Investigating the Methodology of Warning Symbol Design

by

Kristen Michelle Haynes

A thesis submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Master of Science

Auburn, Alabama
May 7, 2016

Keywords: warning symbol, industrial engineering, safety, hazard avoidance

Approved by

Jerry Davis, Chair, Associate Professor of Industrial and Systems Engineering
Richard Sesek, Associate Professor of Industrial and Systems Engineering
Robert Thomas, Professor Emeritus of Industrial Engineering
Adam Piper, Assistant Professor of Industrial Engineering at South Dakota School of Mines and Technology

Abstract

The present study is a continuation of work done by Dr. Adam Piper in his 2010 dissertation on warning symbol design. Warning symbols are important because, if designed correctly, they communicate information about a hazard quickly and effectively so that an individual can react accordingly if exposed to it. Warning symbols themselves, as an alternate to warning messages, have the added benefit of a potential to traverse language barriers and add comprehensibility to a warning message.

Several methods of symbol design have been proposed, but none that utilize the input from the general population from start to finish. Current methods include the designer method, the production method, and the focus group method. The present experiment attempts to expand upon these methods by utilizing the general population throughout the study. A group of general population participants was recruited to draw a symbol based on a given safety referent. These drawings were then semantically annotated by a naive set of individuals representing the general population. The resulting lists of attributes were entered into matrices and clustered via a word clustering software program developed by Feinberg (2014) called Wordle, which identified significant attributes through the size of each attribute in a word cloud (i.e., the larger the word appeared in the cloud, the more often it occurred in the matrices). This procedure resulted in the same general core attribute lists for both referents when compared to the list produced by a panel of safety experts in a previous study.

The most significant limitation of this study was the fact that it took a significant amount of time to check each individual matrix for accuracy, and subsequently consolidate the similar terms into a workable number of attributes for clustering. In the future, an updated method of data consolidation must be employed to reduce the amount of time required for pre-data analysis. Preliminary data analysis revealed that similar attribute lists were created via both the general population method and expert panel method. Therefore, it may be advantageous to utilize an amended version of the general population method presented in this study in order to reduce costs of symbol creation.

Acknowledgements

The author would like to thank Dr. Adam Piper for hosting her while she conducted the research presented in this report as well as his guidance and support throughout this entire process. She would also like to thank Dr. Richard Sesek for sharing his knowledge and experience in safety. Thanks go to Dr. Martha Escobar for introducing her to Human Factors and Ergonomics and encouraging her to apply to an Engineering program as a Psychology major. Lastly, she would like to thank Dr. Jerry Davis for mentoring her not only on this project but continually supporting and encouraging her throughout her entire graduate school experience.

Table of Contents

List of Tables

List of Illustrations

Introduction

The present research is directly related to the work of Dr. Adam Piper in his 2010 dissertation on warning symbol development. Therefore, a summary of the experiments outlined and information presented in the dissertation is described below.

Referent to Symbol Conversion Difficulty

Dorris (2004) notes that when designing a new warning symbol, it is important to select a referent carefully. However, Piper (2010) reported that there was little to no research done on evaluating the effectiveness of a referent before a symbol was created from it. At the time, several studies in the literature measured the communicative effectiveness of a symbol after its creation (Isherwood, McDougall, J.P, & Curry, 2007; S. J. McDougall, Curry, & de Bruijn, 1999; S. J. P. McDougall, de Bruijn, & Curry, 2000; Young & Wogalter, 2001). However, at the time of Piper's (2010) dissertation, there was only one study evaluating the effectiveness of a referent before a symbol is generated. This study was conducted by Hicks, Bell, and Wogalter (2003), who utilized the term "ease of visualization" when asking participants to rate how well they could understand a given referent. The authors concluded that the open-ended comprehension test was the most apt method of measuring a symbol's understandability.

Unfortunately, as Piper (2010) reported, ease of visualization is not a measure of the level of difficulty required to convert a referent to a graphic symbol. For this reason, Piper (2010) conducted a study in order to rank a given set of referents by "ease of symbol conversion." The aim of this study was to provide a list of referents from which experimenters can choose an appropriate referent based on the level of difficulty he or she wishes to use for his or her research purposes. Additionally, testing multiple referents from this list at different difficulty levels allows for comparison between different methods of symbol generation. For this experiment, participants were asked to view written descriptions of different referent pairs, and then asked to choose which one of the proposed referents would invoke more difficulty in drawing a symbol based on the given information about the referent. Five of the nine referents tested were found to be statistically different from one another in terms of conversion difficulty (Piper, 2010). From this procedure, it was found that it is possible to rank a given set of safety referents based on participant comparisons. The list of referents in the above-described experiment was used to select the two referents used in the current research: Hot Exhaust (ranked number two of nine) and Do Not Touch with Wet Hands (ranked number five of nine but found to be equally difficult to convert as three other referents). The table ranking these referents is depicted below in Table 1, originally illustrated in Piper's (2010) dissertation.

Referent
1. No Access for Persons with Metallic Implants
2. Do Not Touch with Wet Hands
3. Warning: Flooring Surface Changes
4. Confined Space; Entry by Permit Only
5. Disconnect Main Plug from Electrical Outlet
No Reaching In
Steel-toed Shoes Required
Hot Exhaust
6. Walk Down Stairs Backwards

Table 1. Final ranked order of warning referents by ease of conversion.

A New Method of Symbol Creation

There have been several methods developed to attempt to solve the problem of confusing warning symbols, but none have used the general population (i.e., individuals unfamiliar with the field of safety and warning symbols) to systematically evaluate the merits of proposed designs. A more detailed description of each method is included in the "Types of Symbol Design" section, but they are briefly outlined here to facilitate a better understanding of the motivation behind Piper's 2010 experiments. The most traditional method of symbol design is what Piper (2010) refers to as the *Designer Method*. In this method, the team of individuals responsible for creating a new symbol decides the important attributes that it should contain, and a professional designer drafts a number of options. This set of designs is then sent to other professionals for their opinions, and a symbol is finally chosen from the list (Wisniewski, E. C., Isaacson, J. J., & Hall, S. M., 2007).

A second method is called the *Production Method*. In the Production Method, a set of individuals who may be affected by the implementation of the new symbol are gathered and asked to draw what they believe to be a valid symbol for the given referent. A professional designer is still utilized to combine these drawings into one final symbol.

The third method of symbol creation is a variant of the Production Method, and is termed the *Focus Group* method. This is the only method that does not utilize the professional designer for decision-making in regards to the symbol's design. As the name suggests, several small groups of individuals are formed in which the participants discuss the referent and create a symbol based off of their discussion (Goldsworthy & Kaplan, 2006; Macbeth, S. A., Moroney, W. F., & Biers, D. W., 2000; Mayhorn & Goldsworthy, 2007). The professional designer is only responsible for creating a professional-looking symbol from the one design produced by the focus group (Dorris, 2004).

While participatory design methods are useful in that they incorporate the opinions of non-expert individuals as well as provide a larger range of feedback for the symbol's design, there are a few shortcomings to these methods. Piper (2010) reports that "culture and language barriers, variations in prior experience and conflicting personality traits" are the most important flaws to consider due to their potentially detrimental effect on the resulting symbol that is created.

In a second experiment, Piper (2010) attempts to control for these shortcomings by introducing technology into the symbol design process. In it, the distributed interactive genetic algorithm (DIGA) method is described, which essentially removes the need for a professional designer. Through the DIGA process, a computer interface processes the designs received from participants and utilizes a genetic algorithm to create a finalized symbol (Dozier, G., Carnahan, B., Seals, C., Kuntz, L. A., & Fu, S. G., 2005a). Participants are given a set of attributes from which to choose in the design process, eliminating the variation in previous experience with safety referents and symbols (Piper, 2010). The use of technology not only has the potential to reach a much higher number of participants, but it also allows for elimination of cultural or language conflicts due to the ability of participants to utilize the system in their most familiar language.

In his experiment, however, Piper (2010) attempts to have participants create this initial set of attributes from which future participants will choose a symbol design. The goal of the experiment was to semantically derive a list of attributes to pre-program into the DIGA system. Initially, participants were asked to draw a symbol from one of two randomly selected referents, Hot Exhaust, and Do Not Touch with Wet Hands. After participants' drawings had been collected, an expert panel of three engineers was recruited to evaluate the drawings. These individuals had been previously trained on the process of semantic annotation and were familiar with both warning symbols in general and the DIGA system. The expert panel evaluated each drawing on the basis of the attributes present. They then omitted some drawings due to either an egregious error or critical confusion. An egregious error resulted if there was a "substantial misunderstanding or misrepresentation of the referent," and critical confusion was defined as a drawing in which the participant drew "the opposite message or a message that could lead to severe injury" (ANSI, 2007a). If the drawing was determined to be a legitimate candidate, the panel created a matrix containing one attribute per column, and one drawing per row. The cells corresponding to each row and column pair represented "a binary response to the question, 'Is this attribute present in this symbol drawing?'" (Piper, 2010). The matrices were then clustered using a K-means clustering algorithm that utilized the median attributes to produce a set of primary attributes for the given referent (Piper, 2010). This algorithm is described in more detail in Piper's dissertation (Piper, 2010). For the referent "Hot Exhaust," the expert panel identified 35 attributes, and three were identified as primary attributes by the clustering algorithm (Piper, 2010). For the referent "Do Not Touch with Wet Hands," 70 attributes were detected, and the algorithm identified five of these as primary (Piper, 2010). Table 2 lists specific primary attributes that the algorithm identified. The results of this study indicate that it is possible to construct a new tool for symbol design that incorporates technology in order to reduce the potentially biased nature of traditional symbol design methods. Piper (2010) suggests that use of the general population can improve symbol design.

There are several conclusions of the experiments described above that are relevant to the current research. First, it is possible to quantify a level of difficulty in symbol conversion from a referent when compared to other referents. It is possible to do this with not only safety professionals, but also naïve participants. Additionally, drawings made from referents can be successfully deconstructed into a list of primary attributes by an expert panel. Lastly, a K-means algorithm can be utilized to cluster these attributes into categories rather than individuals categorizing the attributes themselves.

The drawback to the studies presented above that is of particular interest to the current research is that the general population was not utilized in the semantic annotation of the drawings. While it is advantageous that general population participants created the drawings, the use of safety professionals on the "expert panel" that created the primary attribute lists from these drawings may have introduced bias into the symbol design, as "experts" in the field of safety are likely to have previous knowledge of warning symbols and the way that safety is communicated in the industry. This could result in the use of that prior knowledge, whether intentional or otherwise, to design the symbol instead of relying solely on the drawings presented by the general population. Therefore, the main concern that the author has with this method is that the use of the expert panel in identifying primary attributes of the drawings did not allow for the general population's participation throughout the course of the symbol design process. Additionally, the expert panel is expensive. In Piper's (2010) dissertation, it took three experts nine hours total to identify a set of core attributes from the sketches. It is possible that using the general population will be less expensive because they do not require the same level of compensation than industry professionals or consultants.

The objective of the current research was to expand on the focus group method of design to include general population participants' views throughout the entire process of semantically annotating and identifying primary attributes of the drawings done for each of the two referents that Piper (2010) used for the expert panel method. Instead of using an expert panel to generate primary attributes, general population participants were recruited to semantically annotate the same warning symbol drawings that the expert panel had previously annotated in the experiment described above. The intent of this study was to investigate whether the general population's views on warning symbol components allows for a more objective symbol design.

Literature Review

Significance of Warning Symbols

According to Wogalter (2006), a warning is a way to communicate information about a hazard to a set of individuals. Laughery and Wogalter (2006) report four different purposes for warnings. The first purpose is described above – to communicate information about a hazard. The second purpose is to alter an individual's behavior in response to a hazard, specifically to avoid the reported hazard. Third, warnings are an important reminder of past events that resulted from a lack of knowledge about a hazard. Lastly, and perhaps most significantly, warnings promote a safer environment for all individuals that would otherwise be affected by a given hazard.

Research has shown that when an effective warning symbol is paired with a warning message, it can add comprehensibility to the original message (Wogalter, Silver, Leonard, & Zaikina, 2006). In fact, Kalsher, M. J., Wogalter, M. S., & Racicot, B. M. (1996) reports that individuals prefer warnings with symbols to those without symbols. However, Piper (2010) suggests that pairing a symbol with a warning message may not always add effectiveness. In fact, it may decrease the comprehensibility of the original warning message, potentially resulting in an individual being exposed to a hazard instead of avoiding it. For example, if a symbol does not have enough detail, it may take too long to interpret, or be interpreted incorrectly altogether. This increases an individual's risk of injury. Therefore, it is of utmost importance to design a symbol that communicates the warning message clearly, concisely, and effectively.

Warning Symbol Standards

As the body of research on warnings has grown, there has been a growing interest in promoting safety in the workplace. This has been in part due to increased litigation and workers' compensation claims, as well as the establishment of safety organizations (Clark, Benysh, & Lehto, 2003; Egilman & Bohme, 2006). As a result, ANSI, ISO, and OSHA have approved standards on the use of safety warnings in addition to voluntary guidelines for employers (ANSI, 2011c; ISO, 2003; OSHA, 1996). These standards and guidelines are for use in both industry and in the public environment (ANSI, 2011a, 2011b, 2011c; ISO, 2002, 2004, 2006, 2007, 2008). This was an important development because for the first time, there was a uniform guideline to follow in the development and presentation of warning symbols. However, there are differences between the ANSI and ISO standards developed. For example, because the symbols may be viewed by individuals speaking up to sixteen different languages, the ISO 3864 standard states that a symbol does not necessarily have to be paired with a warning message, and that the symbol can stand alone (Deppa, 2006). However, the ANSI Z535.4 (2011c) standard states that a symbol must include a word message about the hazard indicating type, consequences, and method of avoidance. According to ANSI Z535.4 (2011c), the only situation in which it is appropriate to omit a word message is when a manual or instructions accompany the symbol, or when it passes all requirements of ANSI Z535.3 (2011a) Annex B for symbol comprehension. Because the standards are different in their requirements in addition to lacking in suggestions as to how to develop an effective symbol, the resulting symbols may deviate from one another significantly in both appearance and function (Piper, 2010).

The standards propose the use of an open-ended comprehension test in discerning a given symbol's understandability. ANSI Z535.3 (2011a) dictates that a symbol should pass with an 85% or above among 50 participants that represent the target population. ISO 9186-1 (2007) suggests a 67% passing rate among 50 participants from three different "culturally diverse" countries. The standards do agree that the critical confusion rate among participants must be less than 5% (ANSI, 2011a; ISO 9186-1 (2007), where "critical confusion" is indicated by an individual interpreting a symbol as having the opposite meaning of the warning message or believing that the symbol promotes an unsafe action (Wogalter, M. S., Silver, N. C., Leonard, S. D., & Zaikina, H. 2006).

Symbol Comprehensibility

Piper (2010) reports that warning symbols are only effective when they elicit a reaction, meaning that they are passively – not actively – protective. This is important to note because a symbol must be acted upon in order to serve its purpose, and if a symbol is not comprehended, no action can be taken. The open-ended comprehension test is considered the gold standard in relation to evaluating the comprehensibility of symbols (Hicks et al., 2003). The population estimation technique is a simplification of the open-ended comprehension test that involves participants estimating what percentage of the general population would understand the symbol given the symbol and its meaning (ANSI Z535.3, 2011a, ISO 9186-1, 2007). Young and Wogalter (2001) found that results of the open-ended comprehension test highly correlate with results of the amended version of the test. Lesch (2005), however, conducted a paired-response test where a symbol is shown with a label and users are expected to indicate whether the label correctly describes the symbol. Through this method, Lesch (2005) concluded that symbol comprehensibility is underestimated by the open-ended comprehension test and that semantic relatedness is a better alternative. Due to the conflicting views on open-ended comprehension testing, it may be useful to employ another method.

While the multiple-choice test is also commonly used, it is not as reliable as the open-ended comprehension test (Wolff, 1995). Further, Lesch (2005) found that comprehensibility of a symbol was overestimated when tested via the multiple-choice method. Lesch recommends the use of semantic relatedness as opposed to both the open-ended comprehension test and the multiple-choice comprehension test. Semantic relatedness refers to the process of pairing a symbol with a message and asking individuals whether or not the symbol and message are correlated. Lesch (2005) believes that the semantic relatedness test is optimal in that it neither overestimates nor underestimates symbol comprehensibility. However, there is a possibility of the introduction of bias in semantic relatedness testing because it does not allow for a free response. An individual undergoing this type of testing may be guided by the symbol and message presented, not by what he or she thinks without prompting.

Symbol Familiarity

It is important to consider participants' familiarity with a symbol because the more familiar one is with the given referent or warning symbols in general, he or she will potentially have a different view on what the symbol should look like. The more exposure that an individual has to the symbol or anything related to it, the more familiar he or she is said to be with that given symbol (Isherwood et al., 2007). Another element of symbol familiarity is known as *semantic distance*, which refers to how closely a symbol depicts its referent. If a symbol has good semantic distance, it is said to be more effective than a symbol that is more distant (McDougall et al., 1999). The last important aspect of symbol familiarity is its *ease of visualization*, which is a term defined by Hicks et al. (2003) as how easily a symbol can be visualized by any given individual. Piper (2010) reports that ease of visualization is paramount in a symbol's familiarity because it is the only part of symbol design that does not directly relate to the actual process of designing the symbol. Instead, it is a part of the overall design quality once the symbol is finished (Piper, 2010). Dorris (2004) also mentions the importance of individuals' ability to visualize a symbol in terms of its importance to symbol comprehensibility. Piper (2010) cautions, though, that visualizing a symbol is different from producing a finished design, so the two may not correlate.

Symbol Generalizability

Warning symbols are advantageous in that they may be understandable by a greater number of people than traditional warning messages (Wogalter et al., 2006). In fact, warning symbols have been confirmed as "language-independent" (Liu, Hoelscher, & Gruchmann, 2005) and "culture-neutral" (Edworthy & Adams, 1996). While there have been a number of studies on warning symbols, there is a significant gap in the literature when it comes to cultural and language-related components of warning symbol design, as evidenced by the lack of diversity among participants recruited for studies (Huer, 2000; Lesch, Rau, Zhao, & Liu, 2009; Russo & Boor, 1993).

Symbol design is also not always generalizable to the general population. Dorris (2004) reports that symbol design has traditionally been the responsibility of experts in the field, who use their experience to develop a warning symbol that they think is appropriate for a given warning message. Unfortunately, this means that the general population is not consulted during the design process, reducing the likelihood that the symbol is relevant to the target population (Huer, 2000; Laughery, 2006). For this reason, using the target population as participants in symbol development is now believed to be helpful in improving the effectiveness of the finalized warning symbol (Schuler & Namioka, 1993). One of the main purposes of including a symbol in a warning is to communicate the warning message to individuals with a language barrier, but an ineffective symbol design can have a detrimental effect on these individuals (Dorris, 2004; Huer, 2000). Despite this knowledge, intended users are not often included in the symbol design process. If they are included, it is only in the initial stages, which is not sufficient (Dorris, 2004; Huer, 2000).

Types of Symbol Design

Since visual warnings have traditionally been the focus of much of the body of research on warnings thus far (Smith-Jackson & Wogalter, 2006), the research of Piper's (2010) dissertation focused on visual warnings as well. Since the research done by Piper is the precursor to the current research, it also focuses on visual warnings.

The general three-step process of warning symbol development defined by Dorris (2004) involves the determination of a given warning's message, also known as a referent. This referent tells individuals how to react to or interpret a given warning message. Next, a number of potential symbols must be drawn. Lastly, the drawings must be pooled and narrowed down to one final symbol by a given set of individuals (Dorris, 2004). The process described above has several variants, and each of these has different advantages and disadvantages.

Designer Method

Traditionally, the most common method of symbol design has been termed the "designer method." In this method, an expert panel is recruited to analyze the meaning of a given safety referent. They then come up with a list of attributes that they believe will produce the most effective symbol for that referent. The list is given to a professional designer to produce a final, usable symbol. The symbol is then either tested for comprehension or put directly into use (Ringseis & Caird, 1995; Roberts et al., 2009). If it is tested, it is re-circulated back through the expert panel to update the attribute list, then sent to the designer. After the designer updates the symbol, it is tested once again. The process is repeated until a final symbol is finalized (Wisniewski et al., 2007).

Production Method

A second symbol development method involves recruiting participants to draw sketches from a given safety referent. These sketches are then given to a designer to create a cohesive symbol from them. The utilization of a larger number of participants in the initial stages of design improves the symbol's comprehensibility for the target population (Schuler & Namioka, 1993). When compared to the designer method, Sloan and Eshelman (1981) reported that all symbols resulting from the production method were more effective than those produced by the designer method alone. They concluded that the use of participants in symbol design significantly improves the effectiveness of the final symbol.

Focus Group Method

The focus group method is a further deviation from the designer method. It still employs the use of participants in the initial sketches from the referent. However, the participants compare the drawings to one another themselves instead of sending the drawings to a designer. After comparing all the sketches, the participants decide on one final design to submit to the artist (Dorris, 2004; Goldsworthy & Kaplan, 2006; Macbeth & Moroney, 1994; Macbeth et al., 2000; Macbeth, S. A., Moroney, W. F., & Biers, D. W., 2006; Mayhorn & Goldsworthy, 2007). While this method takes much longer than the production method (Dorris, 2004), it may produce a more effective final symbol since it involves the most participation from the general population and does not allow the designer to have any input in the design process itself (Macbeth & Moroney, 1994).

Introduction of Technology

As technology has progressed, it has been incorporated into symbol design methods (Carnahan & Dorris, 2004; Carnahan, Dorris, & Kuntz, 2005; Dorris, 2004; Dorris, Carnahan, Orsini, & Kuntz, 2004; Dozier et al., 2005b; Parmee, Abraham, & Machwe, 2008). In fact, Piper's (2010) dissertation attempted to "bridge the gap between participatory design and computational technology" through the use of a computational software program. Piper intended to utilize this program to eliminate the focus group and instead crowd source individual participants from around the world. These participants would access the design program individually, and the program would compile the designs into a final symbol, eliminating the designer from the process completely (Piper, 2010).

Semantic Annotation

The main disadvantage of the use of a computer program for symbol design is that individuals cannot freely draw a symbol, and are instead confined to the use of the elements of the design available through the program software. This is a disadvantage because the researcher must select ahead of time the elements of the symbol from which participants may choose so that they can be incorporated into the software (Piper, 2010). To combat this phenomenon, Piper (2010) utilized semantic annotation, which is defined as the method of enlisting an individual or set of individuals to choose important aspects of a given medium, in this case a drawing (Carneiro & Vasconcelos, 2004; Turnbull, Liu, Barringon, & Lanckrie, 2007; Vasconcelos & Lippman, 2000a, 2000b).

Semantic annotation can be applied in a number of ways. For example, Piper, Boelhouwer and Davis (2008) utilized a set of individuals with extensive experience in the field of safety to define the important aspects, or attributes, of each drawing provided by the participants in one of Piper's (2010) studies. A binary matrix was developed from this annotated list of attributes. Each drawing was assigned to a row, and each attribute to a column. If a given attribute was present in a symbol, a one was entered into the corresponding cell. If the attribute was not present, the cell was left blank, indicating a zero value (Piper et al., 2008). Piper (2010) utilized this method in his dissertation research studies in order to find the most important attributes of the drawings, then used a clustering program to further narrow down the attributes into the most important, "core" aspects.

While the method described above was successful in producing a list of core symbol attributes, Aggarwal (2004) reports that the combination of technology with human discretion produces a better product, and suggests that individuals should be responsible for the initial identification of one row of attributes for the software program to cluster into a final symbol. The current research aims to replicate this technique, but instead of using an expert panel, the study will utilize the general population to semantically annotate the drawings. The specific purpose of this is to investigate whether or not the general population and the expert panel produce the same core set of attributes for the same two warning referents.

Material & Methods

For the present study, a volunteer sample of 56 student participants at the South Dakota School of Mines was recruited by email invitation. They were assigned to six groups (A- F) of six to twelve participants each. Each group represented one of the two referents (totaling three groups per referent), and was assigned to five, ten, or fifteen seconds of viewing the drawing. Participants were anonymous, and only basic demographic information was collected. The testing was conducted over three days, and two groups were tested each day. No participant was assigned to more than one group.

Each group was shown a slide show of fifty drawings that was set on a timer set to five, ten, or fifteen seconds per image. Participants were asked to keep their paper turned over for the duration of the time that each drawing was displayed on the slide show. After each image was shown for its allotted time, a new slide appeared instructing participants to annotate the drawing. They then turned their papers over and annotated the drawings into a list of 'significant' attributes that they remembered seeing in the drawing they had just viewed. Participants were instructed to list each attribute as concisely as possible while still describing it completely. They were asked to avoid using complete sentences or combining multiple attributes. For example, "man slipping on water" should be broken into the attributes "man," "slipping," and "water."

After collecting the attribute lists of all participants, they were recorded in matrices. Each attribute list was recorded into one binary matrix. Matrix columns 1-50 corresponded to the sketch numbers 1-50. Matrix rows corresponded to the individual attributes listed. This number varied with each list since participants were not given a lower or upper bound for the number of attributed listed for each sketch viewed. Each time an attribute appeared in a sketch, the number 1 was entered into the corresponding box. All other boxes remained blank, therefore classifying that box as a 0.

Each matrix was then checked for accuracy, and incorrect transcription recordings were corrected. Mistakes in transcription were recorded directly onto the attribute lists as follows: highlight correct attributes and flag illegible attributes. Transcribers were to denote incorrect transcription with a "T," denote a missed attribute with an "M," and denote an attribute in an incorrect sketch number (for example, the participant wrote "fire" in Sketch 4, but the transcriber noted it in Sketch 3) with an "S." Accuracy checkers corrected all discrepancies except illegible attributes directly on the matrix. The author analyzed all flagged illegible attributes and made a decision on whether or not to record that attribute, and if so, what word should be recorded. The number of incorrect transcriptions was recorded for each attribute list.

After all matrices were confirmed to be accurate, they were clustered via Wordle (Feinberg, 2014), which utilizes Microsoft Excel to create word clouds. In the present study, the matrices that were created from the attribute lists were consolidated to eliminate duplicate and redundant attributes. For example, "flame" and "flames" were consolidated into one term. Synonyms, such as "pipe" and "tube," were also consolidated for the referent "Do Not Touch with Wet Hands." However, synonyms for "Warning: Hot Exhaust" were left separate, as this referent was not analyzed past the initial stages. While it may be possible to consolidate the matrices further, the present study wanted to err on the side of caution when combining terms so as to retain as much of the original participants' attributes as possible. These word clouds were then used to determine the core set of attributes identified by participants.

Results

Prior research using an expert panel originally identified a small set of 'core' attributes to be included in the symbol design for each referent. For "Warning: Hot Exhaust," core attributes identified were "emission lines," "pipe/stack," and "flame." For "Do Not Touch with Wet Hands," core attributes identified were "single hand," "water drops," "prohibition symbol," "faucet," and "prohibition X" (Piper, 2010). Preliminary analysis of attribute lists in the current study indicated that while a larger number of untrained participants identified the same set of attributes as most significant, a larger and more varied pool of attributes was identified. Similar (and in some cases the exact) attributes selected by the expert panel appeared in the word clouds, most of them as larger words. This is summarized in Table 2.

Figures 2 and 3 depict the word clouds that Wordle (Feinberg, 2014) produced. The larger the word is, the more often it appeared on the attribute lists corresponding to that referent. Visual analysis of the word clouds revealed that some of the larger words represent the same general set of attributes identified by the expert panel, but some of the largest words (for example, "red" in Figure 1) from the general population data were not identified as primary attributes by the algorithm used by the expert panel. Further data compilation and statistical analysis is required in order to obtain a reliable list.

Referent	Expert Panel	General Population
Warning: Hot Exhaust	Emission Lines	Exhaust/Lines
	Pipe/Stack	Pipe
	Flame	Fire/Flames
Do Not Touch with Wet Hands	Single Hand	Hand
	Water Drops	Water/Drops
	Prohibition Symbol	Circle/Do Not
	Faucet	Faucet
	Prohibition “X”	Cross

Table 2. Comparison of core attributes selected.

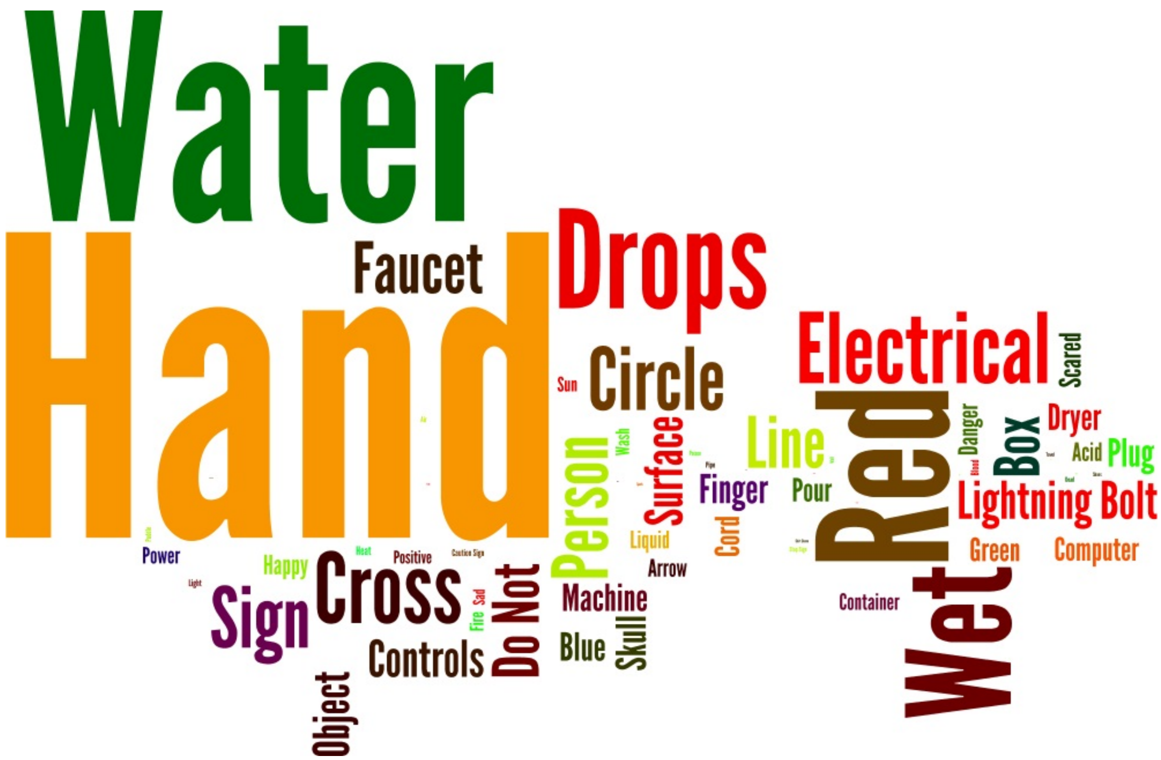


Illustration 1. “Do Not Touch with Wet Hands”

Illustration 2. “Warning: Hot Exhaust”

Discussion

Identified Attributes

For the referent "Do Not Touch with Wet Hands," visual analysis indicated that the words "hand," "water," "drops," "do not," "faucet," and "cross" were all important attributes (Figure 1). These were almost identical to the expert panel's list depicted in Table 2. This indicated that the opinions of the expert panel and the general population might be comparable in symbol design for this particular referent. However, since the same algorithm used in the expert panel method did not cluster results from the present study, further analysis is required in order to gain a more statistical understanding of the true core attribute list created by the general population.

An interesting finding from the word cloud for "Do Not Touch with Wet Hands" was that it also indicated that the attributes "red," "circle," and "line" were important. These words on their own did not make sense as attributes, but by combining them into a comprehensible phrase (i.e., "red prohibition symbol"), it became clear as to why they were represented so largely on the word cloud. At this point, it might have been advantageous for a safety expert to visually inspect the data to ensure that the words produced by the attribute lists were representative of what the general population most likely intended. However, this author believes that this is not necessary. While it is known in the safety field that the correct term is "prohibition symbol," the general population participants are unlikely to know this. It is possible that because the general population is responsible for creation of the symbol from start to finish, there will not be a disconnect because one general population participant will understand what another meant by “red line with circle,” and will be able to incorporate it seamlessly into the final design.

Preliminary analysis on the referent “Warning: Hot Exhaust” indicated that the important attributes identified include “exhaust,” “lines,” “pipe,” “fire,” and “flames,” (Figure 2). These may be more representative of the referent than some of the attributes identified for “Do Not Touch with Wet Hands,” but the words in this cloud were more similar in size, indicating that there was not as much distinction between which attributes were comparatively more important. This is likely due to the fact that no consolidation was done for attribute lists of “Warning: Hot Exhaust.” Further analysis must be done on the attribute lists produced by the general population in order to ascertain whether the core attributes identified truly match those of the expert panel.

Limitations

The first limitation encountered with this study was that the matrices created through the transcription process had errors. This was unexpected because the matrices were simply transcribed directly from the attribute lists to the computer. Since lists were transcribed verbatim, matrices should match the lists with little to no errors, provided that transcribers took an ample amount of time to transfer the lists to matrix form. The author believes that the high number of errors in some transcriptions may be due to both a language barrier and illegible participant handwriting. The transcription process quickly proved to be time-consuming and tedious, so the author recruited as many graduate students as possible to divide the transcription work and prepare the matrices for analysis.

Over twenty volunteers were recruited, making it difficult for the author to individually ensure that each volunteer knew exactly how to transcribe the data from the attribute list to the corresponding matrix in Microsoft Excel. This led to a much longer verification time than expected, as many matrices needed to be corrected, and the individual correcting a matrix could not be the same person that originally transcribed it.

Once the matrices were corrected, the preliminary analysis process began. The original intention was to eliminate some of the synonymous attributes in order to produce a more accurate data set. This was done through the combining of synonymous words as well as assimilating singular and plural attributes into one category for the referent “Do Not Touch with Wet Hands.” Unfortunately, this still resulted in over 600 attributes in the list, which was too large for Wordle (the word cloud program used) to analyze. Eliminating rows with reoccurring words in individual matrices and combining them into one matrix row further reduced the list to 66 attributes, which were analyzed by Wordle and are represented by the word cloud presented in Figure 1. Due to time constraints, at the time of final submission, only matrices for “Do Not Touch with Wet Hands” were analyzed via this method, but the preliminary word cloud for “Warning: Hot Exhaust” is included as well for reference (Figure 2).

Preliminary analysis utilizing the word lists retrieved from the word clouds along with previous expert panel data determined that the most often reported attributes from the semantic annotation method were generally the same as those reported by the expert panel. However, while the same general set of attributes when compared to the expert panel’s list for both referents was identified, there were still “important” attributes for the referent “Do Not Touch with Wet Hands” that did not individually represent the referent.

The author believes this to be a minor limitation as long as the general population is utilized for the creation of the symbol. The attributes for the referent "Warning: Hot Exhaust" appeared to be more representative of the referent, but the list was too large to create one symbol including all the attributes identified as important.

Overall, the untrained participants produced a larger and more varied list of attributes that differ only slightly from one another. Fuzzy analysis methods could be used to evaluate similar but not identical attributes (e.g. "hand" vs. "single hand"). This would eliminate many redundant terms and give a clearer picture of the truly important attributes. This would be necessary before further analysis could be done and a true comparison made with the data from this study.

Future Research

It is recommended to analyze this data set further in order to obtain more results. Analysis could be done among the 5, 10, and 15-second intervals to ascertain whether the amount of time allotted to viewing each drawing made a difference in 'core' attributes identified. This could give insight into whether the amount of time taken to view a symbol increases the level of comprehension or attention to detail. This would be helpful for situations such as road signs, where a person only has a certain amount of time to view and interpret the symbol's message.

Future research involving the general population would require a new method to prepare the data for analysis. Recording annotations on paper proved far more time consuming than anticipated because of the need to transcribe the handwriting, verify the terms, and consolidate the attributes into a master list. It is possible that the participants themselves could look at each other's attribute lists and condense them into what they feel are important attributes. Participants could also transcribe their attribute lists into matrices themselves, reducing the error accounted for by handwriting interpretation.

Conclusions & Recommendations

Overall, the design presented in this study appears to be a potentially effective technique for the creation of warning symbols. The preliminary results obtained are a good building block for future research, and the exploration of a general population method is important in gaining knowledge from individuals who will be interacting with the potential symbol on a daily basis. Since preliminary analysis indicated that both the expert panel and the general population identified the same set of attributes, it did not seem as though bias was introduced by previous knowledge of safety symbols in the expert panel method. Therefore, this study indicated that the expert panel method is feasible for symbol creation. However, in the general population method, the length of time required for the consolidation of data can be reduced by techniques mentioned above, thereby lowering resource costs and making this method less expensive than the fees required to pay three professional consultants to create a symbol. Based on the preliminary results of this study and comparing methods of the expert panel method and general population method, the author recommends utilizing the general population from start to finish as a method of warning symbol creation.

It is recommended that in future studies, the participants create their attribute lists and transcribe them into a matrix themselves. This is expected to significantly reduce the amount of time required to prepare the data for statistical analysis. Additionally, it is recommended that researchers investigate whether 5, 10, or 15 seconds of viewing time has any impact on attribute lists created by participants, and whether the resulting core attribute lists are similar across time intervals. It is also recommended to consider the importance of ratings on warning symbols. For example, it might be important to indicate on a symbol the severity of a hazard so that an individual can decide whether to proceed with caution or avoid a hazard entirely. ANSI Z535.3 (2011a) indicates that of the two messages that safety symbols generally convey, hazard description versus hazard avoidance, hazard description is best. This standard also recommends that symbols are portrayed as literally as possible in order to communicate the exact hazard and to facilitate an understanding of consequences associated with it (ANSI, 2011a). This information suggests that it may be important for a warning symbol to indicate exactly what will happen if an individual comes in contact with a hazard. Implementing a numbered system (for example, severity level 1-4) to accompany a warning symbol may assist in communicating the gravity of a given hazard, especially if every individual cannot recognize the hazard from the symbol alone. Future researchers should consider all of these things when creating a study using this method of symbol design.

References

Aggarwal, C. C. (2004). A human-computer interactive method for projected clustering. *Knowledge and Data Engineering, IEEE Transactions on, 16*(4), 448-460.

ANSI. (2011a). *Criteria for Safety Symbols*: American National Standards Institute Z535.3 - 2011.

ANSI. (2011b). *Environmental and Facility Safety Signs*: American National Standards Institute Z535.2 - 2011.

ANSI. (2011c). *Product Safety Signs and Labels*: American National Standards Institute Z535.4 - 2011.

Carnahan, B. J., & Dorris, N. (2004). *Identifying Relevant Symbol Design Criteria Using Interactive Evolutionary Computation.* Paper presented at the Genetic and Evolutionary Computing Conference (GECCO), Seattle.

Carnahan, B. J., Dorris, N. T., & Kuntz, L. A. (2005). Designing anthromorphic symbols using interactive evolutionary design. *Information Design Journal and Document Design, 13*(3), 179-190.

Carneiro, G., & Vasconcelos, N. (2004). Formulating Semantic Image Annotation as a Supervised Learning Problem. Unpublished Technical Report. University of California, San Diego.

Clark, D. R., Benysh, S. A. H., & Lehto, M. R. (2003). Design of Industrial Warnings. In W. Karwowski & W. S. Marras (Eds.), *Occupational Ergonomics: Principles of Word Design*. Boca Raton, FL: CRC Press.

Deppa, S. W. (2006). U.S. and inernational standards for safety symbols. In M. S. Wogalter (Ed.), *Handbook of Warnings* (pp. 477-486). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Dorris, N. (2004). *The application of interactive evolutionary computation (IEC) to the design of safety symbols.* Unpublished Dissertation, Auburn University, Auburn, AL.

Dorris, N., Carnahan, B., Orsini, L., & Kuntz, L. A. (2004). *Interactive evolutionary design of anthropomorphic symbols.* Paper presented at the Congress on Evolutionary Computation, 2004. CEC2004.

Dozier, G., Carnahan, B., Seals, C., Kuntz, L. A., & Fu, S. G. (2005a). An IDEA for design. Unpublished Technical Report.

Dozier, G., Carnahan, B., Seals, C., Kuntz, L. A., & Fu, S. G. (2005b). *An interactive distributed evolutionary algorithm (IDEA) for design.* Paper presented at the 2005 IEEE International Conference on Systems, Man and Cybernetics.

Edworthy, J., & Adams, A. (1996). *Warning Design: A Research Perspective*: CRC Press.

Egilman, D., & Bohme, S. R. (2006). A Brief History of Warnings. In M. S. Wogalter (Ed.), *Handbook of Warnings* (pp. 11-20). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Feinberg, J. (2014). Wordle [Computer software]. Retrieved from http://www.wordle.net.

Goldsworthy, R., & Kaplan, B. (2006). Exploratory evaluation of several teratogen warning symbols. *Birth Defects Research Part A: Clinical and Molecular Teratology, 76*(6), 453-460.

Hicks, K. E., Bell, J. L., & Wogalter, M. S. (2003). On the Prediction of Pictorial Comprehension. *Human Factors and Ergonomics Society Annual Meeting Proceedings, 47*, 1735-1739.

Huer, M. B. (2000). Examining perceptions of graphic symbols across cultures: Preliminary study of the impact of culture/ethnicity. *Augmentative and Alternative Communication, 16*(3), 180-185.

Isherwood, S. J., McDougall, S., J.P, & Curry, M. B. (2007). Icon Identification in Context: The Changing Role of Icon Characteristics With User Experience. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 49*, 465-476.

ISO. (2002). *Design principles for graphical symbols for use in safety signs*: International Organization for Standardization 3864-1: 2002.

ISO. (2003). *Graphical symbols -- Safety colours and safety signs -- Safety signs used in workplaces and public areas*: International Organization for Standardization 7010:2003

ISO. (2004). *Graphical Symbols -- Design principles for Product Safety Labels*: International Organization for Standardization 3864-2: 2004.

ISO. (2006). *Graphical Symbols -- Design principles for graphical symbols for use in safety signs*: International Organization for Standardization 3864-3: 2006.

ISO. (2007). *Graphical symbols - Test methods - Part 1: Methods for testing comprehensibility*: International Organization for Standardization 9186-1:2007.

ISO. (2008). *Graphical symbols - Test methods - Part 2: Method for testing perceptual quality*: International Organization for Standardization 9186-2:2008.

Kalsher, M. J., Wogalter, M. S., & Racicot, B. M. (1996). Pharmaceutical container labels: enhancing preference perceptions with alternative designs and pictorials. *International Journal of Industrial Ergonomics, 18*(1), 83-90.

Laughery, K. R. (2006). Safety communications: Warnings. *Applied Ergonomics, 37*(4), 467-478.

Laughery, K. R., & Wogalter, M. S. (2006). Designing Effective Warnings. *Reviews of Human Factors and Ergonomics, 2*, 241-271.

Lesch, M. F. (2005). A Semantic Relatedness Paradigm for Assessing Comprehension of Warning Symbols. *Human Factors and Ergonomics Society Annual Meeting Proceedings, 49*, 1795-1799.

Lesch, M. F., Rau, P.-L. P., Zhao, Z., & Liu, C. (2009). A cross-cultural comparison of perceived hazard in response to warning components and configurations: US vs. China. *Applied Ergonomics, 40*(5), 953-961.

Liu, L., Hoelscher, U., & Gruchmann, T. (2005). Symbol comprehension in different countries: experience gained from medical device area. *Mensch Und Computer*, 81-87.

Macbeth, S. A., & Moroney, W. F. (1994). Development and Evaluation of Automobile Symbols: The Focus Group Approach. Human Factors and Ergonomics Society Annual Meeting Proceedings, 38, 978.

Macbeth, S. A., Moroney, W. F., & Biers, D. W. (2000). Development and Evaluation of Symbols and Icons: a Comparison of the Production and Focus Group Methods. *Human Factors and Ergonomics Society Annual Meeting Proceedings, 44*, 327-329.

Macbeth, S. A., Moroney, W. F., & Biers, D. W. (2006). Developing Icons and Symbols: A Comparison of Two Methods. *Ergonomics in Design, 14*, 14-18.

Mayhorn, C. B., & Goldsworthy, R. C. (2007). Refining teratogen warning symbols for diverse populations. *Birth Defects Research Part A: Clinical and Molecular Teratology, 79*(6), 494-506.

McDougall, S. J., Curry, M. B., & de Bruijn, O. (1999). Measuring symbol and icon characteristics: norms for concreteness, complexity, meaningfulness, familiarity, and semantic distance for 239 symbols. *Behav Res Methods Instrum Comput, 31*(3), 487-519.

Parmee, I. C., Abraham, J. A. R., & Machwe, A. (2008). User-Centric Evolutionary Computing: Melding Human and Machine Capability to Satisfy Multiple Criteria. In *Multiobjective Problem Solving from Nature* (pp. 263-283).

Piper, A. K., Boelhouwer, E. J., & Davis, J. (2008). Using hand drawn images to determine warning symbol design parameters within interactive evolutionary computation software. *Human Factors and Ergonomics Society Annual Meeting Proceedings, 52*, 1708-1712.

Piper, A. K. (2010). *Participatory design of warning symbols using distributed interactive evolutionary computation*. (Unpublished doctoral dissertation). Auburn University, Auburn, AL.

McDougall, S. J., Curry, M. B., & de Bruijn, O. (1999). Measuring symbol and icon characteristics: norms for concreteness, complexity, meaningfulness, familiarity, and semantic distance for 239 symbols. *Behav Res Methods Instrum Comput, 31*(3), 487-519.

McDougall, S. J., de Bruijn, O., & Curry, M. B. (2000). Exploring the effects of icon characteristics on user performance: The role of icon concreteness, complexity, and distinctiveness. *Journal of Experimental Psychology: Applied, 6*(4), 291-306.

OSHA. (1996). *Specifications for accident prevention signs and tags* (Vol. 29.1910.145): Occupational Safety and Health Administration.

Ringseis, E. L., & Caird, J. K. (1995). The Comprehensibilitly and Legibility of Twenty Pharmaceutical Warning Pictograms. *Human Factors and Ergonomics Society Annual Meeting Proceedings, 39*, 974-978.

Roberts, N. J., Mohamed, Z., Wong, P.-S., Johnson, M., Loh, L.-C., & Partridge, M. R. (2009). The development and comprehensibility of a pictorial asthma action plan. *Patient Education and Counseling, 74*(1), 12-18.

Russo, P., & Boor, S. (1993). *How fluent is your interface?: designing for international users*. Paper presented at the Proceedings of the INTERACT '93 and CHI '93 conference on Human factors in computing systems.

Schuler, D., & Namioka, A. (1993). *Participatory design: principles and practices*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Sloan, G., & Eshelman, P. (1981). The development and evaluation of pictographic symbols. *Human Factors and Ergonomics Society Annual Meeting Proceedings, 25*, 198-202.

Smith-Jackson, T. L., & Wogalter, M. S. (2006). Methods and procedures in warning research. In M. S. Wogalter (Ed.), *Handbook of Warnings* (pp. 23-33). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Turnbull, D., Liu, R., Barringon, L., & Lanckrie, G. (2007). *A game-based approach for collecting semantic annotations of music.* Paper presented at the International Symposium on Music Information Retrieval.

Vasconcelos, N., & Lippman, A. (2000a). *A Probabilistic Architecture for Content-based Image Retrieval.* Paper presented at the IEEE Conference Computer Vision and Pattern Recognition, Hilton Head, North Carolina.

Vasconcelos, N., & Lippman, A. (2000b). *A Unifying View of Image Similarity.* Paper presented at the 15th International Conference on Pattern Recognition, Barcelona, Spain.

Wisniewski, E. C., Isaacson, J. J., & Hall, S. M. (2007). Development and Evaluation of Safety Symbols for Landscaping Products. *Human Factors and Ergonomics Society Annual Meeting Proceedings, 51*, 1109-1113.

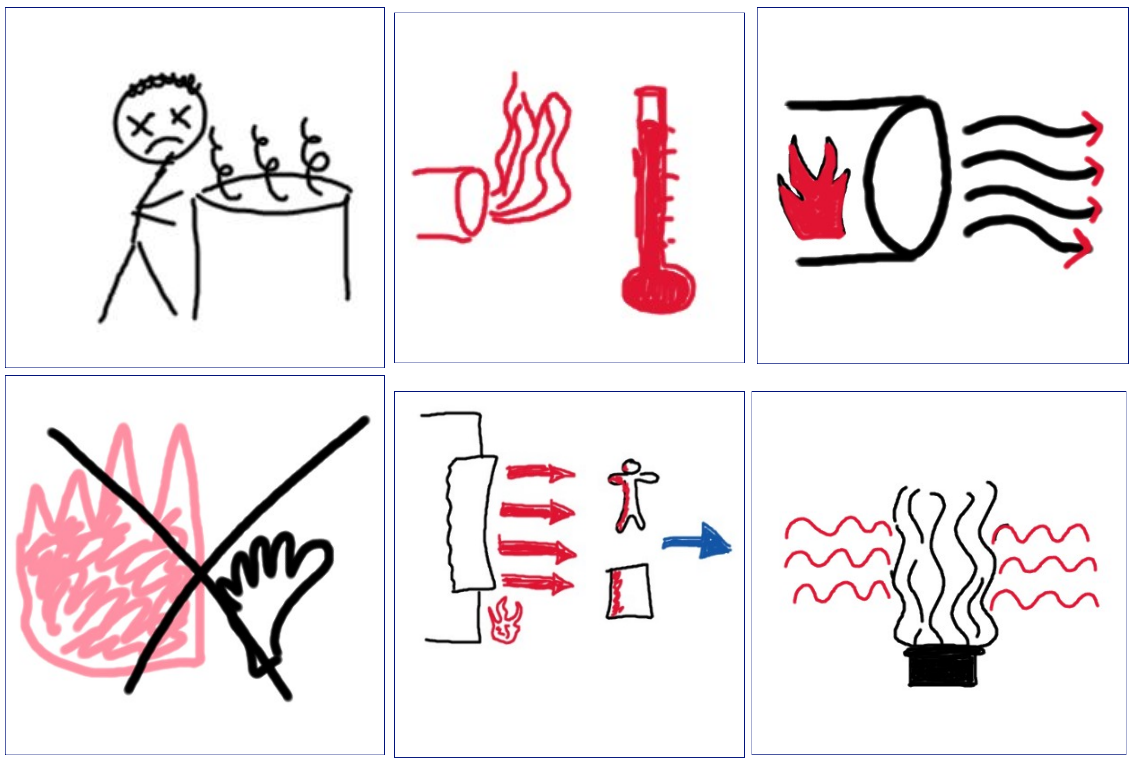
Wogalter, M. S. (2006). Purposes and Scope of Warnings. In M. S. Wogalter (Ed.), *Handbook of Warnings* (pp. 3-9). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Wogalter, M. S., Silver, N. C., Leonard, S. D., & Zaikina, H. (2006). Warning symbols. In M. S. Wogalter (Ed.), *Handbook of Warnings* (pp. 159-176). Mahwah, New Jersey: Lawrence Erlbaum Associates.

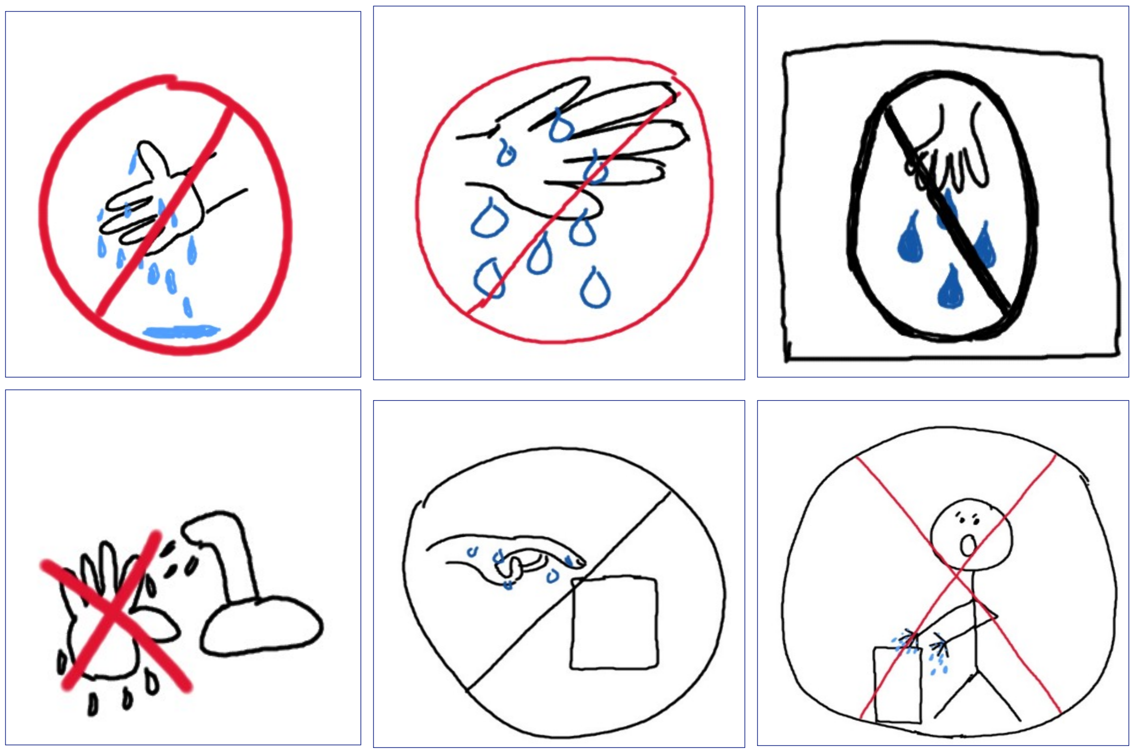
Wolff, J. S. (1995). A Study of the Effect of Context and Test Method in Evaluating Safety Symbols. Georgia Institute of Technology, Atlanta.

Young, S. L., & Wogalter, M. S. (2001). Predictors of pictorial symbol comprehension. *Information Design Journal, 10*, 124-132.

Appendices



Appendix 1. Sample participant drawings for referent “Warning: Hot Exhaust”



Appendix 2. Sample participant drawings for referent “Do Not Touch with Wet Hands”

Sketch #	Components of Sketch

Appendix 3. Semantic annotation transcription sheet

Attributes	Sketch 1	Sketch 2	Sketch 3	Sketch 4	Sketch 5	Sketch 6	Sketch 7
1 Hand	1			1		1	
2 Black X's							
2 brackets							
2 hands		1					
2 lightning bolts							
3 hands							
3 lightning bolts							
Acid							
Arrow							
Arrows			1				
Beaker							
Black circle	1						
Black line							
Blue circle							
Blue line							
Blue water droplets				1			
Box					1		
Button							
Channel							
Circle							
Computer							
Cord							
Cube							
Electricity			1				
Electrocuted person			1				
Faucet		1					1
Finger pointing							
Finger pressing							
Frightened face							
Green circle							
Ground							
Hair dryer			1				
Hand pressing buttons					1		
Hands			1				1

Appendix 4. Sample transcribed matrix