection and have tremendous impact upon all research into the effectiveness of symbols as elements of communication systems.

Another related research concern is the need to determine salience, or the ability of the symbol to attract the user's attention. Clearly, a symbol must command attention -- if its message is worth displaying. Methods for assessing salience or attention-getting, while used extensively by market researchers for product displays (Doblin, 1979), have rarely been applied to symbols for public safety or transportation. Yet the topic is an important one, particularly in situations where background visual clutter can intrude upon and obscure the symbol.

Behavioral effectiveness remains one of the most critical, but largely unresearched areas for assessing symbol excellence. Does the symbol alter the user's behavior in the intended direction? Will he/she follow its message? Word messages must be compared with symbolic messages to see which are more effective in getting people to behave appropriately. In addition, different symbols for a given referent should be compared for their relative effectiveness in determining behavior. Nowhere is the issue more important than in the safety area, where failure to understand a symbol or word message such as "no smoking", "flammable", or "explosive" could lead to a serious injury or disaster such as fire or explosion. Similarly, failure to obey signs for personal protective equipment can lead to injury or death. Thus, the symbol must communicate the seriousness of the hazard situation as well as the nature of the hazard, and direct the user's behavior accordingly. Despite its importance, behavioral effectiveness has rarely been evaluated as systematically as meaningfulness or detectability. As a result, there is a need to develop observational measures and indices which can be as easily studied as Laner and Sell's "hooking back of crane slings," and which provide a similar indication of the effectiveness of the warning message.

Although more extensively studied than processes such as detection, discrimination, salience, and behavioral effectiveness, understandability remains critical. For example, if the symbol tests well in terms of detectability and discriminability but is not meaningful, and no more understandable symbol has been developed, then issues such as ease of learning and retention must be faced. The symbol should then be evaluated in terms of how long it takes to learn it, and how well it is remembered after training. Of the studies reviewed in section 3.0, only Walker, et al., (1965), and Griffith and Actkinson (1979), assessed the ease of learning and length of retention for a set of symbols. The disparity between their results and the problems uncovered by Griffith and Actkinson, while most likely due to extreme differences in the size of the two symbol sets (6 versus 128), indicates the need to develop more effective training methods for learning and remembering symbols. At this point, training typically consists of the hope that the users will eventually learn the meaning of the symbol through repeated association of a word label with the symbol.

In addition to the failure to evaluate performance for all six psychological processes for each application, at least one application has received only limited research attention of any sort. As can be seen from table 2, safety symbols for buildings have been evaluated only in the most limited fashion. Performance for the processes of detection, discrimination, recognition, and attention has not yet been evaluated for any safety symbol set, while that for understandability has only been evaluated for a limited set of fire safety symbols. The only assessment of behavior alteration concentrated upon posters rather than symbols. As a result, the particular application where the use of a poor or misleading symbol could be potentially quite hazardous has received very little research attention. Table 2 makes it clear that the effectiveness of building safety symbols must be researched for all psychological processes, including understandability.

Facing page:

A variety of symbols proposed for safety and directional information

4. SUMMARY

4.1 OVERVIEW OF THE LITERATURE

The need to communicate information without the use of written language has prompted an international movement toward the use of symbolic signs. Graphic symbols now appear in a wide variety of applications, from highway signs to motor vehicles, for a variety of purposes from directional information to hazard warnings. The need for symbols has become more acute with growing international travel, high illiteracy rates, and legal mandates to communicate hazard warnings. This need has generated a concomitant need for standardization.

As noted by Eliot (1960), the earliest modern attempts to use graphic symbols were for highway applications, with international standardization beginning as early as 1909. King (1971) described several international conventions which culminated in a U.N. protocol on traffic signs in 1949. Many of these symbols have finally been adopted for use in the United States. Nevertheless, highways are one of the very few applications for which even a few symbols have been successfully standardized on an international basis.

The process of successful standardization for all applications is heavily dependent upon the testing and evaluation of graphic symbols. Ideally, testing and evaluation should occur at several stages during the process of developing and standardizing symbols and symbol referents. Once selections of the referents are made, the development of symbols should focus on the nature of the graphic symbol itself and the message it is to convey. Symbols can describe a hazard, prohibit an action, or prescribe a remedy. They can indicate the nature of the hazard, the consequences of interacting with a hazard, or the degree of hazardousness. Once a conclusion about the message to be conveyed is reached, then the designer can concentrate on the nature of the symbol and the degree of abstraction. Both Modley (1966) and Eliot (1960) discussed symbol characteristics in terms of the pictorial quality of the image. While a directly pictorial image may communicate its intended meaning rapidly and accurately, such images are difficult to achieve for some concepts, and with rigid standardization may in time appear dated and outmoded. Furthermore some concepts such as "stop", "exit", or "route" can be very difficult to portray pictorially. As a result, abstract symbols have been proposed for such concepts. Because an abstract symbol may also be clearer graphically and more detectable under degraded viewing conditions, the eventual application of the graphic symbol must also be considered during the design and evaluation proccess.

Some pioneers of modern symbol use, such as Mead and Dreyfuss saw few bounds to the use of symbols, and even envisioned symbols as a new language unfettered by linguistic barriers. Others, such as Kolers (1969), objected to the use of symbols as a language because of inherent limitations on combining symbols and on depicting processes, restrictions, or changing ideas. For Kolers, the application of symbols should be restricted to identifying situations or objects and to providing limited instructions or directions. Despite these limitations, however, symbols can be extremely effective, particularly if careful consideration is given to the evaluation of both image quality and intrinsic meaning of a symbol.

Evaluation of symbol design from a graphic viewpoint is an essential step in the development of more effective symbols. Doblin (1979), for example, suggested that the various qualities to be assessed during the design stage should include power, uniqueness, duration, perceptability, and excitement. Whitney (1979) emphasized that design evaluation is most important for providing feedback toward essential improvements. During a graphic review, the designer should focus on the graphic form of the symbol including an assessment of the simplicity of design, consistency of elements, and proportion of legibility of the final product (Dreyfuss, 1966), as well as the need for the message itself (AIGA, 1974). Finally, application of Gestalt principles of perception can further enhance the graphic quality of the image (Easterby, 1967, 1970).

Evaluation of graphic quality is not sufficient in itself, however. Symbols must also be evaluated for understandability. Yet, symbols which are evaluated for graphic consistency are rarely evaluated for understandability--or vice versa. Thus, in the three studies of design evaluation presented in section 2.3.2, no formal attempts were made to collect information about the symbol's meaningfulness or overall effectiveness from a user's standpoint. Similarly, in the studies presented in section 3.0 no attempt was made to use information gained during the assessment of effectiveness for the user to improve the final graphic design. Failure to conduct such bi-directional evaluation has seriously hindered the development of well-designed symbols which also communicate effectively. Furthermore, as pointed out in section 3.8, a complete evaluation of a symbol's effectiveness for the user requires more than just evaluation of understandability. A total evaluation which includes detection, discrimination, recognition, salience, and behavioral effectiveness is rarely accomplished.

Despite the lack of thorough design and research evaluation of symbols, the research studies reviewed in section 3.0 support the use of graphic symbols. The research review indicates that in direct comparisons of word and symbol signs, symbols generally perform better. Specifically, reaction time is faster (Janda & Volk, 1934), understanding is more rapid (Ells & Dewar, 1979), legibility distance is greater (Dewar et al., 1976) and recall is more accurate (Walker et al., 1965). In situations in which word signs perform better than symbols (Dewar & Ells, 1974), the addition of visual distraction and loading tasks, -- an addition which makes the test situation more like the real world -- results in superior performance for symbols (Ells & Dewar, 1976). In addition, the accuracy of symbol identification is greater than for comparable word signs (King, 1971, 1975; Plummer et al., 1974) particularly if interference is added. In one of the few instances where a shorter reaction time was obtained for words (Ells & Dewar, 1979), this difference disappeared when a less verbally-biased response measure was used.

In addition to the direct comparison of word and symbol signs, experimental research has also been directed toward the understandability of one or more sets of graphic symbols. In several cases, researchers have selected and recommended a set of symbols for public use based upon an assessment of their understandability (Brainard et al., 1961; Heard, 1974; Weigand & Glumm, 1979; Griffith & Actkinson, 1978); Easterby & Zwaga, 1976). In only one case was the symbol set found to be difficult to recognize and learn (Griffith & Actkinson, 1978), perhaps due to the large set (128) of symbols tested or the short training time.

When sufficient time and money are available, subjects can produce images or combinations of images for a particular concept. These "production" studies provide valuable insight into the stereotypes that exist for specific concepts. Brainard et al. (1961), Green (1979), and Lerner and Collins (1980) all found that subjects can draw images for a particular referent and that there are common stereotypes for many of the referents. Furthermore, these images were often more accurately identified than a set of standardized images by a subsequent group of subjects (Green, 1979). Easterby and Hakiel (1977a; 1977b; 1977c) also determined that subjects selected images, color, and shapes in a consistent fashion to convey various hazard warnings. Thus, use of a target audience to generate images for subsequent use with that audience may ensure greater understandability of a symbol set. Because this technique can be hampered by the subjects' perceived inability to portray an idea graphically, such production techniques will be most effective if a graphic artist can redraw the various images produced. This also would provide a chance for graphic improvement and refinement of the final image.

Finally, graphic symbols and signs are effective in determining compliance or successful alteration of behavior. Thus, Forbes et al. (1963), Dewar and Swanson (1972), and Freedman and Berkowitz (1976) found that symbols were as effective or more effective than comparable word signs in directing vehicular or pedestrian movement. These data, combined with the results of Laner and Sell (1960), underline the efficacy of behavioral evaluation in determining the ultimate effectiveness of a symbol.

4.2 ISSUES IN THE DEVELOPMENT AND STANDARDIZATION OF MORE EFFECTIVE SYMBOLS

Individual efforts to improve the quality and effectiveness of specific symbols, have typically not been applied to many symbols currently in use. Despite the widespread need for graphic symbols to communicate essential safety, warning, and directional information, various problems have prevented their development and use. These problems arise from the failure to evaluate symbols for both graphic coherence and user effectiveness, and from the current lack of standards for symbols, particularly in the safety information and hazard warning area. Because graphic symbols are widely acclaimed as the cure for written communication failures (Mead & Modley, 1968), they often are developed and implemented without careful design, experimental evaluation, or reference to existing graphic systems. As a result, much of their potential effectiveness is not attained.

The effectiveness of graphic symbols can be enhanced by heeding several key issues. Attention to good graphic design is a paramount issue. Good design is essential to the greater use of symbols, particularly in public buiildings and spaces where much attention is already paid to the design quality of the environment. Yet, good design goes beyond issues of acceptancce by designers and clients; it affects issues critical to the user such as visibility, legibility, and salience. As a result, careful attention must be paid to the graphic quality of any symbol set selected for widespread use and standardization.

A second key issue is the need to determine the intended goal of the symbol. A symbol may be used merely to decorate a space - to make it more interesting visually. In most cases, however, the goal of the graphic symbol is to provide information to an intially uninformed user. This information may be directional, prohibitory, protective, instructive, or prescriptive. In the case of safety symbols, it may be intended to inform the user of the existence of a hazard, the severity of the hazard, the probable consequences of encountering the hazard, or, in some instances, how to avoid the hazard. Different symbol images will often be required to depict each kind of information (see figure 14). Because of the legal need to provide warnings, it is important to select the most accurate and most comprehensive symbolic message for a particular situation. It may be most effective to select an appropriate graphic symbol based upon the specific requirements of the workplace or application.

A third key issue is the determination of the degree of abstraction or realism of the graphic image. Generally, the more abstract the image the less it will be understood, although Easterby and Hakiel (1977c) found that the degree of realistic detail tested in their study did not affect understandability. If the user is taught the symbol's meaning, or if the symbol does not provide information essential to a user's safety, then the degree of abstraction is less critical. A greater degree of abstraction also might be required to convey certain difficult concepts or processes such as "exit" or "buy tickets" or "provide information" or "make reservations". Furthermore, symbols are probably most effective when designed to convey a relatively simple concept which does not demand the combination of ideas. An excessive amount of realism, however, carries with it the possibility that the user will not generalize the message to all intended situations or that it will not be legible or salient. Decisions about the realism or complexity of an image require an assessment of understandability, visibility, legibility, and recognition. Ease of training is another important factor in determining symbolic imagery. If the eventual user can learn the symbol before even confronting it, initial understandability becomes less critical. Training opportunities are often limited, however.

Consideration of the characteristics of the intended user group is the fourth key issue. An image intended particularly for adults might not be satisfactory for children. For example, children tend to associate a skull and crossbones with pirates and games rather than with poison and danger. Yet, a proposed solution, "Mr. Yuk", may seem childish or inappropriate to adult factory workers. (see figure 1). Selecting a symbol image for a given referent as a function of the intended population may cause even further problems. The child trained to recognize "Mr. Yuk" for poison may be puzzled as an adult when encountering the skull and crossbones for the first time. Nevertheless, the characteristics of the intended user group must be determined if undesirable or unexpected associations are to be avoided. Cahill (1976) demonstrated that experience can positively affect symbol recognition for machinery applications. Consequently, symbols indended for specific user groups should be designed with their skills, knowledge, and limitations in mind.

Another important characteristic of the intended user group to consider is disabilities. Many people in the U.S. are sight-impaired with estimates running as high as 6.4 million (Aiello & Steinfeld, 1979). The relative percentage of

Figure 14. Symbol images used to depict different types of information for a single hazard

elderly is increasing, as well. Four studies (Collins & Pierman, 1979; Lerner & Collins, 1980; Hulbert et al., 1979; and Easterby & Hakiel, 1977c) demonstrated that elderly subjects perform significantly worse on tests of symbol understandability. Whether this result occurs because they do not understand the instructions, are reluctant to guess, or have not benefitted from the informal symbol education program provided by current road signs is not clear. Their poorer performance remains an issue to be considered in the implementation of effective symbols, however. Public information symbols such as those for safety, direction, and warning, should be designed with the user's limitations in mind.

A fifth key issue in the development of effective graphic symbol systems is the use of shape and color to code information. Heretofore, symbols have been discussed as though they are simply freestanding images. Yet, most existing symbol systems use both background shape and color to encode supplementary information. The U.S. DoT uses the orange diamond shape to indicate road hazards while restricting the red octagon shape exclusively to stop signs, for example. In the ISO TC 80 system for workplace safety symbols (1978), circles are used to symbolize mandatory or prohibited actions, triangles to provide hazard warnings, and squares or rectangles to indicate general information. Color is used to emphasize the shape-coding system with blue for mandatory action, red for prohibited action (with a slash added to the circle), yellow and black for hazard warning, green for safety, and blue or green for general information. The proposed Australian safety sign standard (1978) followed the ISO system except for substituting a diamond for the triangle for hazard warning.

Yet, despite widespread use in standard symbol systems, the effectiveness of shape and color in encoding information has rarely been assessed. Easterby and Hakiel (1977b) found that subjects chose triangles over circles to convey warning messages. These authors did not test other shape coding options, however. Furthermore, since the triangle is in widespread use in the U.K. as a hazard warning, their subjects may only be reflecting this existing coding system. Use of the triangle for coding hazard has been attacked by the Standards Association of Australia (1978) and others on the grounds that it restricts needed space for symbol placement. Use of the DoT diamond shape for hazard warnings appears to alleviate some of the size problems, but flies in the face of international coding conventions. The question of the need for surround shape to encode level of hazard of nature of the sign remains unresolved.

The selection of the most effective set of shapes and colors to encode information is further complicated by the current use of these coding systems. Such use may not be the "best" means of conveying the desired information. For example, Dewar (1976) who investigated several symbolic codes for prohibition, determined that the conventional prohibitory slash can obscure the symbol, making it less visible and less understandable. Yet, use of a red circle alone would be ineffective for color deficient observers, while a partial slash might be interpreted as only a trivial restriction. Furthermore, the widespread use of the red circle and slash means that this convention, however poor, is difficult to change. Similarly other coding systems, such as the triangles for hazard, may also not be as effective as desired from an experimental point of view. Yet, because these are embedded in various national and international sign standards, it may not be possible to select the most optimal approach for transmitting information. Furthermore, the need to counter the trend toward proliferation of symbol systems may outweigh the need to develop the best possible coding systems. Nevertheless, the issue of developing optimal shape and color coding systems for the most effective transmission of information remains.

Other key issues which must be addressed in the design and use of symbol signs include legibility, durability, and placement. Warning signs are often placed in environments which deteriorate over time. Dust and dirt may affect the visibility of symbols on a day to day basis while smoke may impair legibility during emergency conditions. Thus, the performance of symbols should be evaluated under conditions of blur and dimished contrast. Smith and Weir (1978) for example, found that the performance of arrows of different design varied significantly under blur and glare conditions. Green and Davis (1976) found that placement, or rotation of the symbol, critically affected response time. In buildings, placement of a graphic symbol beyond the normal angle of vision (Follis & Hammer, 1979) can seriously impair its detectability.

Each of the key issues mentioned above must be considered in the development of effective symbols. The various evaluative techniques discussed in section 3.0 and summarized in table 2 can be used to determine the effectiveness of a given symbol. While evaluative research cannot answer all the questions surrounding the development of effective symbols, it can provide insight to the relative performance of a particular symbol and the overall performance of a set of symbols for a given user group. This information can be used to improve the graphic rendition of a particular symbol and to predict the effectiveness of a symbol system for a target population.

Finally, a model for evaluating and predicting symbol performance must be developed. It is not sufficient to determine performance on one of the information processing dimensions discussed in table 2. An integrative model capable of comparing and weighting performance must also be developed and applied across dimensions. Although past decisions about symbol selection and use have typically been unidimensional, centering on one aspect such as understandability, legibility, or graphic coherence, such limited concerns are not adequate for selecting symbols for public use. A legible symbol may not be understandable or vice versa. Yet, because all the processing dimensions noted in table 2 may not be of equal importance, this model must build upon an additive model, or efficiency index, such as that developed by Mackett-Stout and Dewar (1981). As a result, a series of relative weighting factors, related to the demands of the specific application must also be developed to apply to the results from the individual dimensions. Such holistic evaluation is essential for the development and use of more effective building symbols.

The need to conduct evaluative research at both the design and implementation stages is especially critical for supporting the development of standards for symbols. Because of the existing proliferation of symbols and symbol systems, there is an urgent need to develop and standardize symbols which are good graphically and which communicate effectively. Nowhere is the need greater than in the area of safety information and hazard warnings where the failure to communicate can mean serious injury or death.

REFERENCES

Aiello, J. and Steinfeld, E. Accessible buildings for people with severe visual impairments. Washington, D.C: U.S. Department of Housing and Urban Development, HUD - PDR - 404, April 1979.

American Institute of Graphic Arts. Symbol signs - The development of passenger/pedestrian oriented symbols for use in transportation-related facilities. Washington, D.C.: U.S. Department of Transportation, DOT - OS - 40192, November 1974.

American Institute of Graphic Arts. Symbol signs - 2. Washington, D.C.: Department of Transportation, DOT - OS - 60510, March 1979.

Bresnahan, T.F. and Bryk, J. The hazard association values of accident prevention signs. Professional Safety, 1975, January, p. 17-25.

Brainard, R.W., Campbell, R.J., and Elkins, E.H. Design and interpretability of road signs. Journal of Applied Psychology, 1961, 45, 130-136.

Cahill, M.C. Interpretability of graphic symbols as a function of context and experience factors. Journal of Applied Psychology, 1975, 60, 376-380.

Cahill, M.C. Design features of graphic symbols varying in interpretability. Perceptual and Motor Skills, 1976 42, 647-653.

Cantilli, E.J. and Fruin, J.J. Information systems in terminals. Traffic Quarterly, 1972, April, 231-246.

Caron, J.P., Jamieson, D.G., and Dewar, R.E. Evaluating pictographs using semantic differential and classification techniques. Ergonomics, 1980, 23, 137-146.

Collins, B.L. and Pierman, B.C. Evaluation of safety symbols. Washington, D.C.: National Bureau of Standards, NBSIR 79-1760, June 1979.

Dewar, R.E. The slash obscures the symbol on prohibitive traffic signs. Human Factors, 1976, 18, 253-258.

Dewar, R.E. and Ells, J.G. Comparison of three methods for evaluating traffic signs. Transportation Research Record, 1974, 503, 38-47.

Dewar, R.E. and Ells, J.G. The semantic differential as an index of traffic sign perception and comprehension. Human Factors, 1977, 19, 183-189.

Dewar, R.E. and Swanson, H.A. Recognition of traffic control signs. Highway Research Board. 1972, 414, 16-23.

Doblin, J. Part 1: A theoretical model for design evaluation. Industrial Design, 1979, 26, 41-44.

Dorris, A.L., and Purswell, J.L. Warnings and human behavior implications for the design of product warnings. Journal of Products Liability, 1977, 1, 255-264.

Drake, B., Mayblin, B., and Shaw, G. Symbols: Less of an art, more of a science. Design, 1979, 367, 58-63.

Dreyfuss, H. Case study: Symbols for industrial use. In G. Kepes (Ed.), Sign, image and symbol. New York: George Braziller, 1966, pp. 126-133.

Dreyfuss, H. Symbol source book: An authoritative guide to international graphic symbols. New York: McGraw Hill, 1972.

Easterby, R.S. Perceptual organization in static displays for man/machine systems. Ergonomics, 1967, 10, 193-205.

Easterby, R.S. The perception of symbols for machine display. Ergonomics, 1970, 13, 149-158.

Easterby, R.S. and Hakiel, S.R. Safety labelling of consumer products - interim report on shape and colour coding of signs. Birmingham, England: Applied Psychology Department, University of Aston, AP Report 56, March, 1977 (a).

Easterby, R.S. and Hakiel, S.R. Safety labelling of consumer products: Shapes and colour code stereotypes in the design of signs. Birmingham, England: Applied Psychology Department, University of Aston, AP Report 75, December 1977 (b)

Easterby, R.S. and Hakiel, S.R. Safety labelling and consumer products: Field studies of sign recognition. Birmingham, England: Applied Psychology Department, University of Aston, AP Report 75, December, 1977 (c).

Easterby, R.S., and Zwaga, H.T.G., Evaluation of public information symbols ISO tests: 1975 Series. Birmingham, England: Applied Psychology Department, University of Aston, AP Report 60, March, 1976.

Eliot, W.C. Symbology on the highways of the world. Traffic Engineering, 1960, 31, 18-26.

Ells, J.G. and Dewar, R.E. Rapid comprehension of verbal and symbolic traffic sign messages. Human Factors, 1979, 21, 161-168.

FMC Corporation, Product safety signs and labels (2nd Edition). Santa Clara, California: FMC, 1978.

Follis, J. and Hammer, D. Architectural signing and graphics. New York: Whitney Library of Design, 1979.

Forbes, T.W., Gervais, E., and Allen, T. Effectiveness of symbols for lane control signals. Highway Research Board Bulletin. 1964, 244, 16-29.

Freedman, M. Symbol Signs - The testing of passenger/pedestrian oriented symbols for use in transportation related facilities. Washington, D.C.: Department of Transportation, DOT - OS - 60071, December, 1978.

Freedman, M. and Berkowitz, M.J. Preliminary report on laboratory and pilot field testing: Testing criteria and techniques of evaluation for passenger/pedestrian oriented symbols for use in transportation related facilities. Washington, D.C.: Department of Transportation, DOT - OS - 60071, FIRL No. C4448, January, 1977.

Freedman, M., Berkowitz, M., and Gallagher, V.P. Symbol signs: Testing criteria and techniques of evaluation for passenger/pedestrian oriented symbols for use in transportation - related facilities. Washington, D.C.: Department of Transportation, Interim Report, September, 1976.

Giedion, G. Symbolic expression in prehistory and in the first high civilizations. In G. Kepes (Ed.) Sign, image and symbol. New York: George Braziller, 1966, pp. 78-91.

Green, P. Development of pictographic symbols for vehicle controls and displays. Warrendale, Pa: Society of Automotive Engineers, Technical Paper, no. 790 383, February - March, 1979.

Green, P. and Davis, G. The recognition time of rotated pictographic symbols for automobile controls. Journal of Safety Research, 1976, 8, 179-183.

Green, P. and Burgess, W.T. Debugging a symbol set for identifying displays: production and screening studies. Ann Arbor, Mich: University of Michigan Highway Safety Research Institute, Report UM - HSRI - 80-64, Sept. 1980.

Green, P. and Pew, R.W. Evaluating pictographic symbols: An automotive application. Human Factors, 1978, 20, 103-114.

Griffith, D. and Actkinson, T.R. International road signs: Interpretability and training techniques. Proceedings of the Human Factors Society, 21st Annual Meeting, 1977, pp. 392-395.

Griffith, D. and Actkinson, T.R. International road signs: Interpretability and training techniques. Alexandria, Va: Army Research Institute for the Behavioral and Social Sciences, Report no. 1202, September, 1978.

Heard, E.A. Symbol Study - 1972. Detroit, Michigan: Society of Automotive Engineers, Automotive Engineering Congress, February, 25 - March 1, 1974.

Herdeg, W. (Ed). Archigraphia: Architectural and environmental graphics. New York: Hastings House, 1978.

Hoffmann, E.R. and MacDonald, W.A. Short-term retention of traffic turn restriction signs. Human Factors, 1980, 22, 241-251.

Hulbert, S., Beers, J., and Fowler, P. Motorists' understanding of traffic control devices. Falls Church, Va: AAA Foundation for Traffic Safety, March, 1979.

ISO/TC145. Draft addendum - Graphic symbols. 7000/DAD 1, 1980. Geneva: International Organization for Standardization, January, 1980.

ISO/TC21/SC1. Equipment for fire protection and fire fighting - safety signs. Draft Proposal 6309. Geneva: International Organization for Standardization, 1978.

ISO/TC80. Safety colours and safety signs. Draft International Standard - ISO/DIS 3864.3. Geneva: International Organization for Standardization, 1979.

Janda, H.F. and Volk, W.N. Effectiveness of various highway signs. Highway Research Board Proceedings, 1934, 14, 442-47.

Johnston, A.W., Cole, B.L., Jacobs, R.J., and Gibson, A.J. Visibility of traffic control devices: Catering for the real observer. Ergonomics, 1976, 19, 591-609.

King, L.E. A laboratory comparison of symbol and word roadway signs. Traffic Engineering and Control, 1971, 12, 518-520.

King, L.E. Recognition of symbol and word traffic signs. Journal of Safety Research, 1975, 7, 80-84.

Kirsch, I. and Guthrie, J. The concept and meaningfulness of functional literacy. Reading Research Quarterly, 1977-78, 485-507.

Kolers, P.A. Some psychological aspects of pattern recognition. In P.A. Kolers and M. Eden (Eds.). Recognizing patterns - Studies in living and automatic systems. Cambridge, Mass.: MIT Press, 1968, pp. 4-6.

Kolers, P.A. Some formal characteristics of pictograms. American Scientist, 1969, 57, 348-363.

Laner, S. and Sell, R.G. An experiment on the effect of specially designed safety posters. Occupational Psychology, 1960, 34, 153-169.

Lerner, N.D. and Collins, B.L. The assessment of safety symbol understandability by diffferent testing methods. Washington, D.C.: National Bureau of Standards, NBSIR 80-2088, August, 1980.

Lord, K.M. Dashboard symbols - new sign language for car controls. Popular Science. 1980, May, 98.

Mackett-Stout, J. and Dewar, R.L. Evaluation of public information signs. Human Factors, 1981, 23, 139-151.

Mead, M. The future as the basis for establishing a shared culture. Daedalus, 1965, 94, 135-155.

Mead, M. and Modley, R. Communication among all people, everywhere. Natural History, 1968, 77, 56-63.

Modley, R. The challenge of symbology. In E. Whitney (Ed.) Symbology: The use of symbols in visual communications. New York: Hastings House, 1960, pp. 17-30.

Modley, R. Graphic symbols for world-wide communication. In G. Kepes (Ed.) sign, image, and symbol. New York, New York: George Braziller, 1966, pp. 108-125.

Modley, R. Speaking of sign language. Industrial Design, 1976, 23, 60-63.

Modley, R. and Myers, W.R. Handbook of Pictorial Symbols. New York: Dover Publications, 1976.

Morgan, C.T. Introduction to psychology, 2nd ed. New York: McGraw-Hill, 1961.

Plummer, R.W., Minarch, J.J., and King, E.L. Evaluation of driver comprehension of word versus symbol highway signs, Proceedings of the Human Factors Society, 18th Annual Meeting, 1974, pp. 202-208.

Smith, G. and Weir, R. Laboratory visibility studies of directional symbols used for traffic control signals. Ergonomics, 1978, 21, 247-252.

Standards Association of Australia. The design and implementation of public information symbols. Dr 78141 and Dr 78142, Draft, Standards Association of Australia, Committee M 5/3, July 1978.

Walker, R.E., Nicolay, R.C., and Stearns, C.R. Comparative accuracy of response of American and international road signs. Journal of Applied Psychology, 1965, 49, 322-325.

Whitney, P. Part 2: Four real cases of design evaluation. Industrial Design, 1979 26, 44-46.

Wiegand, D. and Glumm, M. An evaluation of pictographic symbols for controls and displays in road vehicles. Aberdeen, Md.: Army Human Engineering Laboratory, Technical Memorandum 7-79, February, 1979.

NBS-114A (REV. 2-80)

U.S. DEPT. OF COMM. BIBLIOGRAPHIC DATA SHEET (See instructions)	1. PUBLICATION OR REPORT NO. NBS BSS 141	2. Performing Organ. Report No.	3. Publication Date May 1982

4. TITLE AND SUBTITLE

The Development and Evaluation of Effective Symbol Signs

5. AUTHOR(S)

Belinda Lowenhaupt Collins

6. PERFORMING ORGANIZATION *(If joint or other than NBS, see instructions)*

NATIONAL BUREAU OF STANDARDS
DEPARTMENT OF COMMERCE
WASHINGTON, D.C. 20234

7. Contract/Grant No.

8. Type of Report & Period Covered

Final

9. SPONSORING ORGANIZATION NAME AND COMPLETE ADDRESS *(Street, City, State, ZIP)*

Same as above

10. SUPPLEMENTARY NOTES

Library of Congress Catalog Card Number: 81-600192

☐ Document describes a computer program; SF-185, FIPS Software Summary, is attached.

11. ABSTRACT *(A 200-word or less factual summary of most significant information. If document includes a significant bibliography or literature survey, mention it here)*

Graphic symbols have recently been widely adopted for sign systems in the United States. Beginning with traffic sign systems, symbols have become widely used for applications ranging from products to buildings. In this report a brief history of the development of symbols is given, followed by a review of research on experimental evaluation of symbols. Some of the general advantages and limitations of symbols are discussed, along with graphic considerations essential in the development of effective symbols. Research on symbols for five areas of application -- highway, automotive/machinery, public information, product hazard, and safety -- is then discussed.

Finally, issues in the research and development of more effective symbols are reviewed. These include the need for good graphic design, characteristics of the intended user group, use of shape and color to encode information, and general visibility considerations.

12. KEY WORDS *(Six to twelve entries; alphabetical order; capitalize only proper names; and separate key words by semicolons)*

communication; design issues; hazard; pictograms; pictorial; safety; signs; standards; symbols; visual alerting; warning

13. AVAILABILITY

[X] Unlimited
☐ For Official Distribution. Do Not Release to NTIS
[X] Order From Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402.
☐ Order From National Technical Information Service (NTIS), Springfield, VA. 22161

14. NO. OF PRINTED PAGES

96

15. Price

USCOMM-DC 6043-P80

☆ U.S. GOVERNMENT PRINTING OFFICE : 1982 O—360-997 (2110)

NBS TECHNICAL PUBLICATIONS

PERIODICALS

JOURNAL OF RESEARCH—The Journal of Research of the National Bureau of Standards reports NBS research and development in those disciplines of the physical and engineering sciences in which the Bureau is active. These include physics, chemistry, engineering, mathematics, and computer sciences. Papers cover a broad range of subjects, with major emphasis on measurement methodology and the basic technology underlying standardization. Also included from time to time are survey articles on topics closely related to the Bureau's technical and scientific programs. As a special service to subscribers each issue contains complete citations to all recent Bureau publications in both NBS and non-NBS media. Issued six times a year. Annual subscription: domestic $18; foreign $22.50. Single copy, $4.25 domestic; $5.35 foreign.

NONPERIODICALS

Monographs—Major contributions to the technical literature on various subjects related to the Bureau's scientific and technical activities.

Handbooks—Recommended codes of engineering and industrial practice (including safety codes) developed in cooperation with interested industries, professional organizations, and regulatory bodies.

Special Publications—Include proceedings of conferences sponsored by NBS, NBS annual reports, and other special publications appropriate to this grouping such as wall charts, pocket cards, and bibliographies.

Applied Mathematics Series—Mathematical tables, manuals, and studies of special interest to physicists, engineers, chemists, biologists, mathematicians, computer programmers, and others engaged in scientific and technical work.

National Standard Reference Data Series—Provides quantitative data on the physical and chemical properties of materials, compiled from the world's literature and critically evaluated. Developed under a worldwide program coordinated by NBS under the authority of the National Standard Data Act (Public Law 90-396).

NOTE: The principal publication outlet for the foregoing data is the Journal of Physical and Chemical Reference Data (JPCRD) published quarterly for NBS by the American Chemical Society (ACS) and the American Institute of Physics (AIP). Subscriptions, reprints, and supplements available from ACS, 1155 Sixteenth St., NW, Washington, DC 20056.

Building Science Series—Disseminates technical information developed at the Bureau on building materials, components, systems, and whole structures. The series presents research results, test methods, and performance criteria related to the structural and environmental functions and the durability and safety characteristics of building elements and systems.

Technical Notes—Studies or reports which are complete in themselves but restrictive in their treatment of a subject. Analogous to monographs but not so comprehensive in scope or definitive in treatment of the subject area. Often serve as a vehicle for final reports of work performed at NBS under the sponsorship of other government agencies.

Voluntary Product Standards—Developed under procedures published by the Department of Commerce in Part 10, Title 15, of the Code of Federal Regulations. The standards establish nationally recognized requirements for products, and provide all concerned interests with a basis for common understanding of the characteristics of the products. NBS administers this program as a supplement to the activities of the private sector standardizing organizations.

Consumer Information Series—Practical information, based on NBS research and experience, covering areas of interest to the consumer. Easily understandable language and illustrations provide useful background knowledge for shopping in today's technological marketplace.

Order the **above** *NBS publications from: Superintendent of Documents, Government Printing Office, Washington, DC 20402.*

Order the **following** *NBS publications—FIPS and NBSIR's—from the National Technical Information Services, Springfield, VA 22161.*

Federal Information Processing Standards Publications (FIPS PUB)—Publications in this series collectively constitute the Federal Information Processing Standards Register. The Register serves as the official source of information in the Federal Government regarding standards issued by NBS pursuant to the Federal Property and Administrative Services Act of 1949 as amended, Public Law 89-306 (79 Stat. 1127), and as implemented by Executive Order 11717 (38 FR 12315, dated May 11, 1973) and Part 6 of Title 15 CFR (Code of Federal Regulations).

NBS Interagency Reports (NBSIR)—A special series of interim or final reports on work performed by NBS for outside sponsors (both government and non-government). In general, initial distribution is handled by the sponsor; public distribution is by the National Technical Information Services, Springfield, VA 22161, in paper copy or microfiche form.