Abridgment

Highway Sign Meaning as an Indicator of Perceptual Response

KENNETH A. BREWER, ALICE A. THIEMAN, WILLIAM F. WOODMAN, and LLOYD L. AVANT

ABSTRACT

Semantic differential scaling has been used as a method of evaluating and assessing driver understanding and comprehension of traffic signs in the past. Litigation and other operational pressures on traffic engineering agencies have created an interest in finding a laboratory method for quick and easy estimation of driver performance in processing communication via signs. This paper contains data on research attempts to correlate the meanings assigned to road signs through semantic differential scales. These scales are correlated with drivers' abilities to detect, recognize, and react to road signs. Significant correlations were most often found between meanings attributed to signs in semantic differential scales and the performance of drivers in recognizing signs. No semantic differential scales were found for any sign tested for which a significant correlation existed in detection, in recognition, and in decision-reaction tests. It was concluded that semantic differential scaling has little or no relationship to perceptual response to highway signs by drivers.

During the past decade, tort litigation has made those agencies that are responsible for signing and traffic control of streets and highways very sensitive to the problem of traffic sign effectiveness and driver communication. Although substantial discussion about this heightened sensitivity of state agencies has taken place, the authors' experience has been that local agencies are as much or more affected than state agencies. As engineering organizations have become more interested in examining the fundamental effectiveness of existing and proposed signs, or new applications of existing signs, a concern has arisen as to how testing and evaluation of signs should be carried out.

The typical engineering approach has been to create a prototype and make a pilot plant installation. The design of a sign and test installation on a limited portion of the street and highway system that is suggested by this philosophy has become quite risky as a result of the threat of tort litigation over accidents during testing. Thus, concerns over potential safety hazards inherent in full scale sign testing as well as the potential financial loss during subsequent litigation has increased interest in the laboratory testing of signs.

The Manual on Uniform Traffic Control Devices (1) identifies the generally accepted five basic requirements of an effective traffic control device. Engineering studies can determine whether the need for traffic control devices exists. Traffic enforcement and the judicial process are the primary mechanisms by which road users develop respect for traffic control devices, and likewise, the authors are not concerned with a laboratory method to test respect for traffic control devices. However, it would seem that if laboratory experiments can be conducted that measure differences among signs related to commanding attention, conveying a clear and simple meaning, and giving adequate time for proper response, then much can be learned about the effectiveness of a sign without the necessity of using prototype field testing.

A technique suggested as providing a simple, inexpensive method for evaluating traffic signs is that of the semantic differential (2). The semantic differential technique developed by Osgood et al. assumes that an underlying structure exists for the meanings (semantic context) assigned to elements in a perceived environment (3). Osgood et al. wrote that these underlying or subconscious structures of meanings may be studied by means of a scaling technique similar to a questionnaire. Although Osgood et al. used exploratory factor analysis to find four dimensions of meaning among the set of scales by which the respondents rated a test item, Nunnally has defined analysis validity for each scale (4). Because factor analysis of semantic scales is only a qualitative or arguable assessment of the interaction of scale responses, the authors have chosen for this analysis a portion of their research data set to follow Nunnally and examine each scale separately.

If semantic differential scales of perceived meaning of signs are to be useful in addressing, via laboratory tests, the three basic sign requirements of interest identified previously, then it should be possible to demonstrate some relationship between semantic scales and quantitative tests designed to measure responses to these sign requirements. This paper reports one of a number of analyses performed in the course of a research project funded by the Iowa Department of Transportation Highway Division and demonstrates that caution must be exercised in attempting to extrapolate perceived highway sign meaning into driver response.

EXPERIMENTATION

Three laboratory experiments were designed to test driver responses to a set of 16 signs. The fundamental focus of the research was to examine differences between word legend and symbol legend Stop Ahead warning signs. However, to test the significance and sensitivity of any experimentally determined differences between these signs, it was necessary to incorporate a larger sign set into the design. The total sign set consisted of the 16 signs shown in Figure 1.

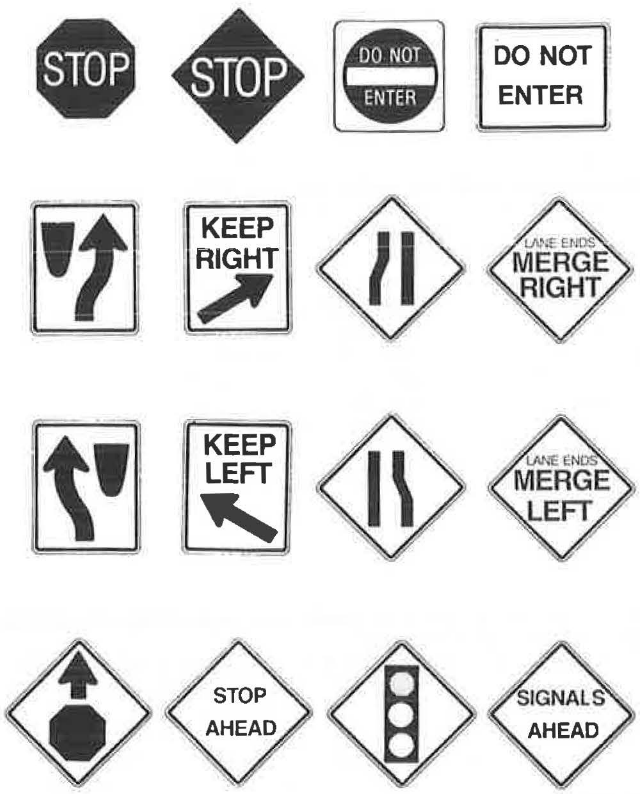

FIGURE 1 Matrix of signs for detection, recognition, and reaction experiments.

Respondents who participated in the experiments described in the following sections were volunteers from undergraduate courses as well as faculty and administrative staff at Iowa State University. Faculty and staff members (16 of 108 persons) ranged from late 30s to early 60s in age. All participants had to possess a valid driver's license. Because the design of the experiments and the testing equipment made potential differences in visual acuity among subjects an irrelevant consideration, no measurement of visual acuity was conducted. Age was not asked of the respondents because a measure of driving experience was obtained (found not to be a significant influence on performance in any of the authors' analyses).

Experiment 1: Detection

A detection experiment was conducted first. Each of 30 persons was presented a series of pre- and post-masked tachistoscopic inputs and asked, after each trial, whether the input was a road sign or a blank flash. Subjects began each trial viewing a mask slide consisting of randomly assembled pieces of various road signs, and the test input for each trial was essentially a brief interruption in the viewing of the mask slide. Each series of trials included presentations of the 16 signs listed previously and 16 blank presentations in a random order. For each subject, the first series of trials began with 110-msec presentations that were clearly visible to the subject. On succeeding series of trials, exposure durations were reduced until the subject performed at no better than chance level in deciding whether each presentation was a blank or a road sign. The criterion of acceptable consistency for a given subject was performance at or below chance level on three consecutive sequences of 32 presentations. Once this criterion was met, three additional series of 32 presentations each were administered to the subject and recorded along with the results of the previous three series.

For each sign, the measure submitted to statistical evaluation was the number of times the sign was correctly detected over the six series at chance-level exposure duration. For the analysis reported here, the probability of correct detection was correlated with semantic differential scale results. The mean chance-level exposure duration for all 30 subjects was 24 msec.

Experiment 2: Recognition

The same sample of 16 signs was used in a second experiment designed to test for differences in recognizability among signs. The experiment was designed to determine whether, after a sign's presence is detected, differences exist in the perceptual operations involved in the recognition process that make the driver aware of the sign. A total of 36 subjects participated in the experiment.

The general procedure was to present the subject a road sign tachistoscopically and then have the subject decide which of the two signs (the just-presented sign and another sign randomly selected from the set) shown outside the tachistoscope in clear vision was the sign that had just been presented. Each trial began with the subject viewing the previously described mask slide; as in the preceding detection experiment, the stimulus presentation was essentially an interruption of the subject's viewing of the mask. The experiment required 240 trials for each subject. This permitted 15 test trials for each sign, that is, 15 trials on which a given sign was presented tachistoscopically and then paired with each of the other signs for the subject's forced choice identification of which sign has been presented tachistoscopically on that trial. The performance measure was the number of errors, of a possible 15, that each subject made. For the analysis reported here, the probability of correct recognition was correlated with the semantic differential scale results.

The 36 subjects were assigned to three groups of 12 subjects each. This made it possible to evaluate the effect of viewing time on sign recognition. Exposure durations were based on the results of Experiment 1 (Detection). Recognition experiment exposure times for Groups 1, 2, and 3 were 32, 41, and 49 msec, respectively. These exposure durations were, respectively, 1, 2, and 3 standard deviations about the mean exposure duration for chance-level presence-absence detection in Experiment 1 (24 msec).

Experiment 3: Decision Reaction Times

This experiment was designed to measure the speed with which subjects could decide on appropriate driver actions for various road signs once the signs were recognized. Forty-eight subjects participated in the experiment. Each subject was provided a response box that housed four response buttons. Respondents were seated in front of a screen onto which road sign slides were projected. At the beginning of the experiment, they were told that road signs would be projected onto the screen and that, for each sign, one of four action decisions would be appropriate. The response decisions would be to stop, to go right, to go left, or to slow down. The subjects were asked to indicate, by pressing the appropriate response button as rapidly as possible, what driver action they would take in response to each of the projected signs.

Proper experimental control required that the assignment of the four response buttons to the four decision actions be varied across subjects. The 48 subjects were accordingly assigned to four groups of 12 subjects each, and assignment of decision actions was counterbalanced across the four groups. As positioned from left to right, the response buttons indicated the following action decisions for the four groups of subjects:

Group	Action Decision
1	Stop, left, right, slow
2	Slow, stop, left, right
3	Right, slow, stop, left
4	Left, right, slow, stop

The performance measure was each subject's mean response reaction time for each sign over 10 randomly ordered presentations of each of the 16 signs. As might be expected, the reversal of decision associated with button position for "go left" and "go right" for Group 3 produced such aberrant values that the results from Group 3 were deleted for this reported analysis.

Semantic Differential Tests

Each subject in the detection, recognition, and decision-reaction experiments was instructed to go to another laboratory to complete a second test. There they were administered the semantic differential scale. Not all subjects did so and the exclusion of subjects in Experiment 3 with reversed left-right response buttons (Group 3) provided 27 subjects from Experiment 1, 35 subjects from Experiment 2, and 23 subjects from Experiment 3 who completed the semantic differential and whose performance could be correlated across the experiments.

To limit the time required in the semantic differential test and minimize subject resistance, the authors decided to utilize only a portion of the complete set of 16 signs. Because the contract focus of the research revolved around the differences between the word and the symbol Stop Ahead signs, both of those were included. Driver behavior using the Stop sign as a "slow" rather than a "stop" driver action was also an issue in the research question, so it was determined that the set of signs to be tested would be the four "slow down" driver action signs and the four "stop" driver action signs.

Twelve 7-point scales were created for each subject to mark in response to each of the eight signs. The extreme ends of each scale were identified with the following pairs of descriptors: good to bad; familiar to unfamiliar; active to passive; predictable to unpredictable; beautiful to ugly; meaningful to meaningless; fast to slow; strong to weak; valuable to worthless; important to unimportant; sharp to dull; simple to complex. These descriptors were selected after consulting original work by Osgood et al. (3) and considering the application previously made by Dewar and Ells (2).

A random number generator was used to select two different sequences of the eight signs to produce slide set A and slide set B to be displayed to respondents. Trial measurements indicated that no more than one person would be expected to be waiting while a subject was participating in the semantic scale test. A random number generator was used to select the order in which the scales were placed on the answer sheet with the same answer sheet being used for all signs viewed and all subjects. Each subject was seated in a room with subdued lighting and shown slides of the previously described eight signs. Each subject was allowed to study each sign as long as he or she wished, but the instructions given at the beginning of the test informed each subject that each scale was to be marked with the first impression about the sign. A randomized order to the scales also included a randomization of the "positive" or the "negative" descriptor as the left end of the scale. The positive end of the scale was given a weight of 7 and the negative end was given a weight of 1 in the data reduction.

RESULTS

It should be pointed out that there were extremely few statistically significant correlations where 192 calculations per table were carried out. In Table 1 there were 18 statistically significant correlations (9.37 percent), whereas in Table 2, only 4 of the correlations were significant (2.08 percent). In Table 3, 10 of 192 possible correlations were significant (5.20 percent), and in Table 4, there were again 10 statistically significant correlations (5.20 percent). Thus, an average 5.46 percent of the possible correlations were statistically significant.

At the same time, the only meaningful patterns of significant correlations were found in relation to the signs bearing the following legends:

- Stop Ahead (symbol)
- Signal Ahead (symbol)
- Stop Ahead (word)
- Do Not Enter (word)

Given that the purpose of the authors' research was to examine formats of the Stop Ahead warning to motorists, the authors found this pattern of findings interesting but puzzling. One possible interpretation of these results might be that all four signs are not seen with great frequency and are likely not thought about when seen. Unlike standard Stop signs that have been so frequently seen that they may have become functionally invisible, these signs may still bear sufficient freshness that they engender responses and meaning attribution. At the same time, the semantic differential scales that generate substantial patterns of correlations (three or more significant correlations) included only active to passive and predictable to unpredictable.

Why these two meaning dimensions would produce these patterns of correlations is also unclear. Given the preceding comments regarding the frequency of sign usage, it may well be that these less frequently seen signs generated both respondent certainty and uncertainty as well as the vitality or robustness of the message contained.

CONCLUSIONS

The basic hypothesis of this research was that tests of perceptual detection, recognition, and action decision latency would correlate with measures of perceived meaning of signs (i.e., that the ability to

TABLE 1 Semantic Differential Scale Correlations with Detection Experiment Results by Sign Shown

	Signal Ahead (Sym)	Signal Ahead (Word)	Stop Ahead (Sym)	Stop Ahead (Word)	Do Not Enter (Sym)	Do Not Enter (Word)	Stop (Oct)	Stop (Diam)
Good - Bad								
Perf Same	--	--	--	+0.39	--	--	--	--
Perf Opp	--	--	--	+0.52	--	--	--	--
Familiar - UNF								
Perf Same	--	--	--	--	--	-0.40	--	--
Perf Opp	--	--	--	--	--	--	--	--
Active - Passive								
Perf Same	+0.40	--	--	+0.37	--	-0.52	--	--
Perf Opp	--	--	--	+0.54	--	-0.52	--	--
Pred - Unpred								
Perf Same	+0.41	--	--	+0.42	--	-0.37	--	--
Perf Opp	+0.50	--	--	--	--	--	--	--
Beautiful - Ugly								
Perf Same	--	--	--	--	--	--	--	--
Perf Opp	--	--	--	--	--	--	--	--
Mean'ful - Mean'less								
Perf Same	--	--	--	--	--	--	--	--
Perf Opp	--	--	--	--	--	--	--	--
Fast - Slow								
Perf Same	--	--	--	--	--	--	--	--
Perf Opp	--	--	--	+0.44	--	--	--	--
Strong - Weak								
Perf Same	--	--	--	--	--	-0.48	--	--
Perf Opp	--	--	--	--	--	--	--	--
Val - Worthless								
Perf Same	--	--	--	--	--	--	--	--
Perf Opp	--	--	--	+0.40	--	--	--	--
Imp - Unimp								
Perf Same	--	--	--	--	--	--	--	--
Perf Opp	--	--	--	--	--	--	--	--
Sharp - Dull								
Perf Same	--	--	--	--	--	--	--	--
Perf Opp	+0.55	--	--	--	--	--	--	--
Simple - Complex								
Perf Same	--	--	--	--	--	--	--	--
Perf Opp	+0.43	+0.46	--	--	--	--	--	--

"--" = Not significant at 0.05 or better level.

Perf Same = detection, recognition or decision-reaction performance on sign with same lexical status to legend as the one scaled.

Perf Opp = detection, recognition or decision-reaction performance on sign with opposite lexical status in legend as the one scaled.

32ms and 49ms = milliseconds exposure duration in tachiostoscopic presentation during recognition experiment, etc.

TABLE 2 Semantic Differential Scale Correlations with 32 msec Recognition Experiment Results by Sign Shown

	Signal Ahead (Sym)	Signal Ahead (Word)	Stop Ahead (Sym)	Stop Ahead (Word)	Do Not Enter (Sym)	Do Not Enter (Word)	Stop (Oct)	Stop (Diam)
Good - Bad								
Perf Same	--	--	--	--	--	--	--	--
Perf Opp	--	--	--	--	--	--	--	--
Familiar - UNF								
Perf Same	--	--	--	--	--	--	-0.57	--
Perf Opp	--	--	--	--	--	--	--	--
Active - Passive								
Perf Same	--	--	--	--	--	--	--	--
Perf Opp	--	--	--	--	--	--	--	--
Pred - Unpred								
Perf Same	--	--	--	-0.67	--	--	--	--
Perf Opp	--	--	--	-0.57	--	--	--	--
Beautiful - Ugly								
Perf Same	--	--	--	--	--	--	--	--
Perf Opp	--	--	--	--	--	--	--	--
Mean'ful - Mean'less								
Perf Same	--	--	--	--	--	--	--	--
Perf Opp	--	--	--	--	--	--	--	--
Fast - Slow								
Perf Same	--	--	--	--	--	--	--	--
Perf Opp	--	--	--	--	--	--	--	--
Strong - Weak								
Perf Same	--	--	--	--	--	--	--	--
Perf Opp	--	--	--	--	--	--	--	--
Val - Worthless								
Perf Same	--	--	--	--	--	--	--	--
Perf Opp	--	--	--	--	--	--	--	--
Imp - Unimp								
Perf Same	+0.69	--	--	--	--	--	--	--
Perf Opp	--	--	--	--	--	--	--	--
Sharp - Dull								
Perf Same	--	--	--	--	--	--	--	--
Perf Opp	--	--	--	--	--	--	--	--
Simple - Complex								
Perf Same	--	--	--	--	--	--	--	--
Perf Opp	--	--	--	--	--	--	--	--

"--" = Not significant at 0.05 or better level.

Perf Same = detection, recognition or decision-reaction performance on sign with same lexical status to legend as the one scaled.

Perf Opp = detection, recognition or decision-reaction performance on sign with opposite lexical status in legend as the one scaled.

32ms and 49ms = milliseconds exposure duration in tachiostoscopic presentation during recognition experiment, etc.

TABLE 3 Semantic Differential Scale Correlations with 49 msec Recognition Experiment Results by Sign Shown

	Signal Ahead (Sym)	Signal Ahead (Word)	Stop Ahead (Sym)	Stop Ahead (Word)	Do Not Enter (Sym)	Do Not Enter (Word)	Stop (Oct)	Stop (Diam)
Good - Bad								
Perf Same	--	--	--	--	--	--	--	--
Perf Opp	--	--	--	--	--	--	--	--
Familiar - UNF								
Perf Same	--	--	--	--	--	--	--	--
Perf Opp	--	--	--	--	--	--	--	--
Active - Passive								
Perf Same	--	--	--	--	--	--	--	--
Perf Opp	--	--	--	--	--	--	--	--
Pred - Unpred								
Perf Same	--	--	--	--	--	--	--	--
Perf Opp	--	--	--	--	--	--	--	--
Beautiful - Ugly								
Perf Same	--	--	+0.61	--	--	--	--	--
Perf Opp	--	--	--	--	+0.70	--	--	--
Mean'ful - Mean'less								
Perf Same	--	--	--	--	--	--	--	--
Perf Opp	--	--	--	--	--	--	--	--
Fast - Slow								
Perf Same	--	--	--	--	--	--	--	--
Perf Opp	--	--	--	--	--	--	-0.73	--
Strong - Weak								
Perf Same	--	--	+0.69	--	--	--	--	--
Perf Opp	--	--	--	--	--	--	--	--
Val - Worthless								
Perf Same	--	--	+0.63	--	--	--	--	--
Perf Opp	--	--	--	--	--	--	--	--
Imp - Unimp								
Perf Same	--	--	--	--	--	--	--	--
Perf Opp	--	--	--	--	+0.68	--	--	--
Sharp - Dull								
Perf Same	--	--	+0.61	--	--	--	--	--
Perf Opp	--	-0.87	--	--	--	--	--	--
Simple - Complex								
Perf Same	--	--	+0.66	--	--	--	--	--
Perf Opp	--	--	+0.60	--	--	--	--	--

"--" = Not significant at 0.05 or better level.

Perf Same = detection, recognition or decision-reaction performance on sign with same lexical status to legend as the one scaled.

Perf Opp = detection, recognition or decision-reaction performance on sign with opposite lexical status in legend as the one scaled.

32ms and 49ms = milliseconds exposure duration in tachiostoscopic presentation during recognition experiment, etc.

see and recognize signs in very short time durations was somehow related to semantic differential measures of stored meaning). Data that the authors will report elsewhere clearly show that sign detection, recognition, and action decision latency are all clearly related to sign meaning. However, for this report, the authors computed a total of 1,152 correlations between laboratory tests of perception and 12 semantic differential meaning scales and so few were found to be significant that it is clear that semantic differential measures of attributed meanings of a sign are not systematically related to laboratory tests of the ability to detect, recognize, and decide on driver actions. The clear suggestion of these findings is that the semantic differential, as an adjunct and verification device for laboratory detection/recognition research is of questionable reliability and validity.

ACKNOWLEDGMENTS

The authors wish to acknowledge the support of Iowa Department of Transportation Highway Division Project HR-256 from which the data in this paper were taken. In addition, The continuing support of both the Engineering Research Institute and the Sciences and Humanities Research Institute at Iowa State University is gratefully acknowledged.

REFERENCES

1. Manual on Uniform Traffic Control Devices. Federal Highway Administration, U.S. Department of Transportation, 1978.
2. R.E. Dewar and J.G. Ells. The Semantic Differential as an Index of Traffic Sign Perception and Comprehension. Human Factors, 1977, Vol. 19, No. 2, pp. 183-189.
3. C.E., Osgood, G.J. Succi, and P.H. Tannebaum. The Measurement of Meaning. University of Illinois Press, Champaign-Urbana, 1967.
4. J. Nunnally. Psychometric Theory. McGraw-Hill, New York, 1978.

The opinions, interpretations, findings and conclusions taken from these data are solely those of the authors.

Publication of this paper sponsored by Committee on User Information Systems.

Drivers' Unconscious Errors in the Processing of Traffic Signs

LESLIE A. WHITAKER

ABSTRACT

Human information processing is divided into two processing modes. One is a conscious, attention-demanding method that is flexible and can be readily controlled. The second is an unconscious, essentially uncontrolled processing that is triggered by well-practiced stimulus-response associations. This paper contains a description of two types of errors to which unconscious processing is prone: illusory combinations of display elements and interference from conflicting irrelevant display elements. Traffic guide signs that may be susceptible to unconscious (automatic) processing errors are also presented as well as research results that are consistent with the hypothesized errors.

Unconscious behaviors during driving are a common experience for some motorists. A person may drive a familiar route and arrive at the destination without being aware of the frequent turns and stops along the way. Drivers are most often aware of this unconscious behavior when they have intended to alter a familiar route by stopping, for example, at the golf course or the gas station. Then, they arrive at the office after having missed the intended stop completely.

Well-learned tasks move from attention-demanding ones to the effortless nature of automatic unconscious processing. This processing is susceptible to a different set of problems than is conscious processing. Driving is a task that may be particularly susceptible to the tricks of automatic processing because it is so well-practiced.

Study of the unconscious was an active area of investigation early in this century (1). It fell subsequently into disfavor and it has been only re-