ENHANCING COMPREHENSION AND RETENTION OF SAFETY-RELATED PICTORIALS

John W. Brelsford
Department of Psychology
Rice University
Houston, TX 77251-1892

Michael S. Wogalter
Department of Psychology
North Carolina State University
Raleigh, NC 27695-7801

James A. Scoggins
Department of Psychology
North Carolina State University
Raleigh, NC 27695-7801

ABSTRACT

Because of their relatively universal information transmission potential, pictorials have been suggested as a common means of safety communications across heterogeneous groups of users and uses. The present study used a training paradigm designed to enhance comprehension and retention of pharmaceutical and industrial-safety pictorials. Manipulated were time of testing (prior to training, immediately following training, and after a one-week delay), content of instruction (supplying the associated verbal label vs. the verbal label plus an extra explanatory statement), and difficulty level ("easy" vs. "difficult" to understand pictorials according to comprehension rates in earlier studies). Using an incomplete factorial mixed-model design experiment, the results showed substantial training effects. There was little change in scores between the test immediately after training and the test after a one-week delay (and the final test scores did not differ between participants who took or did not take the immediate post-training test). Easy pictorials were comprehended (both initially and following training) better than difficult pictorials, although the latter showed the most dramatic increase in understandability after training. Additionally, the instruction content manipulation (adding the explanatory statement to the verbal label)—which had been expected to influence the degree of encoding—had no effect on retention. The substantial gains in understanding the more difficult pictorials suggest that brief training, as little as giving the pictorial's verbal meaning once, can have a large impact in facilitating comprehension for pictorials that would otherwise not be understood by many people.

INTRODUCTION

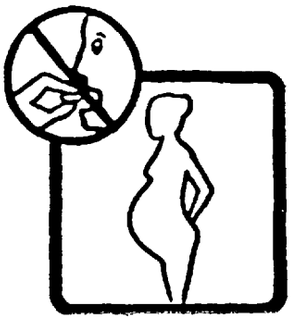
Warnings, risk communications, and operating instructions increasingly use pictorials to convey safety-related information (Laux, Mayer, and Thompson, 1989; Young and Wogalter, 1990). Most warnings guidelines and standards (e.g., ANSI, 1991; FMC, 1985; Westinghouse, 1981) recommend that warnings include descriptive pictorials. Pictorials are increasingly used in consumer products ranging from pharmaceuticals to high-tech vehicles, photographic, and home entertainment equipment. They also appear in professional contexts ranging from health to industrial machinery. Concern for user diversity as related to language and comprehension abilities is also increasing. In part, this interest is derived from multi-national marketing approaches, changes in life styles and usage patterns, and increasingly complex features of modern life.

Pictorials are potentially useful ways to convey safety information, but only if they are accurately comprehended and understanding is maintained over time. Several studies (e.g., Collins, Learner, and Pierman, 1982; Laux, Mayer, and Thompson, 1989; Wolff and Wogalter, 1993) have shown that some pictorials in use today are not well understood. The problem is that while pictorials can be developed for some concepts that people will readily understand (mainly for concrete ideas), there are other concepts (mainly of abstract ideas, e.g., passage of time, radiation, biohazards, etc.) for which many people will not understand when pictorially depicted. Indeed, some concepts may never be adequately communicated via simple pictorials. Wolff and Wogalter (1993) and Wogalter, Wolff, Magurno, and Kohake (1994) describe a set of studies in which a set of pharmaceutical pictorials are iteratively evaluated across a series of test and redesign cycles. After several iterations of improvements to some poorly understood pictorials, it became apparent that some of the concepts (e.g., Take Until Gone, Take 2 hours after/before meals) may never have an associated pictorial that reaches the ANSI or ISO criteria of 85% correct comprehension (on recall-type tests). The dilemma is what to do when certain pictorials can not be redesigned successfully such that the concept does not have a pictorial representation that is comprehended at high levels. One possibility is to train people on the meanings of various pictorials in order to enhance people's understanding of them (Cairney and Sless, 1982; Green and Pew, 1978). Furthermore, if the training is to be considered beneficial then pictorial comprehension should be retained over time. Also, as a practical concern, one would prefer the training to be simple and brief.

This study used a training paradigm to determine whether both comprehension and retention of heterogeneous pictorials can be enhanced. Content of Instruction and the Difficulty

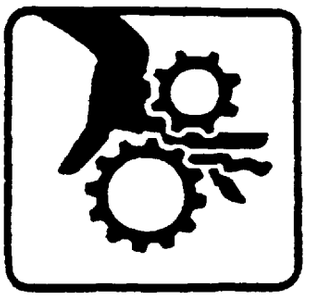
Figure 1
Example Easy and Difficult Pharmaceutical and Industrial-Safety Pictorials

Pharmaceutical—Easy



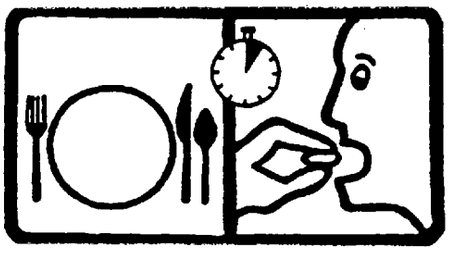
Verbal Label: Do Not Take if Pregnant
Explanation: Medications are passed to the baby from the mother and may affect or hurt the baby. Effects can include substantial birth defects, stillbirths, or spontaneous abortions.

Industrial-Safety—Easy



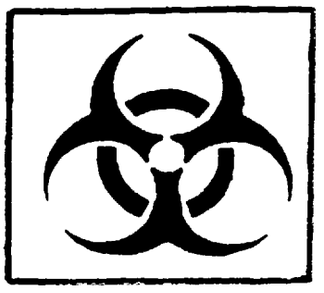
Verbal Label: Danger! Entanglement Hazard: Keep Hands Clear
Explanation: Wearing loose clothing and/or jewelry can cause entanglement to occur, resulting in loss of limbs or death. Long hair should be tied back to avoid entanglement.

Pharmaceutical—Difficult



Verbal Label: Take 1 Hour after Meals
Explanation: This medication may upset an empty stomach irritating the stomach lining and interfering with digestion. If taken immediately before or with meals, food my bind with the drug and alter its chemical composition.

Industrial-Safety—Difficult



Verbal Label: Biohazard
Explanation: This hazard may cause birth defects or some other type of genetic abnormalities, illness, or disease.

Level of the pictorials were manipulated, as was the presence of an immediate post-training test. The content of the instruction itself consisted of either: (a) giving the associated verbal label for each pictorial, or (b) giving the verbal label plus an explanatory statement that included additional detail on the reason for the message conveyed by the pictorial. This enhanced description, should, according to the verbal learning literature, enhance people's memory codes, leading to increased retention performance (D'Agostino, O'Neill, and Paivio, 1977; Paivio, 1975). Difficulty Level of the pictorials (easy vs. difficult to understand) was included to determine whether the training procedures benefit only the more difficult pictorials or do they also increase comprehension of those known to be understood at acceptable levels. Both pharmaceutical and industrial-safety pictorials were included for the purpose of maximizing generalizability to other pictorial categories. Finally, the manipulation of the presence vs. absence of an immediate post-instruction test was threefold: (a) to determine the potential influence of this test on performance on the later test, and (b) to determine whether performance on the immediate post-test predicts comprehension of the pictorials at the later point in time. The dependent measure was comprehension test scores at different points in time (before training, immediately after training, and after 7-10 days following training).

METHOD

Participants

Sixty North Carolina State University undergraduates participated for research credit in introductory psychology courses. Four between-Participant groups were formed each containing 15 randomly-assigned participants.

Table 1
Mean Proportion Correct as a Function of Pictorial Training Conditions

	Verbal Label Only				Verbal Label Plus Explanation			
	Pharmaceutical		*Industrial*		*Pharmaceutical*		*Industrial*	
Time of Test	*Easy*	*Difficult*	*Easy*	*Difficult*	*Easy*	*Difficult*	*Easy*	*Difficult*
Initial Test	.92	.50	.66	.33	.86	.43	.69	.32
Test Immediately Following Training	.98	.84	.94	.82	.93	.85	.84	.73
Delayed Test 7-10 Days Later	.99	.84	.93	.73	.95	.79	.87	.71

Materials and Design

Forty safety-related pictorials were used. Twenty were selected from a set used to communicate pharmaceutical safety instructions and 20 represented various industrial-safety hazards. Within each set, 10 pictorials were classified as easy to comprehend and 10 were classified as difficult to understand. Selection and classification was based on comprehension data from previous studies. The pharmaceutical pictorials included were drawn from comprehension test results of Wogalter et al. (1994) and Wolff and Wogalter (1993). The industrial safety pictorials are standard pictorials found in industry (FMC, 1985; Westinghouse, 1981). Selection and categorization was based on the research of Collins, Lerner, and Pierman (1982), Frascara and Yau (1986), and Mayer (1992). The pictorials were printed in black and shown on white backgrounds.

The experiment involved the crossing of two between-subjects independent variables: (a) Content of Instruction (verbal label only vs. verbal label with explanatory statement), and (b) Test Immediately Following Training (presence vs. absence). These factors produced four groups:

(1) verbal label only with no immediate post-training test,
(2) verbal label only with an immediate post-training test,
(3) verbal label plus explanation with no immediate post-training test, and
(4) verbal label plus explanation with an immediate post-training test.

There were also three repeated-measures independent variables: (a) Time of Testing (prior to training, immediately following training—which was assessed for half of the participants, and final test after a 7-10 day delay), (b) Difficulty Level (easy and difficult to understand pictorials), and Pictorial Type (pharmaceutical and industrial-safety).

Procedure

Testing and training sessions consisted of groups no larger than six and began when participants were told their initial task was to identify (and explain the significance of) concepts represented by 40 pictorials. Participants received pictorials assembled in a booklet each in a differing unique random order and were provided adequate time to write responses to each pictorial.

Following the initial comprehension test, all participants received one of two types of pictorial training. Half of the participants viewed the set of 40 pictorials printed on transparencies (shown on an overhead projector) along with its defining verbal label printed below each pictorial. The other half viewed the pictorials in which each included the defining verbal labels plus a short statement explaining the nature of the concept or hazard. Example pictorials and the corresponding verbal labels and explanatory statements are shown in Figure 1. As each transparency was presented, the experimenter read aloud the verbal statements (either the label alone or the label and the explanation). The experimenter paused for approximately 3 sec after saying each statement before presenting the next pictorial.

Following the brief training session, participants engaged in a set of filler tasks. They were told that a secondary purpose of the study was to determine whether performance on the pictorial test was related to scores on a perceptual-speed task. In this task, participants circled target letters on a

page containing a large array of randomly-ordered alphabetic characters. The purpose was to: (a) insert a non-rehearsal distractor activity between the training phase and the test immediately followed training, and (b) to prevent stimuli from being retained in short-term memory. Even though only half the participants took the immediate post-training test, all participants performed the intervening perceptual-speed task to keep all aspects of the procedure identical except for the manipulated variables. Both speed and accuracy were emphasized in the instructions for the perceptual-speed activity. After working for three min. on the first sheet, participants were told to stop. A second sheet with a different set of target letters was given to participants and the procedure was repeated. After the search tasks, participants completed a demographic questionnaire. Together, the filler tasks (the two perceptual-speed tasks and the demographic questionnaires) took about 10 min.

Depending upon participants' randomly-determined group assignments, one half were released after the filler activities with the understanding that they would return seven days later to receive full credit for their research participation. The other half of the participants received another identification/comprehension test on the same set of pictorials. The immediate post-training test was identical to the initial (pre-training) test except that each participant received the pictorials in a different random order. After completing the immediate post-training test, they were told to return to the laboratory in seven days to receive full credit for their research participation.

Seven to 10 days later, all participants took part in the final comprehension test. The testing followed the identical procedures as before (with new random orders). Participants were then debriefed, given research credit, and thanked.

RESULTS

Responses by participants were scored with a "1" if correct and "0" if incorrect. Two judges scored all of the response sheets without knowing the conditions from which they were taken (i.e., blind). Reliability of the scoring (the number of scores matching between the two judges divided by the total number X 100) was 90.3%. The scores in this report are based on one judge's scores.

Means for conditions in the experiment can be seen in Table 1. The experiment was an incomplete factorial design (half the participants were not tested immediately following training); therefore, it was necessary to analyze the entire set of data using two analyses of variance (ANOVAs).

The first analysis used all 60 participants but did not include data from the immediate post-instruction test. A five-way mixed-model ANOVA included the factors: (a) Content of Instruction: verbal label only vs. verbal label plus explanation, (b) Immediate Test: having or not having an immediate post-training test, (c) Time of Test: initial vs. final test, (d) Pictorial Type: pharmaceutical vs. industrial-safety pictorials, and (e) Difficulty Level: easy vs. difficult to understand pictorials. The ANOVA showed a significant main effect of Time of Test, $F(1, 56) = 408.84$, $p < .0001$, with comprehension scores higher in the final ($M = .85$) than in the initial ($M = .59$) test; a main effect of Pictorial Type, $F(1, 56) = 107.52$, $p < .0001$, with comprehension better on the pharmaceutical ($M = .79$) than the industrial-safety pictorials ($M = .65$), and a main effect of Difficulty Level, $F(1, 56) = 628.99$, $p < .0001$, with comprehension better for the easy ($M = .86$) than the difficult pictorials ($M = .58$).

The ANOVA also showed a significant two-factor interaction of Time of Test and Difficulty Level, $F(1, 56) = 106.26$, $p < .01$. The pattern of this interaction can be seen within the set of means shown in Table 1. The Time of Test x Pictorial Difficulty means show a substantial difference between the easy and difficult pictorials at the initial test, and while both showed increased comprehension following training at the final test, the difficult pictorials showed a more dramatic increase. That is, the difference between easy and difficult pictorials decreased in magnitude on the final test. The ANOVA also showed a significant three-factor interaction of Time of Test, Pictorial Type, and Difficulty Level, $F(1, 56) = 7.90$, $p < .01$. The pattern of this interaction is similar to the two-way interaction described above except that it specifically indicates that the less-well-known pictorials according to the first test (which tended to be the more difficult industrial-safety pictorials) showed the most impressive increase at the time of the final test whereas better known pictorials (of either pictorial type) at the first test did not show as large of an increase.

The second ANOVA included scores from the intermediate post-instruction test, and, as a consequence, excluded participants who did not take this test. The four-way mixed-model ANOVA included the factors: (a) Content of Instruction: verbal label only vs. verbal label plus explanation, (b) Time of Test (with 3 levels as opposed to the earlier ANOVA which had two levels for this factor): initial test vs.test immediately following training vs. final test 7-10 days later, (c) Pictorial Type: pharmaceutical vs. industrial-safety pictorials, and (d) Difficulty Level: easy vs. difficult to understand pictorials. The pattern of results is similar to the description previously presented. However, this ANOVA confirmed a pattern apparent in Table 1, that of a significant Time of Test x Difficulty Level interaction, $F(2, 56) = 62.90$, $p < .0001$. Simple effects analysis and pairwise comparisons using the Newman-Keuls Multiple Range tests showed a significant increase in scores from the initial test to the test immediately following training and that the higher rate of comprehension was maintained (and did not significantly differ from the immediate post-training test) one week later at the final test.

DISCUSSION

Results support the expectation that comprehension and retention of pictorial meanings can be influenced in several

ways. Significant effects of training are seen by comparing scores from the initial test (prior to training) and scores in the two subsequent post-training tests. Comprehension scores immediately following training show an increase compared to the initial test prior to training. Moreover, the increased comprehension was maintained one week later at the final test (irrespective of having taken a test immediately after training). This result is somewhat surprising, given what is known about memory of verbal materials where there tends to be a substantial decline after a several-day delay. However, the failure to see a drop in performance is not entirely unexpected because research using other kinds of complex visual stimuli (e.g., faces) also does not show a rapid forgetting curve (e.g., see Laughery and Wogalter, 1989 for a review). Nevertheless, it is important to determine the rate of forgetting. For this purpose, subsequent research should probe for longer-term retention to evaluate declines in pictorial comprehension by testing after a longer delay (e.g., after a year). It would be surprising if some losses in pictorial comprehension did not take place over lengthy retention periods. However, if the decline is very small then in both theoretical and empirical terms, it is suggested that highly meaningful visual pictorial stimuli may act on memory systems differently than verbal materials. If extremely long-term retention were found, then clearly the training is successful and the findings practically useful.

A few other points should be mentioned. First, the pre-training test means (as seen in Table 1) provide an excellent manipulation check for evaluating the selection process used to categorize "easy" and "difficult" pictorials; initial selection of pictorials was based on the comprehension rates in earlier studies, and the present results showed it to be a valid method of categorization.

Second, high levels of performance for easy pictorials on the pre-training test indicates a ceiling effect. When performance is initially at high levels, only small effects of training are possible. Only for difficult pictorials is the initial level of performance low enough to show substantial training effects. Nevertheless, positive training effects were found for even highly comprehensible (easy) pictorials.

Third, content of instruction did not produce the expected effect of using an elaborated verbal code; there was no difference between the presence and absence of the extra explanatory statement versus training provided by the verbal label alone. Such "expanded" feedback was intended to lead to increased retention through enhanced memory codes. Possible reasons for the null finding might be (a) an inability to adequately encode the explanatory statement itself, (b) the failure of such information to provide additional memory codes beyond those existing otherwise, or (c) the retention measure was not sensitive enough to assess the effect of the elaborative encoding that did take place. At this point, definitive conclusions on the effects of supplementary information can not be given.

Fourth, and perhaps the most important point of this research, is that very brief instruction with only the associated verbal label substantially increased comprehension of difficult pictorials. This finding is important because pictorials are being created and used in a variety of contexts in which their meaning is not always readily apparent. The present results suggest that while highly understandable pictorials are the best pictorials to use when they are available (or can be developed), very brief training of poorly-understood pictorials appears to raise comprehensibility dramatically. Furthermore, such memory appears to be maintained over long periods of time. Training of unfamiliar pictorials might be accomplished, for example, through television spots (e.g., public service messages). The television medium affords an excellent opportunity for pairing pictorials with verbal meanings. Moreover, the verbal description could conveyed in a variety of languages in multi-cultural communities.

ACKNOWLEDGMENTS

The authors would like to thank Hal Cockerham, Cash Johnson, Tammy Taylor, Chris Hill, and Amy Magurno for their help in the early stages of this research. The authors also wish to thank the U.S. Pharmacopoeia for permission to use their pictorials in research.

REFERENCES

Cairney, P. T., and Sless, D. (1982). Communication of effectiveness of symbolic safety signs with different user groups. *Applied Ergonomics, 13*, 91-96.

Collins, B. L., Lerner, N. D., and Pierman, B. C. (1982). *Symbols for industrial Safety*. (Tech. Report NBSIR 82-2485). Washington, DC: U.S. Department of Commerce.

D'Agostino, P. R., O'Neill, B. J., and Paivio, A. (1977). Memory for pictures and words as a function of level processing: Depth or dual coding? *Memory and Cognition, 5*, 252-256.

Frascara, J., and Yau, T. (1986). *Evaluation of Safety Symbols, Appropriateness Ranking Tests and Comprehension Recognition Tests.* Department of Art and Design, University of Alberta, Edmonton, Canada (Report submitted to Research Contribution Advisory Committee, Standards Council of Canada).

Green, P., and Pew, R. W. (1978). Evaluating pictographic symbols: An to motive application. *Human Factors, 20*, 103-114.

Laughery, K. R., & Wogalter, M. S. (1989). Forensic applications of face research. In A. W. Young, & H. D. Ellis (Eds.), *Handbook of Research on Face Processing* (pp. 519-555). Amsterdam: Elsevier.

Laux, L., Mayer, D. L., and Thompson, N. B. (1989). Usefulness of symbols and pictorials to communicate hazard information. In *Proceedings of Interface '89* (pp. 79-83). Santa Monica, CA: Human Factors Society.

Mayer, D. L. (1992). *Identifiability and effectiveness of graphic symbols used in warning messages.* Unpublished manuscript. Houston, TX: Rice University.

Paivio, A. (1975). Perceptual comparisons through the mind's eye. *Memory and Cognition, 3*, 636-647.

Westinghouse (1981). *Westinghouse product safety label handbook.* Trafford, PA: Westinghouse Printing Division.

Wogalter, M. S., Wolff, J .S., Magurno, A. B., and Kohake, J. R. (1994). Evaluation and redesign of pharmaceutical pictorials. In *Proceedings of the 12th Triennial Congress of the International Ergonomics Association*, in press.

Wolff, J. S., and Wogalter, M. S. (1993). Test and Development of Pharmaceutical Pictorials. In *Proceedings of Interface '93* (pp. 187-192). Santa Monica, CA: Human Factors Society.

Young, S. L., and Wogalter, M. S. (1990). Comprehension and memory of instruction manual warnings: Conspicuous print and pictorial icons. *Human Factors, 32*, 637-649.