Fall 2023

The Right Sign for the Right Time: A Look at the Effectiveness of Different Types of Signage

Samantha Emily Feldstein
Bard College

The Right Sign for the Right Time: A Look at the Effectiveness of Different Types of Signage

Senior Project Submitted to

The Division of Science, Math, and Computing

of Bard College

by Samantha Feldstein

Annandale-on-Hudson, New York

December 2023

Table of Contents

Abstract

Warning signs are present in the day to day lives of most individuals. While there has been past research into what makes a warning sign more or less effective, there is still no clear answer as to what the most effective warning sign looks like. This study looks at specifically the words being used on a sign and the sign's color to determine which has a larger impact on the hazard perception of the sign. An online study was conducted in which participants were shown different signs and asked to rate how hazardous of a situation they thought each sign represents. Participants were shown eight different stimuli, ones with just the color red or blue, ones with just the word “danger” or “notice”, and ones with each combination of those colors and words. It was hypothesized that participants will find signs with the word “danger” more hazardous than signs with the word “notice”, find signs with the color red more hazardous than signs with the color blue, and finally that the color of the sign has a greater impact than the word on the individual’s perception of it’s hazard level. However, only the first two hypotheses could be supported with the data collected. Applications of the results of the study and directions for future study are discussed.

Introduction

We encounter many different types of warning signs in our day to day lives. Regardless of where the sign is and what it is indicating it is imperative that warning signs are able to effectively convey their message in a noticeable and accurate manner. This is why when it comes to signs indicating workplace hazards, The Occupational Safety and Health Administration (OSHA) has strict guidelines in terms of color, wording, inclusion of symbols, and a number of other factors. These guidelines are intended to ensure that all signage is accurately conveying hazard information to workers. However, while these guidelines are very strict, they place the same level of importance on all the different aspects of the sign. This raises the question, are all elements of a sign created equal? Do they all have the same impact on an individual's perception of the hazard they represent or do some elements have more of a sway than others?

It is important to note that there are more than one type of sign used to communicate hazards. A 2020 study by Jun and colleagues looked at three distinct types of signage: warning signs (signs that indicate the danger present in an environment), prohibitory signs (signs that only indicate what not to do), and mandatory signs (signs that only indicate what has to be done). Researchers found that participants reported warning signs as being indicative of higher levels of hazard than prohibitory or mandatory signs. Furthermore, the OSHA guidelines refer specifically to warning signs. For these reasons this study will look at warning signs

specifically. This is not to say that similar research cannot or should not be done with prohibitory or mandatory signs. That is simply outside the scope of this experiment.

There has been a number of past research studies focusing on what makes a warning sign more or less effective. One meta-analysis showed that the presence of warning signs does increase safe behaviors especially when the warning signs have increased vividness enhancing characteristics (Argo et al., 2004). These “vividness enhancing characteristics” included font size, the color of the sign, spacing, the level of specificity, and the presence of symbols. However, even these researchers stated that “The ideal combination of warning factors has yet to be identified” (Argo et al., 2004).

Another study looked at the presence of the “hazard alert symbol”, an exclamation point in a triangle, on signs to see if that impacted individuals’ perception of the hazard level of a situation (Jensen et al., 2003). This study found that the presence of the symbol did increase the perceived hazard of the situation, supporting the finding of the meta analysis that vividness enhancing characteristics, in this case the presence of a symbol, can increase the perception of a warning sign. However, this study only looked at the absence or presence of this symbol. It used only yellow signs that all had the word “caution” to decrease the number of variables in the study.

Color

While it is clear that a number of factors play a role in the efficacy of warning signs, there is also past research that indicated the background color of the sign specifically has an impact on its ability to alert individuals to hazards. In a study by Yuan et al. (2021) participants were shown a number of different hazard signs with varying background colors. Of the three colors, white, yellow, and blue, yellow signs were more likely to alert participants to hazards than the other colors. From these results researchers were able to conclude that even if an individual's attention is not readily directed towards the sign they are still able to identify the differences in sign color and the differences in hazard information associated with each color. The study run by Yuan et al. highlights the impact of sign color but did not look at the colors red or blue specifically, indicating that there is reason for a difference in perception but that difference has not been found between these specific colors.

Beyond knowing that the color of a sign can impact how it conveys hazard messages, there is also evidence to suggest that the color red specifically would facilitate a higher perceived hazard level from participants. Using the known attention capturing hierarchy of colors, a 2017 study from Blizzard and colleagues looked at the potential for that same hierarchy to be extended to higher levels of processing. Using a stop signal task researchers found that red stop signals facilitated more response inhibition than green stop signals. This indicates that response inhibition is sensitive to differences in color salience, including the color red.

Not only does the general color of a sign matter, there is evidence to support that getting the specific color of the sign is important as well. Researchers looked into how the human brain categorizes colors, specifically ones that are similar to each other (Boynton, R., 1989). Specifically, this study looked at recall of colors after 10 s had passed. Participants were shown a color, then waited 10 s before being shown either the same color or one that was only a few degrees different. The study found that "...the greater the degree of categorical difference between two colors…the less likely they are to be incorrectly called 'same' in a task that forced categorization" (Boynton, R., 1989). When it comes to warning signs, an area in which distinction between categories is highly important, these findings indicate that the color of the sign not only matters, but it matters how different two signs colors' are from each other. If signs that are intended to be the same background color are too different they may be categorized as different hazard levels leading to misinterpretation. On the other hand, if two signs that are intended to be different categorically have colors that are too similar they may be given the same hazard rating, again leading to potentially dangerous misinterpretations. This supports our study's use of two very distinct colors, red and blue, as these are colors that have enough degrees of difference to not be mistaken for each other.

When it comes to red specifically there is physiological evidence as to why this color may indicate a higher level of hazard. According to Birren in his 1961 book *Color Psychology and Color Therapy,* the color red will stimulate the autonomic

nervous system. The autonomic nervous system is comprised of both the parasympathetic and sympathetic systems. When the parasympathetic system is aroused, as Birren says is the care when an individual sees the color red, there is an increase in production of adrenaline which can trigger the fight or flight instinct. This fight or flight instinct is likely to provoke a hazard response when viewing a red sign. While the color red arouses the autonomic nervous system the color blue relaxes it. This means that while red signs trigger fight or flight responses potentially leading to higher perceived hazard levels, the color blue will have the opposite effect, calming down the body and likely producing far lower levels of perceived hazard. Based on the difference between how the colors red and blue interact with the autonomic nervous system, there is physiological evidence to support red signs having higher levels of perceived hazard than the blue signs.

Word

It has been shown that people connect different words with different colors. In a survey, when asked what color they most associated with the word fear, 41% of individuals responded with the color red, more than any other color (Hallock, J., 2003). Simultaneously, no individuals chose the color blue to represent the word fear. Based on the results of this study it makes sense that individuals would associate red signs with fear, more so than blue signs. This fear association could potentially be linked to higher hazard rating among red signs as opposed to blue signs.

Past research has also shown that changing the signal word on signs can change the level of hazard they are perceived to represent (Zhu et al., 2020). This study looked at two variables, the warning word on the sign and the sign's shape. Using EEG data, scientists were able to determine that there is a "high hazard" combination of shape and warning word that gives the sign the ability to attain high priority in an individual's cognitive processing. The combination of an upright triangle shape along with "high hazard words" was the most able to capture an individual's attention according to the EEG data. However, while this study does give insight into the ability of the word on a sign to impact an individual's hazard perception, it is not helpful when it comes to specifically the words "danger" and "notice". The study by Zhu and colleagues was run in China and as a result all the words used on the signs were in Chinese. The resulting language differences make it impossible to generalize the ranking of different words found in this study to the English language. Instead, the basic finding of a change in word resulting in a change in hazard perception is what must be taken into consideration.

A study by Heller et al. (2007) looked at 17 different signal words and how they were rated by individuals in relation to each other. Researchers found that there were multiple dimensions to the way participants rated the signal words. They rated the words in a way that was concurrent with the hazard ratings that could be found in literature, based on if the word indicates specifically a hazardous situation or just that one needs to draw attention, and based on the level of specificity of the signal word.

Consistent with the results of this study the signal words being used in the current study, “danger” and “notice”, are different across all three of these dimensions, indicating that there should be a significant difference between the hazard ratings of these two words.

Interactions

Ours is not the first study to look at the interaction between color and signal words. There is a previous study that looked at the combination of three different signal words with four different background colors (Chapanis, A., 1994). This study found that the word “danger” was rated the most hazardous when compared with “warning” and “caution”. The color red was also rated the most hazardous compared to the colors white, yellow, and orange. Importantly to this study, Chapanis found that the word danger combined with the color red was consistently indicative of the highest hazard rating.

In 1995 a study done by Braun and Silver looked at how college students rated the hazard level of different signal words printed in different colors. Researchers found that both the word and the color it was printed in had a role in how hazardous of a rating the stimulus received. This study indicates that both color and word have an impact on the hazard rating of a sign, however, in this case researchers were looking at the color of the word, not the color background it was presented on. This difference is important as OSHA regulations focus more on the background color of the sign than the color of the word. Despite this key difference, the Braun and Silver

study does indicate that there is potential for the combination of background color and word to have an impact on the hazard ratings of signs.

The current study examines the impact of sign color and word used on perceived hazard levels. It also looks at the interaction between these two factors and if one factor has more of an impact on perceived hazard levels than another.

Methods

Setting

All participants in this study were recruited from United States of America servers through Prolific.com. Prolific is an online participant recruiting platform that matches researchers with participants, allowing for more diverse sample groups in online studies. The study itself was designed, programmed, and hosted on the online platform Gorilla. All data was also collected through Gorilla. Gorilla was chosen to host the study due to its ability to easily collect reaction time data. Due to the importance of language and color recognition, participants were screened to ensure English was their first language and they were not diagnosed with any form of colorblindness. Researchers set out to include 100 participants. This number was determined by budget and funding limitations. This study and all procedures were approved by the Bard University Institutional Review Board (Appendix A).

Stimuli

This study looks at the colors blue and red. According to the OSHA guidelines, (Creative Safety Supply, 2016) the color red is to be used for signs representing the highest potential hazard level and blue is to be used on signs conveying important information that is not hazard related. These two colors were chosen because they are at opposite ends of OSHA's spectrum which will hopefully allow for the greatest difference in perceived hazard levels. For the color red the hex code #AF241C was used and for the color blue the hex code #005983. These codes were taken from the

color matching website *My Perfect Color* which was used as it had color codes for both colors that were visually consistent with the colors used in the guidelines and were specifically noted to be OSHA colors. It was important that both codes come from the same website to eliminate potentially confounding variables around the sourcing of the colors.

The same factors were used to choose the words “danger” and “notice”. According to OSHA guidelines, the word “danger” is to be used for high hazard situations and “notice” is to be used for important information that is not hazard related. No font is specified in OSHA guidelines so the font Arial was used as it is common and not distracting to viewers.

Procedure

Piloting

Before official participant enrollment began, three individuals participated in pilot testing. These individuals were selected based on ease of access for the researcher. These three participants were asked to complete the full study and after completion were asked their thoughts on the study and any questions they had while participating. The first issue pilot participants described was that the speed of the trials was too fast. It is important to keep the speed of the trials fast so that the study tests a participant’s first reactions to the signs. The decision was made that rather than slowing the speed of the trials to add a disclaimer to the instruction portion of the trial to inform participants of the rapid nature of the study.

Pilot participants also expressed some concerns about the rating scale. The faces were moved from being below the line to above the line. It was made clearer in the instructions that participants needed to click directly on the line, not just near it. The size of the line on the scale was also increased. Originally the most hazardous end of the scale was on the left with the least hazardous end on the right. Pilot participants were confused by that orientation, so it was changed to the least hazardous face on the left of the scale and the most hazardous on the right side.

The data collected from the pilot participants was used to ensure that the study's data collection was working properly and confirmed that the coding for all aspects of the study worked as intended.

Participant Enrollment

Prior to enrollment participants were first asked to read and digitally sign an informed consent form (Appendix B). This form told participants why they had been chosen to participate in the study, what their participation would entail, any potential costs or benefits to completing the study, and gave them information on who to contact if they have any questions about their participation. At the bottom of the form participants were asked to digitally sign by checking boxes asserting that they willingly volunteered to participate in the study and that they were over the age of 18. If they agreed to participate and were over the age of 18 they were then directed to a demographics survey (Appendix C), which asked participants their age, the current employment status, current education level, and if the majority of their time working

was spent in the United States. They were also asked again to confirm that English is their first language and they have not been diagnosed with any form of colorblindness.

Instruction

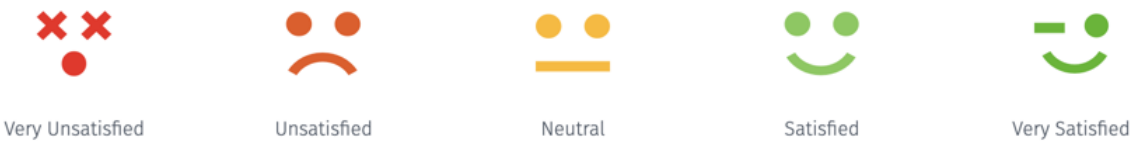
Next, participants were given instructions on how to complete the study. They were informed that they would be shown a sign and then given 5 s to rate it. We included a warning that the study will move quickly. Then participants were shown the scale they would be using to rate the different stimuli (Figure 1) and given a chance to practice using the scale. This scale is presented to participants as a gray line they click along to choose their rating, along with faces, one at the midpoint and one at each endpoint. The face on the left endpoint, representing the least hazardous rating, is a smiling face. The midpoint is represented by a face with a straight line mouth. The face on the right endpoint, representing the most hazardous rating, is a face with Xs for eyes and a circle mouth.

Figure 1. Rating Scale

The images used in the scale were taken from an article on the website QuestionPro that discussed the creation of likert scales. They were taken from an

example of a likert scale which contained five color images of different faces (Figure 2).

Figure 2. Reference Scale



The three faces were chosen based on their ability to accurately represent hazard, something most other face scales did not have. The images selected were those labeled very unsatisfied, neutral, and satisfied in the original scale. Those faces were then converted to black and white from their original colors. Due to color being one of the variables under study it was important that no color associations be present in the rating scale.

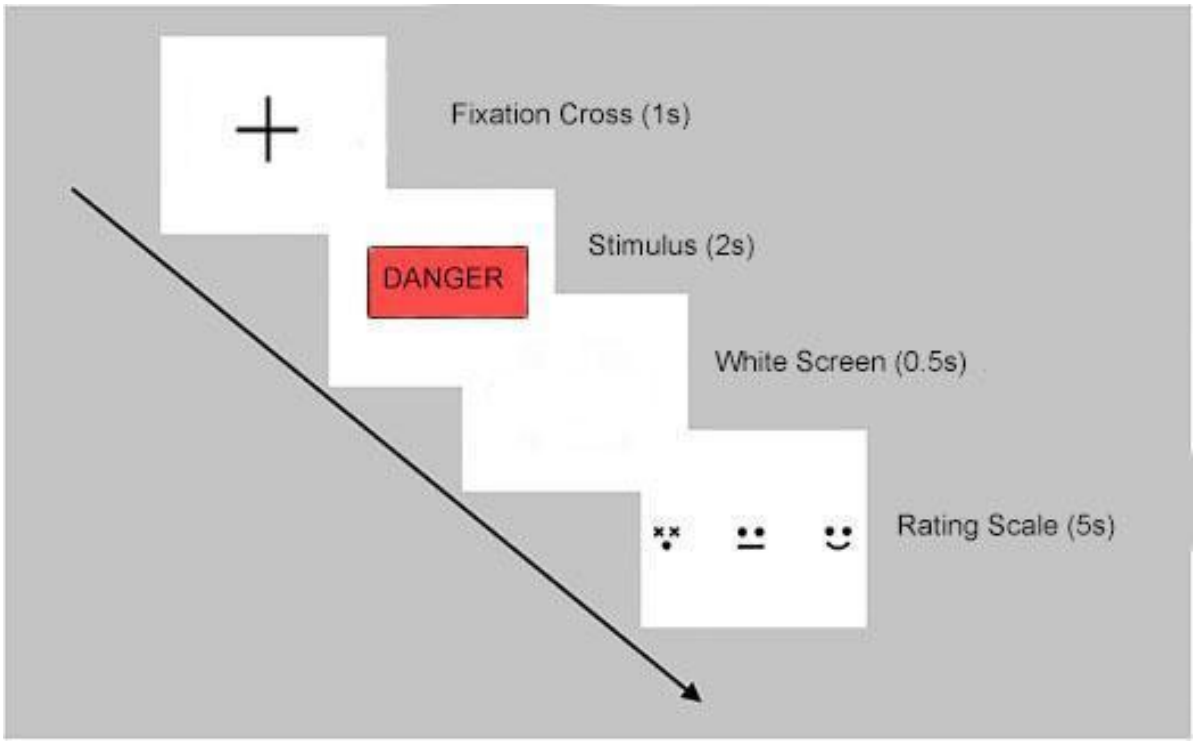
Participants were also informed they would be seeing and rating the same stimulus multiple times during the study as this was a question that came up multiple times during the pilot testing of the study. Finally, participants were told that if they did not want to rate a sign for any reason they could choose to skip it.

Intervention

Participants were first shown a fixation cross for 1 s. They were then shown one of eight possible stimuli (Appendix D) for 2 s. The stimuli are: a blue sign with no

text, a red sign with no text, a white sign with the word “notice”, a white sign with the word “danger”, a blue sign with the word “notice”, a blue sign with the word “danger”, a red sign with the word “notice”, and a red sign with the word “danger”. Then there was a blank screen for .5 s before participants were shown the rating scale and asked to rate the given stimulus. After participants clicked the continue button or 5 s passed, whichever came first, a second blank screen was shown for .5 s as a spacer between each trial (Figure 3). After being shown each stimulus five times, for a total of 40 trials, participants were shown a screen informing them they had reached the end of the study.

Figure 3. Study Design Layout



Debrief

After completing the study, participants were shown the debriefing form (Appendix E). This form includes a description of the study's aim and the researchers' hypothesis as well as contact information if the participant has any questions. After confirming that they read the debriefing form participants were directed back to Prolific where they were compensated $1.82 for their participation.

Outcomes

All the data in this study was downloaded directly from Gorilla. The first data that was collected was the demographic information from the demographic survey. This included age, current occupation status, current education level, and if their previous work experience was primarily within the United States. Next, data was collected on the ratings of each stimulus. While participants do not see any numeric value attached to their rating, the programming of the study allowed each point on the rating line to be connected to a number 1-100. When participants choose their rating point on the scale the number associated with that point is recorded. Those numbers are the values used for analysis. Simultaneously, Gorilla was able to collect data on how quickly participants completed their ratings.

Analysis

Every participant was shown each stimulus five different times. The average of their scores were then calculated and that value used in analysis. The intention of this design was to mitigate any priming factors that may come from an individual seeing

each stimulus only one time in a certain order. Due to the rapid pace of the test, this also allowed individuals to miss one rating of a given stimulus and still allow their results to be used in the final analysis. The only time participant's scores were excluded from analysis was if they missed three or more individual ratings of a given sign or if their final averaged score was more than two standard deviations away from the mean. Both these exclusion criteria were included in the study's preregistration.

For the main analysis a repeated measures ANOVA was run with the rating data from the stimuli that had both color and signal word. This analysis was used due to the number of variables and the within subject design of the study. After this initial analysis post-hoc tests were run to look at the interactions between specific combinations of word and color. We hypothesize that the color red will receive higher hazard ratings than the color blue and that the word "danger" will receive higher hazard ratings than the word "notice". We also hypothesize that the color of a sign will matter more to the hazard rating of the sign than the word on the sign. It is proposed that this will manifest by the red sign with the word "notice" being rated as more hazardous than the blue sign with the word "danger".

Results

Participants

Data was collected from 100 participants. 12 had responses more than two standard deviations away from the mean and thus were excluded. The mean age of participants was 37.8 years (Table 1). The frequency results of the demographic survey are listed below (Tables 2-4). 60.9% of participants were employed, with all who chose to specify saying they had spent the majority of their time working in the United States. 84.1% of participants also had at least some post high school education.

Table 1. Descriptive Statistics for Age Data

Descriptives

	N	Missing	Mean	Median	SD	Minimum	Maximum
Age	88	0	37.8	35.5	13.2	20	77

Table 2. Frequencies of Employment Status

Employment Status	Counts	% of Total
Employed	53	60.9 %
Retired	3	3.4 %
Self-Employed	2	2.3 %
Student	2	2.3 %
Unemployed	27	31.0 %

Table 3. Frequencies of Work Experience in US

Work Experience in US	Counts	% of Total
N/A	3	3.4 %
Yes	85	96.6 %

Table 4. Frequencies of Education Level

Education Level	Counts	% of Total
High school diploma/GED/equivalent	13	14.8 %
Some post high school education	16	18.2 %
Associate degree	13	14.8 %
Trade school degree	5	5.7 %
Bachelor's degree	28	31.8 %
Graduate degree	11	12.5 %
Doctorate or professional degree	1	1.1 %
N/A	1	1.1 %

No significant interactions were found between demographic data and rating results in between subjects analyses. The p value for the between subjects effect of participant employment status was p =0.178 (Table 5).

Table 5. Between Subjects Effect of Employment Status

Between Subjects Effects

	Sum of Squares	df	Mean Square	F	p
Employment Status	1835	4	459	1.62	0.178
Residual	23254	82	284		

Note. Type 3 Sums of Squares

The p value for the between subjects effect of participant work experience in the United States was p =0.467 (Table 6).

Table 6. Between Subjects Effect of Work Experience in the United States

Between Subjects Effects

	Sum of Squares	df	Mean Square	F	p
Work Experience	156	1	156	0.533	0.467
Residual	25072	86	292		

Note. Type 3 Sums of Squares

The p value for the between subjects effect of participant education level was p=0.098 (Table 7).

Table 7. Between Subjects Effect of Education Level

Between Subjects Effects

	Sum of Squares	df	Mean Square	F	p
Education Level	3442	7	492	1.81	0.098
Residual	21786	80	272		

Note. Type 3 Sums of Squares

Rating Results

Color

The repeated measures ANOVA indicated a significant effect of color on the perceived hazard level rating of the stimuli ($p < .001$) (Table 7). A look at the estimated marginal means shows that red signs had a mean rating of 78.2 (SE=0.925) while blue signs had a mean rating of 52.2 (SE=1.344) (Figure 4, Table 8).

Figure 4. Estimated Marginal Means for Color in Stimul

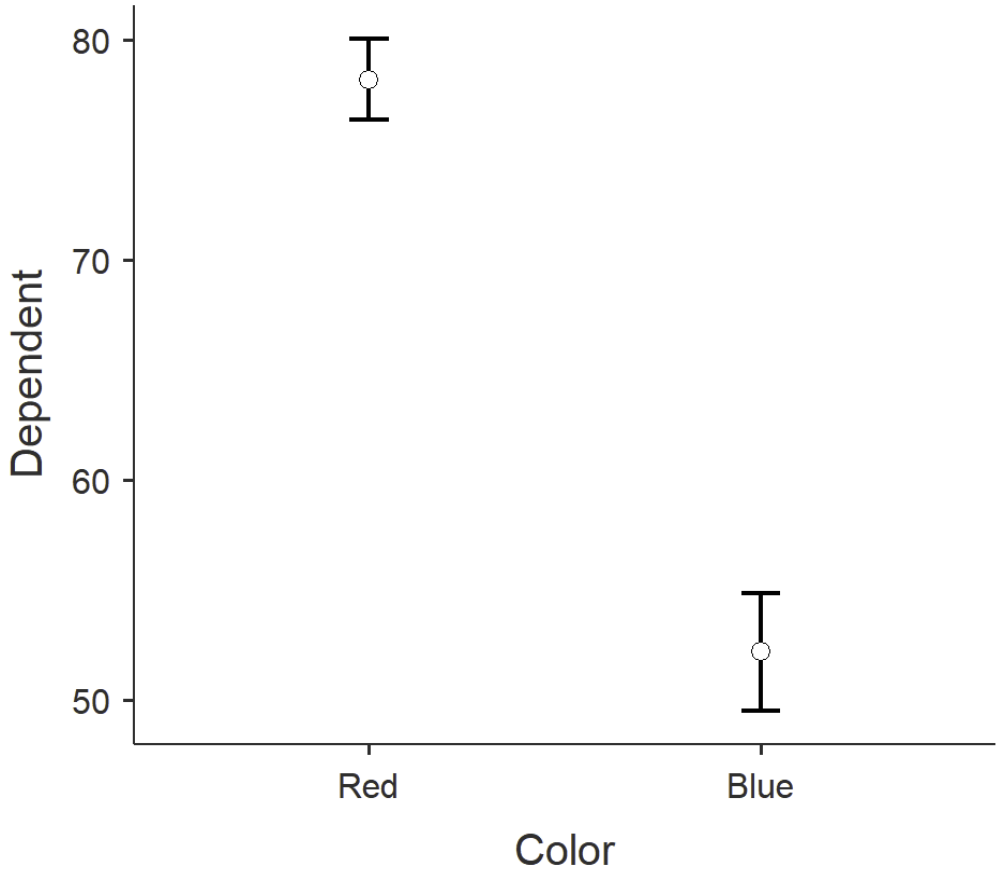


Table 8. Estimated Marginal Means for Color in Stimuli

Estimated Marginal Means - Color

			95% Confidence Interval	
Color	Mean	SE	Lower	Upper
Red	78.2	0.925	76.4	80.0
Blue	52.2	1.344	49.5	54.9

Word

The repeated measures ANOVA also indicated a significant effect of signal word on the perceived hazard level rating of the stimuli ($p < .001$) (Table 7). A look at the estimated marginal means shows that signs with the word “danger” had a mean rating of 78.4 (SE=1.16) while signs with the word “notice” had a mean rating of 52.0 (SE=1.20) (Figure 5, Table 9).

Figure 5. Estimated Marginal Means for Word in Stimuli

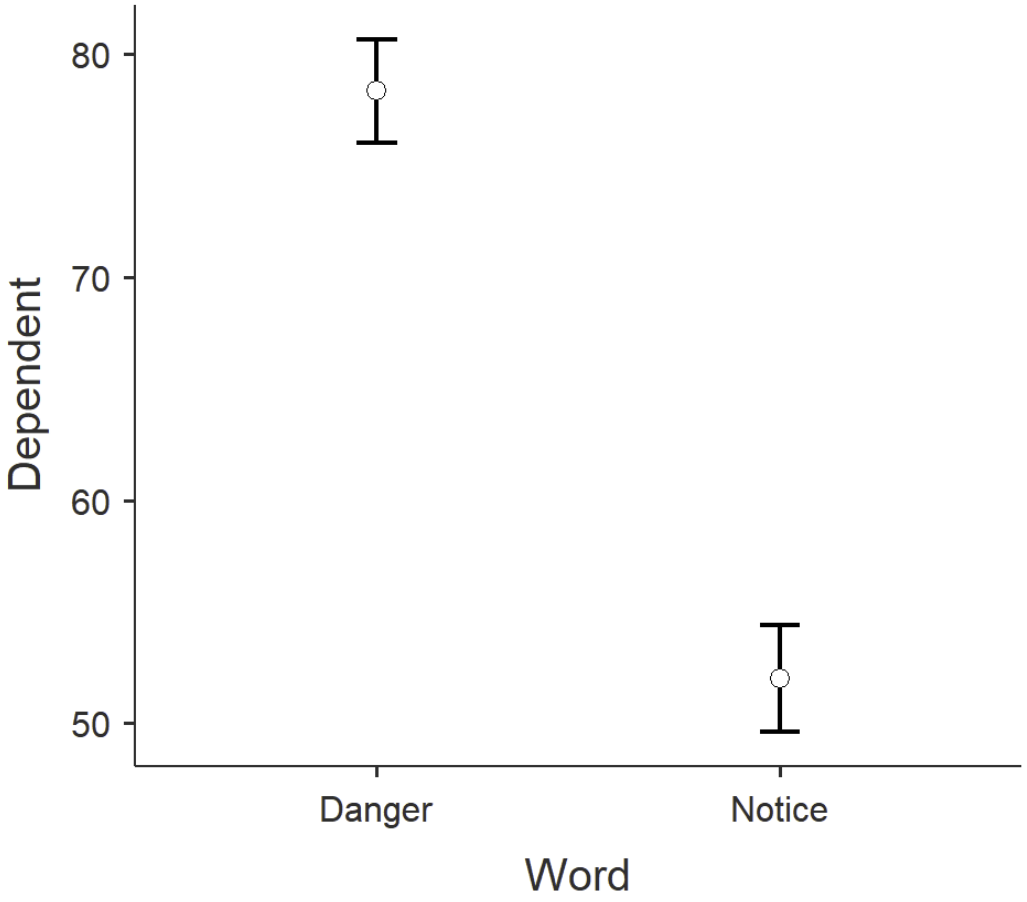


Table 9. Estimated Marginal Means for Word in Stimuli

Estimated Marginal Means - Word

			95% Confidence Interval	
Word	Mean	SE	Lower	Upper
Danger	78.4	1.16	76.1	80.7
Notice	52.0	1.20	49.6	54.4

Color and Word Interaction

Finally, the repeated measures ANOVA indicated a significant effect in the interaction between color and word of the stimulus (p =0.041) (Table 10).

Table 10. Results of Repeated Measures ANOVA

Within Subjects Effects

	Sum of Squares	df	Mean Square	F	p
Color	59478	1	59477.6	333.43	< .001
Residual	15519	87	178.4		
Word	61021	1	61021.4	300.77	< .001
Residual	17651	87	202.9		
Color ✻ Word	278	1	278.1	4.30	0.041
Residual	5631	87	64.7		

Note. Type 3 Sums of Squares

Post Hoc analysis was performed to determine which combinations of word and color had significant differences in their ratings. The analysis revealed that all combinations with the exception of one combination, a red sign with the word “notice” and a blue sign with the word “danger”, had significantly different ratings (Table 11). While all other combinations had significant differences, the combination with the largest mean difference was the red sign with the word “danger” and the blue sign with the word “notice” (MD =52.331). All combinations had p values of <0.001.

Table 11. Results of Post Hoc Comparisons

Post Hoc Comparisons - Color ✻ Word

Comparison									
Color	Word		Color	Word	Mean Difference	SE	df	t	p_{tukey}
Red	Danger	-	Red	Notice	24.555	1.37	87.0	17.888	< .001
		-	Blue	Danger	24.220	1.63	87.0	14.886	< .001
		-	Blue	Notice	52.331	1.90	87.0	27.487	< .001
	Notice	-	Blue	Danger	−0.335	2.25	87.0	−0.149	0.999
		-	Blue	Notice	27.776	1.70	87.0	16.373	< .001
Blue	Danger	-	Blue	Notice	28.111	2.05	87.0	13.721	< .001

Order of Ratings

When looking at the descriptives of the rating data for the stimuli with both word and color (Table 12), It is possible to see in which order, from most to least hazardous, participants rated the different stimulus. In order from most to least hazardous the signs were rated:

1. A red sign with the word “danger”
2. A blue sign with the word “danger”
3. A red sign with the word “notice”
4. A blue sign with the word “notice”

This data is also shown graphically in Figure 6.

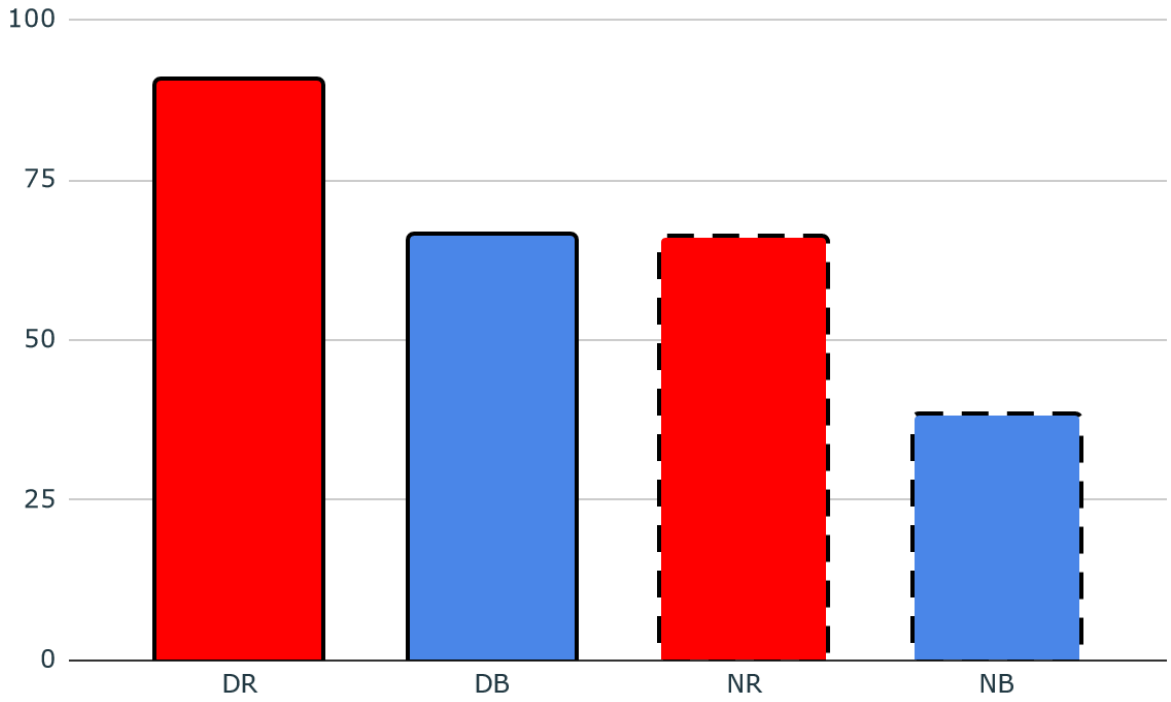
Table 12. Descriptives of Color and Word Stimuli Ratings

Descriptives

	DR Rating	DB Rating	NR Rating	NB Rating
N	88	88	88	88
Missing	0	0	0	0
Mean	90.5	66.2	65.9	38.1
Median	92.9	62.9	66.2	41.6
Standard deviation	9.15	16.4	12.2	15.2
Minimum	64.6	36.8	41.4	7.40
Maximum	99.8	99.8	93.2	67.0

Note: DR=red sign with the word “danger”, DB=blue sign with the word “danger”, NR=red sign with the word “notice”, and NB=blue sign with the word “notice.

Figure 6. Graph of Rating Means



Note: DR=red sign with the word “danger”, DB=blue sign with the word “danger”, NR=red sign with the word “notice”, and NB=blue sign with the word “notice”

Reaction Time

No significant effects were found in relation to reaction time data. The table below shows the descriptive data for reaction times for the four stimuli with both word and color (Table 13). There were also no significant differences between the reaction times of different stimuli. The mean rating times ranged from 2579 ms to 2886 ms.

Table 13. Descriptives of Color and Word Stimuli Reaction Times

Descriptives

	DR Time	DB Time	NR Time	NB Time
N	88	88	88	88
Missing	0	0	0	0
Mean	2579	2886	2649	2762
Median	2354	2620	2490	2464
Standard deviation	846	1603	876	1101
Minimum	1364	1525	1309	1394
Maximum	5009	15503	5004	8448

Note: DR=red sign with the word “danger”, DB=blue sign with the word “danger”, NR=red sign with the word “notice”, and NB=blue sign with the word “notice”. All values are in milliseconds.

Completion Time

Researchers expected the study to take participants about 15 minutes to complete. Instead, according to Prolific’s data, the study took participants a mean time of 6 minutes 43 s with times ranging from 27 minutes 33 s to 4 minutes 48 s.

Excluded Data

Before participant’s data points were excluded based on the preregistered exclusion criteria the raw data was run through a repeated measures ANOVA. This preliminary analysis also found that the color of the sign, the signal word on the sign,

and the interaction between the color and word of the sign were all significant (Table 14).

Table 14. Results of Repeated Measures ANOVA Before Exclusion of Data

Within Subjects Effects

	Sum of Squares	df	Mean Square	F	p
Color	60498	1	60497.8	227.10	< .001
Residual	26106	98	266.4		
Word	72368	1	72368.2	263.13	< .001
Residual	26953	98	275.0		
Color ✻ Word	324	1	323.8	4.70	0.033
Residual	6753	98	68.9		

Note. Type 3 Sums of Squares

Discussion

This study tested three primary hypotheses and five secondary hypotheses. The primary hypotheses were:

1. Signs with the color red will be rated more hazardous than signs that were blue
2. The word “danger” will be rated more hazardous than the word “notice.”
3. The color of the sign will be more important than the word on the sign in terms of rating

The secondary hypotheses were:

1. That the age of the participant will impact their reaction time- with older participants having faster responses due to more exposure over time to signage.
2. That the current employment status of the participant will impact their reaction time– with currently employed individuals having faster response times due to their familiarity with signage.
3. The age of the participant will impact how much their responses are in alignment with OSHA guidelines, with older participants being more in line with the standard due to their higher level of exposure to signage.
4. The current employment status of the participant will impact how much their responses are in alignment with OSHA guidelines, with individuals who are

currently employed being more in line with the standards due to their higher level of exposure to signage.

5. Participants who have spent the majority of their time working in the United States will have responses more in line with OSHA guidelines as they are most familiar with those standards as opposed to those held by the governing bodies of other countries.

Primary Hypotheses

This study found support for two of its three main hypotheses. The first was that signs with the color red would be rated significantly more hazardous than signs with the color blue. This was supported by the results of the repeated measures ANOVA. This means that the current OSHA guidelines are correct in telling companies to use red signs for higher hazard situations than blue signs. This is also corroborated by past research which has indicated that the color red has higher hazard ratings than other colors.

The second hypothesis supported by this study's findings is that signs with the word "danger" are rated as significantly more hazardous than signs with the word "notice". This also supports current OSHA guidelines which state that the word "danger" should be used for higher hazard situations than the word "notice".

The final hypothesis of the study could not be supported by the data. That is the hypothesis that the color of the sign matters more to the hazard rating than the word on the sign. In order for this hypothesis to be supported the red sign with the

word “notice” would have to be rated as significantly more hazardous than the blue sign with the word “danger”. In the end, the opposite was true. Not only was there no significant difference between these two signs, the blue “danger” sign was actually rated as being slightly more hazardous than the red “notice” sign.

While the color of the sign mattering more than the word on it could not be supported, the data did show that there was a significant interaction between the color of the sign and the word used on it. This indicates that while one may not matter more than the other, getting the combination of color and word correct is important. This is a potential explanation for why OSHA guidelines put the same emphasis on all aspects of the sign. If companies believed that they could put any words on a red sign and have it represent the same level of hazard they could inadvertently cause confusion among employees. While researchers originally disagreed with the importance of treating all elements of the sign equally in the guidelines, the current study supports careful consideration of all elements.

Secondary Hypotheses

The first and second secondary hypotheses were about the relationship between reaction time and the participants age and employment status respectively. These hypotheses could not be supported by the data collected. One potential explanation for this is that the online nature of the study affected the response times of individuals. It is possible that compute lag or loading times between screens resulted in inconsistent reaction times across individuals. Another potential

explanation for the lack of significance of reaction times is that there may have been a ceiling effect, or in this case a basement effect. Participants may have already been giving ratings so quickly that it was not possible for them to rate them any quicker. All individuals taking the study may have been responding generally as quickly as possible leaving no room for older participants or those who were currently employed to give ratings any faster.

There are a few possible explanations when it comes specifically to the first hypotheses of older individuals giving faster response times due to increased familiarity with signage. It is possible that older individuals had lower levels of computer literacy and thus found it harder to give ratings, slowing down their recorded reaction times. It is also possible that older individuals have generally slower reaction times meaning even with their increased experience with this type of signage they were unable to give faster ratings than those participants who had less experience but were younger.

The third and fourth secondary hypotheses could also not be supported. Analysis showed that there was no significant correlation between age and hazard ratings or employment status and hazard ratings. This could potentially be due to the fact that all individuals seemed to give ratings that were in line with OSHA guidelines.

Finally, the fifth secondary hypothesis regarding work experience in the United States and rating adherence to OSHA guidelines could not be supported. This was due to no data being collected from individuals who did not spend the majority of

their time employed working in the United States. In order to fully test this hypothesis data would have to be gathered that specifically included individuals who had spent the majority of their time working outside of the United States.

Further Importance

The fact that both color and word had significant effects on the hazard rating of signs, while not surprising based on the previous literature, is important beyond just the realm of OSHA compliant signs. There are far more areas that involve communicating hazards to individuals than just workplace signage. One example is roadside signage. It is important that drivers know what signs require their attention and which signs include less immediate information. Knowing not only what color signs should be but also what words should be used on them is important for maintaining road safety. The fact that this study included responses from individuals of all education levels makes it more representative of the general population, those who are using the roads. Thus, it is not a stretch to extrapolate the results of this study from warning signs to road signs.

Another area where colors are used to indicate hazards is in the communication of public health information. During the COVID-19 pandemic individuals were constantly being shown graphs, bar plots, and other descriptions of scientific data and were expected to fully understand them for their own health. However, depending on how this information was conveyed, there is potential for the importance or gravity of the data to be lost on individuals, especially those with low

scientific or health literacy. It is thus important, especially in a situation like this, to ensure that the most important information is being conveyed in a clear manner, in other words its level of hazard or risk. The data collected in this study regarding the differences in perceived hazard levels of different colors, is directly relevant to this area. Knowing what colors to use on infographics is a matter of public safety and making sure that the information is being conveyed accurately could in some cases be a matter of life or death. This also goes beyond just the COVID-19 pandemic. Information from weather stations regarding the safety of driving conditions or the importance of staying inside during a storm need to be conveyed with the proper amounts of hazard otherwise people will not realize how serious the situation is and may not take precautions when there is actually a dangerous situation.

It is important to note that while this study found support for the current order of two of the colors used in OSHA guidelines it is not definitively saying that these are the best colors to be used in their given scenarios. Indeed, past research has been done looking at the use of fluorescent and neon colors in comparison with the standard safety colors promoted by OSHA (Zielinska, O. A., 2017). That study found that the fluorescent colors tested were rated as high as the standard red colors in both hazard level and importance. This indicates that while the color red is rated as more hazardous than the color blue, there may be other colors that convey the intended hazard levels better.

Biology vs. Habituation

What is the cause of red being seen as more hazardous than blue. While physiology studies support the notion that color can cause a visceral reaction in people causing a sympathetic nervous response, this does not tell us if this is an inherent condition or a learned condition.

Previous research has indicated that infants are able to detect colors on the red-green spectrum from before the age of 2-3 months (Maule, J., 2023). The fact that even infants can recognize the color red and identify it from such a young age points to this color being important evolutionarily. The color red is also used in nature as a color of warning. Berries that are red are often poisonous. The same is true of insects, snakes, and other wildlife. If an animal wants to camouflage itself and pretend to be poisonous it choses red or other bright red like colors. Fire is red and is generally seen as dangerous. Human blood, which when seen indicates danger or injury, is also red. As humans have evolved they have been surrounded by red things that are indicative of danger. Humans from a young age are able to identify that red objects are dangerous. The question still remains however, is this in our biology, or is it something we learn?

Just like an argument can be made that the color red representing negative or hazardous situations is a biological response, there is also an argument to be made that this response to the color red is learned. From a young age we are taught to avoid things that are red. We are surrounded by red warning labels, stop signs and light, and more. Then are the warning signs we see from such a young age what

teaches us to avoid the color red? It would then be possible to make any color a warning color. If all of a sudden all stop, warning, and danger sings turned purple would future generations not then learn to avoid the color purple the same way people now avoid the color red?

While we may not fully understand why red conveys greater hazard, what is important is how people react to it and how it changes their behavior. When it comes to keeping people safe in the workplace and beyond accuracy is of the utmost importance. What matters most is that we know to present information that is especially important when it comes to conveying high hazard situations. This is not to say that further research should not be done on the topic of biology vs. habituation.

Limitations

There are a number of limitations to the study. First, we were only able to gather data from 100 participants. After exclusions, this meant that only the data of 88 participants could be used in final analyses. This was due to funding limitations. However, power analyses indicate that 140-200 participants would be ideal. This could have impacted the negative results we saw for the red sign with “notice” versus the blue sign with “danger”. If the study were to be replicated with a larger budget, it is recommended that more participants be enrolled to confirm the findings of the current study.

Another limitation for this study is that all participants are residents of the United States. This was done to reduce the number of variables present. However, it does mean that the results of this study are limited to the United States.

The program used to run the study, Gorilla, has a number of faults that could potentially impact data collection. First, in order for their rating to be recorded participants had to click directly on the gray line of the scale. Researcher practicing and pilot testing showed this was sometimes difficult to do, resulting in ratings occasionally misrepresentative of the participants intended rating. Pilot testing also indicated that it was difficult for participants to understand that they had to directly click on the gray line as opposed to near it. Due to this feedback, the instruction section was amended to tell participants specifically that they had to click directly on the line for their submission to count. Despite this additional direction, this was still a difficult mechanic to use. It is possible that some participants, especially those who did not thoroughly read the instructions, still had difficulty using the scale properly.

Due to the online nature of the study, participants were required to perform the tasks on their own computers. This could lead to potential errors in data resulting from network or other computer issues. This would be most likely to impact reaction time data.

Another drawback to the online setting of the study is that researchers could not observe participants while they completed the tasks. It is possible that some participants may have not given adequate attention to the tasks or were distracted

while performing the study tasks. While we did exclude participants whose results were more than 2 standard deviations from the mean, it is possible that others may have not been fully engaged in the tasks.

In a similar vein, it is possible that some individuals completing the study did not fully read the instructions provided to them. The total time taken by participants was shorter than what was expected given the amount of time it should have taken to read the consent, instruction, and debriefing materials along with completing the main timed section of the study. If this study was done in a controlled lab setting, researchers could have spent time making sure that individuals clearly understood all instructions.

Directions For Future Study

One of the limitations mentioned for this study is that all participants were residents of the United States. While this was intentional on the part of researchers it would still be interesting to repeat this study with individuals from other countries. OSHA guidelines are only used in the United States, and while organizations like the European Agency for Safety and Health at Work (EU-OSHA) in the EU have similar guidelines, they have some differences. It would be interesting to look at these different agencies' standards to see if one is more supported by research than others. Furthermore, this study should be conducted in languages other than just English. While English is the most popular language spoken in the United States many people living in the United States are non-native English speakers or do not speak English at

all. It is important to know if signs containing two languages, for example English and Spanish, can use direct translations from one language to another or if there are different signal words that work better to convey the intended hazard level. This would be critical to make sure that non-English speakers are receiving the same level of warning. By running a similar study on people outside of the United States and in different languages there is also a chance to understand if the hazard perception of colors is in human biology or habituated. If the same results are found across a number of cultures then it is likely that there is at least some element of biology in the perception of colors. However, if there are significantly different results across cultures it is more likely that habituation is playing the main role in color perception.

Another group that was intentionally excluded from this study was individuals who are diagnosed with a form of colorblindness. While these individuals were excluded from our study there have been studies looking at how colorblind people are able to interpret graphical representations of data that are color coded. It would be interesting to look at warning signs specifically and see if the colors currently being used are accessible to individuals with some form of colorblindness. These are individuals who are also in the workforce and ensuring that safety signs can be understood by everyone is an important endeavor.

Another avenue of potential study is to look at the differences between how men and women view safety signs. Past research has shown that women and men have different associations with colors (Hallock, J., 2003). However, there seems to

be no research looking at color perceptions in contexts. This means that while men and women may associate different things with the color red, they may both understand and rate red as being more hazardous than blue for example. Our study did not collect gender data from participants so this is not an analysis we can conduct but it may be important to look into in a future study. This comes back to one of the main reasons studies like this one are important; to ensure that all individuals in a workplace are able to perform their job safely and with accurate awareness of all potential hazards.

One variable that was controlled in this study but can be manipulated is the font used on the sign. A 2002 study by Bernard and colleagues showed that there were no significant differences in reading efficiency between the eight fonts tested. However, there were significant differences in reading times, not only between the fonts but between the font sizes. They found that Times New Roman and Arial fonts were read faster than others tested. They also found that 12-point fonts were read faster than 10-point fonts. This indicates that the font used on a given sign may have an impact on how fast the signal words can be read. When it comes to hazard signs, it is especially important that signs are both easy to read and can be read quickly. While Bernard indicates that the font used does not impact how easy the sign is to read, it tells us that both font type and font size do impact the speed at which individuals can read the words on the signs. Further research should be done to

determine if font type and size actually play a role in reading speed when it comes to hazard signage.

One final potential direction for future study would be to look at the use of words as opposed to symbols on hazard signage. This study chose words as they are still highly emphasized in the current OSHA guidelines. However, there has been an increasing emphasis alongside text to convey hazard information. New OSHA standards indicate that symbols be added to traditional only word signs. The use of symbols does allow for a more inclusive workplace environment. Individuals who are not fluent in the primary language of the workplace may not be able to read the words on hazard signs but will be able to understand the intended message based on the symbols used. The question remains; are the symbols suggested for use based on current OSHA guidelines able to accurately convey their intended level of hazard? A study similar to this one but with symbols, either in addition to or in the place of words, may be able to shed light on this question.

In conclusion, this study found that both the color of a sign and wording on it impact the sign's perceived level of hazard. Not only do both elements impact the hazard rating of signs but there is also an interaction effect between the two. Careful attention to both elements is necessary in development of signs to clearly convey the appropriate level of hazard for a given situation.

Appendix A: IRB Approval Letter

Bard College Institutional Review Board

Date: 9/1/2023
To: Samantha Feldstein
Cc: Justin Hulbert; Nazir Nazari
From: Ziad M. Abu-Rish, IRB Chair
Re: The Right Sign for the Right Time: A Look at the Effectiveness of Different Types of Signage

DECISION: APPROVED

Dear Samantha Feldstein:

The Bard IRB committee has reviewed your revised proposal. Your application is approved through August 31, 2024. Your case number is 2023SEP1-FEL.

Please notify the IRB if your methodology changes or unexpected events arise. We

wish you the best of luck with your research.

Ziad M. Abu-Rish, Ph.D.
IRB Chair
Associate Professor of Human Rights and Middle Eastern Studies
Bard College (zaburish@bard.edu)

Appendix B: Informed Consent Form

Bard College

Research Participant Information and Consent Form

Study Title: The Right Sign for the Right Time: A Look at the Effectiveness of Different Types of Signage

Principal Investigator: Samantha Feldstein (Email: sf1588@bard.edu)

Faculty Advisor: Justin Hulbert (Email: jhulbert@bard.edu)

Description of the experiment

You are invited to participate in a research study investigating the level of perceived hazard indicated by differently styled safety signs. The purpose of the research is to learn more about the effectiveness of different safety signs.

Why you have been chosen to participate

You have been chosen to participate in this study because you have been identified as an English-speaking adult living in the United States who is willing to voluntarily fill out the survey.

What will my participation involve?

If you decide to participate, you will be asked to look at different safety signs and rank on a scale of “not hazardous at all” to “extremely hazardous” how hazardous you think the situation represented by each one is. Your participation will last approximately 15 minutes.

Are there any risks to me?

While minor, there is a potential risk that participants may feel a slightly elevated sense of anxiety during and/or after the study from viewing images that are supposed to indicate hazardous scenarios.

Are there any benefits to me?

There are no direct benefits to participants; however, participants will be paid $1.82 for their time and will be taking part in research that has the potential to change regulations regarding safety signs.

Will I be compensated for my participation?

For participating in this survey you will be paid $1.82.

How will my confidentiality be protected?

This study is confidential. Neither your name or any other identifiable information will be published. The only people who will have access to the data you provide in your responses will be the researchers. All data will be collected and stored on secure servers used by a program called Gorilla which is being used to run this survey. After it has been downloaded it will be kept on the researcher's encrypted drive which no one else will have access to. The final research thesis (Senior Project) will be available permanently and publicly at the Bard College library and online through the Bard Digital Commons; however, the specific responses participants provide will not be included in that report, only the summary results.

Whom should I contact if I have questions?

If you have any questions or concerns about the survey you may contact any of the following individuals.

The primary researcher, Samantha Feldstein, by email at sf1588@bard.edu

The faculty advisor, Dr. Justin Hulbert, by email at jhulbert@bard.edu

The Bard College IRB chair at irb@bard.edu for questions about your rights as a research participant

Your participation in this survey is completely voluntary. If you decide not to participate or to withdraw from the study, you may do so at any time without penalty even after agreeing in this consent form. You may also choose to skip any question in the survey without penalty.

By checking this box you indicate that you have read this consent form and voluntarily consent to filling out this survey:

- ☐ I have read the above form and voluntarily consent to filling out this survey

By checking this box you confirm that you are at least 18 years of age:

- ☐ I am at least 18 years of age

Appendix C: Demographics Survey

Demographic Questions for Warning Sign Study

What is your age? (Enter N/A if you prefer not to answer this question)
(free response)

What is your current employment status?

- ☐ Employed
- ☐ Unemployed
- ☐ Other (please specify):_____
- ☐ Prefer not to answer

What is your current education level?

- ☐ Some high school education
- ☐ High school diploma, GED, or equivalent
- ☐ Some post high school education
- ☐ Trade school degree
- ☐ Associate degree
- ☐ Bachelor's degree
- ☐ Graduate degree
- ☐ Doctorate or professional degree
- ☐ Prefer not to answer

If you have worked previously, has the majority of your time working been spent in the United States of America?

- ☐ Yes
- ☐ No
- ☐ Not applicable
- ☐ Prefer not to answer

Is English your first language?

- ☐ Yes
- ☐ No

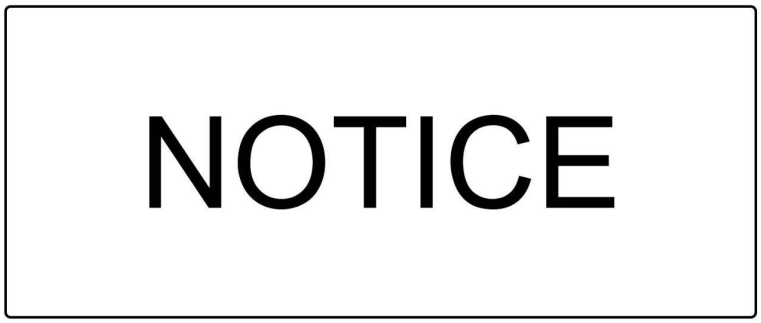
Have you ever been diagnosed with any form of colorblindness?

- ☐ Yes
- ☐ No

Appendix D: Stimuli

Blue Sign

Red Sign



Sign With The Word “Notice”

Sign With The Word “Danger”

Blue Sign With The Word “Notice”

Blue Sign With The Word “Danger”

Red Sign With The Word “Notice”

Red Sign With The Word “Danger”

Appendix E: Debriefing Form

Thank you for participating in our research study. This study, titled *The Right Sign for the Right Time: A Look at the Effectiveness of Different Types of Signage*, is looking at different parts of safety signs to see which makes them seem more hazardous. Specifically, this study is looking at sign color and the word being used on the sign. We think that the color of the sign will matter more than the word on the sign. This is because humans have learned through evolution that the color red indicates danger. People understand this from a very young age, before they are able to understand words. This may make colors more important to making decisions than words are. We believe that because people recognize the color of the sign before they read the word on it, the color will matter more to how hazardous the situation represented by the sign is.

This study is being run by Samantha Feldstein under the supervision of Dr. Justin Hulbert at Bard College located in Annandale-on-Hudson, New York. If you have any questions or concerns about your participation please reach out to Samantha Feldstein by email at sf1588@bard.edu or her supervisor at jhulbert@bard.edu. If you have any questions about your rights as a participant, you may contact Bard's Institutional Review Board (IRB) at irb@bard.edu.

Bibliography

Argo, J. J., & Main, K. J. (2004). Meta-Analyses of the Effectiveness of Warning Labels. Journal of Public Policy &Amp; Marketing, 23(2), 193–208. https://doi.org/10.1509/jppm.23.2.193.51400

Bernard, M.L., Lida, B., Riley, S.R., Hackler, T., & Janzen, K. (2002). A Comparison of Popular Online Fonts: Which Size and Type is Best? https://api.semanticscholar.org/CorpusID:59144734

Birren, Faber., Color Psychology and Color Therapy. New Hyde Park, New York: University Books, Inc., 1961

Blizzard S, Fierro-Rojas A, & Fallah M (2017) Response Inhibition Is Facilitated by a Change to Red Over Green in the Stop Signal Paradigm. Front. Hum. Neurosci. 10:655. doi: 10.3389/fnhum.2016.00655

Boynton, R. M., Fargo, L., Olson, C. X., & Smallman, H. S. (1989). Category effects in color memory. Color Research & Application, 14(5), 229–234. https://doi.org/10.1002/col.5080140505

Braun, C. C., Sansing, L., & Silver, N. C. (1994). The interaction of Signal word and color on warning labels: Differences in perceived hazard. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 38(14), 831–835. https://doi.org/10.1177/154193129403801407

Chapanis, A. (1994). Hazards associated with three signal words and four colours on warning signs. Ergonomics, 37(2), 265–275. https://doi.org/10.1080/00140139408963644

Creative Safety Supply (2016) A Guide to OSHA Safety Signs

Hallock, J. (2003). Colour assignment. Joe Hallock. https://www.joehallock.com/?page_id=1281

Hellier, E., Aldrich, K., Wright, D. B., Daunt, D., & Edworthy, J. (2007). A multi dimensional analysis of warning signal words. Journal of Risk Research, 10(3), 323–338. https://doi.org/10.1080/13669870601066963

Hu, L., Feng, D., Li, Y., Xu, J., & Zheng, J. (2022). The effect of safety signs on the monitoring of conflict and erroneous response. Frontiers in Psychology, 13. https://doi.org/10.3389/fpsyg.2022.830929

Jensen, R. C., & McCammack, A. M. (2003). Severity Message from Hazard Alert Symbol on Caution Signs. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 47(14), 1767–1771. https://doi.org/10.1177/154193120304701411

Jun Bian, Huijian Fu & Jia Jin (2020) Are We Sensitive to Different Types of Safety Signs? Evidence from ERPs, Psychology Research and Behavior Management, 495-505 https://doi.org/10.2147/PRBM.S248947

Maule, J., Skelton, A. E., & Franklin, A. (2023). The development of color perception and cognition. Annual Review of Psychology, 74(1), 87–111. https://doi.org/10.1146/annurev-psych-032720-040512

McCorry, L. K. (2007). Physiology of the autonomic nervous system. American journal of pharmaceutical education. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1959222/

MyPerfectColor.com OSHA Safety Blue. https://www.myperfectcolor.com/paint/440237-behr-osha-1-osha-safety-blue

MyPerfectColor.com OSHA Safety Red. https://www.myperfectcolor.com/paint/440241-behr-osha-5-osha-safety-red

QuestionPro. (n.d.). Likert scale questions, survey and examples. https://www.questionpro.com/article/likert-scale-survey-questions.html

Yuan, J., Song, Z., Hu, Y., Fu, H., Liu, X., & Bian, J. (2021). Electrophysiological correlates of processing warning signs with different background colors: An event-related potentials investigation. Frontiers in Psychology, 12. https://doi.org/10.3389/fpsyg.2021.648871

Zhu, L., Ma, Q., Bai, X., & Hu, L. (2020). Mechanisms behind hazard perception of warning signs: An EEG study. Transportation Research Part F: Traffic Psychology and Behaviour, 69, 362–374. https://doi.org/10.1016/j.trf.2020.02.001

Zielinska, O. A., Mayhorn, C. B., & Wogalter, M. S. (2017). Connoted hazard and perceived importance of fluorescent, neon, and Standard Safety Colors. Applied Ergonomics, 65, 326–334. https://doi.org/10.1016/j.apergo.2017.07.011