Routledge
Taylor & Francis Group

Journal of Organizational Behavior Management

ISSN: 0160-8061 (Print) 1540-8604 (Online) Journal homepage: http://www.tandfonline.com/loi/worg20

Applied Behavior Analysis and Occupational Safety

Thomas E. Boyce & E. Scott Geller

To cite this article: Thomas E. Boyce & E. Scott Geller (2001) Applied Behavior Analysis and Occupational Safety, Journal of Organizational Behavior Management, 21:1, 31-60, DOI: 10.1300/J075v21n01_03

To link to this article: http://dx.doi.org/10.1300/J075v21n01_03

SPECIAL ARTICLE ON SAFETY RESPONSE MAINTENANCE AND MANAGEMENT

Applied Behavior Analysis and Occupational Safety: The Challenge of Response Maintenance

Thomas E. Boyce
E. Scott Geller

ABSTRACT. Methods of generalization (Stokes & Baer, 1977) are applied to investigate contingencies that produce long-term response maintenance in large-scale behavioral interventions for industrial safety. It is proposed that characteristics representing four methods originally outlined by Stokes and Baer (1977) have an additive effect on the duration and amount of response maintenance obtained. To evaluate the validity of this conceptualization, a sample of occupational safety research published in refereed journals between 1974 and 1996 is de-

Thomas E. Boyce and E. Scott Geller are affiliated with Center for Applied Behavior Systems, Department of Psychology, Virginia Polytechnic Institute and State University. Thomas E. Boyce is now at the University of Nevada, Reno.

Preparation of this report was supported in part by Grant Number 1 R01 OH03374-02 from the National Institute for Occupational Safety and Health (NIOSH), Centers for Disease Control and Prevention.

Correspondence concerning this article should be addressed to: Thomas E. Boyce, Department of Psychology/296, University of Nevada-Reno, Reno, NV 89557 or E. Scott Geller, Center for Applied Behavior Systems, 5100 Derring Hall, Department of Psychology, Virginia Tech, Blacksburg, VA 24061-0436 (E-mail: teboyce@unr.edu).

Journal of Organizational Behavior Management, Vol. 21(1) 2001

scribed and analyzed. Regardless of intervention technique, noteworthy consistencies among all studies successful at producing maintenance were the use of on-site workers to administer the program and general feedback in the context of multiple target behaviors; or a single-target behavior that was easy and convenient to emit. It is recommended that applied behavior analysts set out to program response maintenance at the expense of publishing fewer short-term demonstrations of functional control.

KEYWORDS. Response generalization methods, long-term response maintenance, large-scale industrial safety interventions, behavioral safety

The Occupational Safety and Health Act of 1970 was a catalyst for psychological research related to proactive injury reduction. Initial research consisted of examining behavioral and psychological maladies related to exposure to toxic substances, workers' abilities to detect occupational hazards, individual differences related to injury proneness, job stress, and intervention techniques for improving occupational safety and health behaviors (Cohen & Margolis, 1973). This latter research domain is most closely related to the theme of this paper.

For the purpose of this paper, safety is defined as injury control. Although many researchers use the terms injury and accident interchangeably, not all accidents result in injury nor are all accidents chance occurrences as implied by the word "accident" (Geller, 2001). Regardless, injuries are the permanent products (Cooper, Heron, & Heward, 1987) or outcomes of behaviors in the context of a dynamic physical and social environment. Reducing the at-risk behaviors which lead to injury and promoting safe alternatives necessarily results in a decrease in injuries (Reber & Wallin, 1983). The challenge is to produce maintenance of safety-related behaviors through a change in natural environmental contingencies.

BEHAVIORAL MAINTENANCE

Definition

Maintenance of behavior change has been defined in many ways, using many different terms. Early definitions from applied behavior

analysis referred to response maintenance as generalization of behavior change across time (Baer, 1982; Stokes & Baer, 1977). Others have referred to maintenance as resistance to extinction (Kazdin, 1994; Stokes & Baer, 1977). For the purpose of this paper, maintenance is defined as the occurrence of one or more target behaviors above baseline levels for a minimum of two weeks after the withdrawal of contrived intervention contingencies. Levels "above baseline" must be detectable through visual inspection of time-series data or reported in terms of statistical significance. Thus, occurrence of increased levels of the target behavior at times when no formal treatment is in effect (Kazdin & Polster, 1973) will be considered maintenance. This definition should distinguish maintenance as the topic of this conceptual analysis from "institutionalization" which can be seen as the continuation of program-related contingencies by on-site workers after the outside intervention agents or researchers have left the setting.

Response maintenance is not necessarily a natural by-product of behavior change techniques, but needs to be programmed to occur as a result of natural environmental contingencies operating in various settings (Baer, Wolf, & Risley, 1968; Bandura, 1986; Kazdin, 1977; Stokes & Baer, 1977). It has been argued that even natural social reinforcement must be systematically programmed (Kazdin, 1973). We presume that techniques of generalization (Baer, 1982) or transfer of training (Kazdin, 1994) are the same as those needed to produce maintenance of behavior change by contingencies naturally occurring in the workplace. From the perspective of applied behavior analysis, the majority of these technologies rely on variations in presentation of behavioral antecedents and consequences.

Programming Maintenance of Behavior Change with Behavior Analysis

The current thesis is that the contingencies by which behavioral maintenance is achieved are different than those required for acquisition of behavior. Efficient response acquisition requires successful discrimination (Skinner, 1953; Stokes & Baer, 1977). In contrast to the mechanisms of response acquisition, we argue that successful behavioral maintenance relies on the failure of people to accurately discriminate contingencies of reinforcement.

Stokes and Baer (1977) wrote extensively about the necessity of programming generalization of responding. Baer (1982) reconceptual-

ized the generalization techniques introduced by Stokes and Baer (1977) and distinguished between those that were generalization methods from those that were not scientifically-based methods of programming generalization (e.g., "train and hope"). These antecedent and consequence control strategies can be applied to maintenance of behavior change since maintenance can be considered generalization to occasions after the intervention has been terminated. The four techniques, identified by Baer (1982) as true "methods of generalization," are discussed below with regard to response maintenance. Each is illustrated with an example from industrial safety.

The method of training loosely. This technique involves programming several stimuli to evoke discriminated operants by reinforcing all desired behavior change (Baer, 1982). Specifically, all possible stimulus dimensions within the experimental setting are manipulated unpredictably and thus, set the occasion for contingent consequences produced by a variety of responses. The objective is to prevent any single stimulus from gaining unquestionable discriminative properties (Baer, 1982). Studies that target multiple behaviors and used more than one antecedent to occasion behavior change qualify as "training loosely."

Sulzer-Azaroff and de Sanatamaria (1980) delivered feedback regarding improvements in 18 behaviors related to the prevention of environmental hazards to 128 employees of a custom-product manufacturing plant. Feedback was provided twice a week to department supervisors who posted the number and location of hazards for their departments. Feedback was also accompanied by more frequent safety meetings and a neater plant appearance. The presence of the supervisor, feedback signs, safety meetings, and plant appearance can be considered antecedent stimuli in this context. Housekeeping improvements were maintained for up to four months after the termination of the formal program.

The method of indiscriminable contingencies. Although training loosely precludes specific stimulus-response-consequence discriminations, the primary emphasis is on manipulating antecedent stimuli and reinforcing multiple target behaviors. The method of indiscriminable contingencies places emphasis on the consequences of behavior. That is, consequences are presented in a manner to prevent discriminations as to the setting, time, or specific response on which they are contingent (Baer, 1982). Variable schedules of reinforcement and delayed

reinforcement procedures provide excellent examples of this technique.

The effectiveness of delayed or variable probability reinforcement was demonstrated by Rudd and Geller (1985) in their large-scale intervention to increase the use of safety belts on a university campus. This incentive/reward program used police to record the license plate numbers of drivers who were buckled-up as they entered or left campus parking lots. Ten weekly raffle winners were drawn from the recorded licenses for prizes ranging from $20 to $500. Conceptually, this program provided a variable schedule of reinforcement dependent on the number of times a particular person was seen buckled and the proportion of other safety-belt users noticed during the week preceding a raffle. Since prizes were not delivered immediately to safety-belt users, the rewards were delayed. An increase in belt use of 46% above baseline was maintained for five months after initial intervention period ended.

The method of common stimuli. This method involves a functional analysis of the stimuli most often occurring in the natural environment of the target individual(s). Discrimination training is provided in the presence of these stimuli. Thus, discriminations are made in the presence of a variety of environmental antecedents. A practical example of training common stimuli in industry would be to reinforce desired behaviors in the presence of employees with whom people work on a daily basis. In this situation, the physical presence of certain peers could exert stimulus control (Baer, 1982).

Peer influence was a primary component in a study reported by Ludwig and Geller (1991). In their promise-card commitment intervention to increase the use of safety belts among pizza deliverers, store cooks and dispatchers reminded drivers to buckle-up as they left for their deliveries. Thus, after the formal termination of the intervention, the mere presence of the store employees could have helped to maintain the desired behavior. That is, it could be argued that the cooks, dispatchers, and store managers exerted some stimulus control over safety-belt use, which was presumably negatively reinforced through avoidance of oral or written reprimands. In fact, increases up to 200% above baseline levels of safety-belt use were maintained up to 11 weeks after termination of the education, goal-setting, and promise-card commitment intervention.

The method of mediated generalization. This is a ''self-control'' technique. Self-control is defined by Rachlin (1991) as a pair of dis-

count functions that cross as time goes by. Specifically, as long as more than one reinforcer is available and all are delayed, the largest is presumed to control behavior. With regard to maintenance of safety-related behaviors, individuals are said to reinforce their own target behaviors on a schedule of reinforcement that places control on a larger delayed reinforcer (Baer, 1982). It seems then, in mediated generalization, verbal behaviors exert stimulus control over the target response and more probable behaviors (Premack, 1962) often function as reinforcers.

A large-scale application of this method is exemplified by the ''Promise Card Technique'' described by Geller and Lehman (1991). This simple intervention strategy asks participants to make a written commitment to perform a specific response. Such commitments, when voluntary, create a saying-then-doing sequence for which a consistency between saying and doing has likely been reinforced in the past (Biglan, 1987; Rogers-Warren & Baer, 1976). Others have argued that such behavior is ''rule-governed'' as when verbal behavior describes ''indirect-acting'' contingencies of reinforcement. In this conceptualization it is proposed that the rules lead to a ''direct-acting'' escape or avoidance contingency that maintains behavior (Malott, 1992). It is our contention that this is simply different language for Stokes and Baer's (1977) method of mediated generalization and is thus consistent with the present argument.

Geller, Kalsher, Rudd, and Lehman (1989) combined a promise-card commitment technique with incentive/rewards to motivate safety-belt use among faculty, staff, and students on a university campus. Specifically, participants who signed buckle-up promise-cards (antecedent) were given the opportunity to win prizes (consequences) donated by local merchants. The promise-card was designed so that one portion could be used as a raffle entry and the other portion could be hung from a vehicle's rear-view mirror as a reminder of the commitment. Faculty promise signers significantly increased their use of safety belts from a preintervention mean of 59.1% to a follow-up mean of 75.8%. Student promise signers also increased their belt use to 68.7% from an initial baseline of 50.4%.

Research Characteristics Needed to Demonstrate Maintenance

In behavior change or acquisition research, it seems investigators are most concerned with demonstrating a functional relationship between the intervention contingencies and behavior change. In contrast,

demonstrating maintenance requires concluding that behavior change persists after the intervention is withdrawn. A demonstration of maintenance, however, first requires a demonstration of intervention effects (Rusch & Kazdin, 1981) or acquisition of the target response. In time series designs, behavioral acquisition is minimally demonstrated by statistically significant differences in the frequency, rate, or strength of a target response from baseline to intervention. A visual inspection of the data should reveal changes in the level, trend (upward or downward), or degree of change (slope) from baseline to intervention (Jones, Vaught, & Weinrott, 1977). Rusch and Kazdin (1981) suggested that the lack of studies reporting maintenance may be due to a lack of research designs that allow the systematic evaluation of factors contributing to response maintenance.

Consider, for example, the A-B-A reversal design (Baer, Wolf, & Risley, 1968). In reversal designs the first A assesses the rate or frequency of a target response prior to the introduction of any contrived experimental contingency. These data provide a baseline for comparison with the effects of an intervention.

Thus, when using reversal designs, a return to baseline levels of the target behavior during withdrawal is requisite for demonstrating unequivocal functional control of the intervention contingencies. That is, control over behavior is demonstrated if altering the experimental operations in turn reliably alters the behavior (Kazdin, 1973). This requirement by nature precludes demonstrations of maintenance because failure of the target behavior to reverse (maintenance in the absence of contrived contingencies) prevents conclusions of functional control of the treatment, and thus, it could be suspected that other variables might be operating on the target behavior.

The failure to conclude that the results of an experiment are due to the experimental manipulation is a failure to demonstrate internal validity (Campbell & Stanley, 1963). Accordingly, Kazdin and Polster (1973) suggest that variables resulting in maintenance must be evaluated after a successful reversal has occurred. According to these criteria, it is possible that successful demonstrations of maintenance where no reversal occurred were not accepted for publication because functional control was not adequately demonstrated.

Another design used commonly in applied settings is the multiple baseline design (MBD) (Baer, Wolf, & Risley, 1968). This design is used when reversals are neither practical nor ethical. The MBD allows

for the demonstration of functional control (internal validity) through collection of data across multiple behaviors, settings, or individuals. As such, in MBDs, baseline measurements are taken on more than one behavior, in more than one setting, or on more than one individual at a time. The defining characteristic of the MBD is that the experimental contingencies are introduced at different times to the different baselines (Kazdin, 1994). A change in the treated behaviors while untreated behaviors remain at original baseline levels demonstrates the control of the contrived intervention. A more powerful demonstration then occurs when previously untreated behaviors are targeted with the intervention and subsequently change in the desired direction. In such presentations across behaviors it is assumed a priori that the behaviors to be treated are independent (Kazdin, 1973).

As mentioned above, the goals of acquisition and maintenance research differ. Rusch and Kazdin (1973) put it quite eloquently. Acquisition "depend[s] upon the ability of the subject to discern and respond to changes in the environment when that environment is altered; [maintenance] relies upon subject's failure to discriminate between those very same stimuli or, possibly, upon the subject's failure to discriminate among functionally similar stimul[i] . . . If the investigator is evaluating acquisition or maintenance, he or she should be able to conclude which variables are responsible for behavior change or maintenance" (pp. 131-132). Thus, maintenance cannot be hoped for, it must be programmed (Stokes & Baer, 1977).

Research Inclusion Criteria

The criteria used to include articles for critical review were that the research target at least one on-the-job or traffic-related safety behavior among employees at work, and that the study was published in a refereed journal between 1974 and 1996. Additionally, the investigators must have measured and reported follow-up data for the purpose of assessing post-intervention maintenance of behavior change. Follow-up phases are defined as the continuation of observation procedures (observing and recording the target behavior) at least two weeks after the intervention was formally withdrawn.

Certain components of some intervention packages cannot be withdrawn. Education and training are the most common examples. These situations are noted where applicable. Research not meeting the above

criteria is used for theoretical discussion, but was not included in the formal review of interventions for behavioral maintenance.

Literature was searched by reviewing the electronic databases PsycLit and the Social Sciences Index, as well as the psychological abstracts from 1974 to 1996 for items containing the key terms ''behavior-based'' and ''safety'' in various forms and combinations. The original search of the literature revealed 47 behavior-based studies (designed to improve industrial or traffic safety) since 1974. Of the 29 studies conducted in industry, 19 reported no follow-up data and were therefore eliminated, leaving a sample of 10 studies.

Of the 18 original traffic studies, five were eliminated because they targeted children or were not associated with a workplace. Interestingly, of the 13 remaining studies, all reported some follow-up data. The disparity in measuring follow-up between these two domains of research may be due to the nature of the observation protocols, the willingness of organizations to allow observations in their business, and costs associated with different observation procedures.

Twenty-three studies were included for review in the final sample. Follow-up periods lasted a minimum of two weeks to more than one year after the termination of the contrived intervention contingencies. The vast majority of follow-up periods lasted for more than one month after the termination of interventions.

Common research designs and interventions. Reversal designs were used predominantly in the traffic studies. Of the 13 traffic studies in the final sample, 12 used some form of an A-B-A design. The remaining study was a multiple baseline design (MBD) across settings. A larger variety of designs were used for the industry-based research, which included five MBDs, two reversals, and three (pre-post) designs with a follow-up phase after a period of time during which no observations were recorded.

The intervention strategies in both groupings of studies varied considerably. Incentives were the primary component in seven of the 13 traffic studies (54%). All of these provided participants with the opportunity to win cash or a prize in a lottery. A promise-card commitment was used in five studies targeting safety-belt use, including three that also used a raffle drawing incentive contingent on signing the promise and performing the target behavior. One study provided performance feedback as the primary intervention component.

Behavioral feedback was the most common intervention studied in the industry-based research. It was used six times (60%). In these six studies, feedback was combined with contingent one-on-one praise in three interventions and with goal setting in three others. A training component was also used in one of these latter intervention packages. Individual information feedback (in the form of amount of temporary personal hearing loss) was used in one study to increase the use of hearing protection. Of the remaining studies reviewed, token economies were used twice, and a promise-card commitment technique once. Because intervention *packages* have been used most often to improve occupational safety, discussion of these change programs are organized according to the key components for behavior change to have occurred.

A CRITICAL REVIEW OF MAINTENANCE IN OCCUPATIONAL SAFETY RESEARCH

Written Behavioral Commitment

Commitment strategies presumably gain their effectiveness from people's history of reinforcement for maintaining a consistency between saying and doing (Rogers-Warren & Baer, 1976). Similar to "goal statements," commitment strategies are more effective when they are described by participants to be freely chosen, public, and require a moderate amount of effort (Cialdini, 1993). Thus, a written commitment is more powerful than a spoken commitment and may be even more powerful if it is controlled by social contingencies, as when displayed publicly (cf. Hayes, 1989). As such, some scholars would presume that commitments are rules describing contingencies of negative reinforcement (e.g., Malott, 1992). A written commitment in the form of a pledge or promise card has been frequently used for large-scale behavior change interventions (e.g., Geller & Lehman, 1991; Katzev & Pardini, 1987).

Of the studies included in the final analysis, written commitment was the primary component in six interventions. Only one of these studies did not report some maintenance of behavior change. It is instructive to explore the characteristics shared by the more successful interventions, and investigate how they differed from those reporting less successful maintenance.

Written commitments for safety have been used in combination with: (a) education and awareness (Kello, Geller, Rice, & Bryant, 1988; Streff, Kalsher, & Geller, 1993), (b) incentives (Geller, Kalsher, Rudd, & Lehman, 1989; Nimmer & Geller, 1988), (c) disincentives (Kalsher, Geller, Clarke, & Lehman, 1989), and (d) reminder messages (Ludwig & Geller, 1991). The duration of the commitment interventions reviewed varied greatly, ranging from one week to six months. Follow-up data were collected at times ranging from a minimum of seven weeks following the termination of a pledge period to one year after withdrawing all contingencies. No noticeable differences or consistencies between the programs reporting maintenance and the one failing to produce maintenance were observed with regard to these phase durations. The level of maintenance obtained ranged from a low of approximately 15% over baseline to a three-fold increase. The average level of maintenance achieved was 75% over baseline levels of performance. Appendix A includes a table showing the approximate level and duration of maintenance per study reviewed.

Successful vs. unsuccessful maintenance with commitments. Streff, Kalsher, and Geller (1993) used a promise-card commitment strategy in combination with safety awareness sessions to increase the use of safety glasses among hourly employees at an electronics component firm (n = 51). The target behavior (safety-glasses use) and a non-target behavior (safety-belt use) were observed systematically during a baseline, intervention, and follow-up period. The specific durations of each separate experimental phase were not reported. Education, plus a two-month promise-card commitment, successfully improved the number of employees using safety glasses and also increased safety-belt use. Safety belts were never discussed with workers.

Observations taken one month and two months after the commitment period revealed that the use of both safety glasses and safety belts had returned to pre-intervention levels. This is the only study implementing a promise-card commitment technique that did not demonstrate some maintenance of the target behavior. However, ceiling effects may have contributed to this result since the baseline level of safety-glasses use was near 90%.

Compared with the other commitment interventions reviewed, only this program and the study reported by Kello, Geller, Rice, and Bryant (1988) did not publicize their promise program with a "promotional

blitz.'' Thus, lack of visibility (publicity) may have prevented necessary changes in the natural (e.g., social) contingencies controlling these behaviors. The visibility deficiency of the Kello et al. (1988) program may have been overcome by the fact that these researchers emphasized group participation and discussion in their awareness sessions, thus changing verbal behavior about safety in the workplace. That is, participative involvement may have functioned as an establishing operation, enhancing the social consequences of maintaining the use of safety belts for the 141 white-and blue-collar participants at this polyurethane manufacturing plant. This notion is supported by data indicating the duration of the promise-card commitment (i.e., one week, one month, or three months) did not influence safety-belt use during the intervention period or during the six-week follow-up (Kello et al., 1988).

All other studies reporting the use of promise-card commitment also reported successful maintenance of behavior change. For example, Nimmer and Geller (1988) increased safety-belt use among hospital workers by combining promise-card signing with cash rewards, contingent on signing the promise and emitting the target response. Similarly, Geller et al. (1989) increased safety-belt use on a university campus by combining buckle-up promises with a prize lottery in which raffle entry was contingent on signing the promise card. Each of these programs was highly publicized through various media before and throughout the six-month and three-week interventions, respectively. Safety-belt use remained at approximately 50% above baseline among promise-card signers for up to a year after termination of the intervention at the hospital, and nearly 20% above baseline for two months at the university after two applications of the promise-card lottery.

Similar maintenance of safety-belt use (20% above baseline) was obtained when incentives or disincentives were combined with a highly-publicized buckle-up promise at two Naval bases (Kalsher, Geller, Clarke, & Lehman, 1989). It is noteworthy that the baseline levels of above 50% safety-belt use in the two latter studies on a university campus and naval base were much higher than the baselines in the hospital setting studied by Nimmer and Geller (1988).

The successful programs had the following characteristics in common, regardless of the components combined with the promise-card technique: (a) signing the promise to emit the target response involved

individual choice; (b) the promise-card program itself was customized for the culture in which it was implemented; (c) when rewards were added, on-site workers delivered them; and (d) the single target behavior (i.e., use of a vehicle safety belt or available safety glasses) was easy and convenient to perform. In addition, with the exception of Streff et al. (1993) and Kello et al. (1988), the programs were highly publicized, and thus, highly visible prior to and during the intervention phase.

Methods of maintenance. From the perspective of the technologies of generalization, commitment strategies are best understood within the context of the method of mediated generalization (Baer, 1982; Stokes & Baer, 1977). That is, in making a commitment to emit a certain safety-related behavior, people provide their own antecedent conditions to set the occasion for the desired response. This may be facilitated, at first, by periodic reminders (Ludwig & Geller, 1991) or physically displaying the promise card (Geller & Lehman, 1991), but can be conceived of as rule-following.

A rule is simply a verbal description of a contingency (Malott, 1992). The rule invoked by self-selected antecedents in the commitment studies reviewed may describe some negative social consequences for failure to follow through on a promise. Powerful social contingencies may ensue as social benefits are observed among people successfully performing the target response, and cultural contingencies change such that performing the target behavior is now reinforced. It is likely that written commitments promote rules which shape the behavior and the social benefits of keeping one's promise enhance response maintenance.

Incentive/Rewards

Large-scale incentive/reward programs often involve a raffle or lottery for cash or prizes. Seven studies reviewed used incentive/reward contingencies as the primary component of a large-scale safety intervention. Four of these programs reported some maintenance of behavior change. Those unsuccessful at obtaining maintenance combined antecedent prompts with: (a) occasional contingent rewards (Geller, Davis, & Spicer, 1983); (b) feedback and intermittently scheduled cash prizes (Cope, Smith, & Grossnickle, 1986); and (c) a bingo-like game to win prizes (Geller, 1983).

In general, the incentive/reward interventions most successful at producing maintenance were shorter in duration, ranging from three to five weeks; the less successful interventions varied in length from approximately four to ten weeks. For all seven studies reviewed, follow-up data were collected at times ranging from 11 workdays (two weeks) following the formal intervention to ten months later. All of the incentive/reward interventions were successful in producing desired behavior change while the program was in place. And, all of these safety programs targeted only one behavior (i.e., safety-belt use). Of the incentive/reward studies reporting successful maintenance of behavior change, the level obtained ranged from a low of 40% over baseline to a high of almost four times the baseline level. The average level of maintenance obtained across the successful studies was approximately 100% over baseline or two times the baseline performance.

Successful vs. unsuccessful maintenance with incentives. A consistent finding in the incentive/reward programs reviewed is that the incentive component announced a low probability of receiving a reward contingent on emitting the target safety-related behavior. A noteworthy difference between successful and unsuccessful maintenance-producing programs is that the researchers delivered the consequences in the unsuccessful maintenance interventions, whereas in the successful programs on-site workers delivered the prize or drew the winning tickets in the prize or cash lotteries. Campus police, for example, delivered prizes to safety-belt users in an attempt to institutionalize an incentive program for promoting safety-belt use on a university campus (Rudd & Geller, 1985). Furthermore, at a large munitions plant of 3023 employees, plant safety personnel helped to administer the contingencies by delivering prizes for safety-belt use to the raffle winners (Geller, 1984). These incentive programs produced maintenance of safety-belt use 46% and 75% over baseline, respectively.

A worker drew the winning ticket in the license plate lottery used by Geller and Hahn (1984). This intervention produced maintenance of safety-belt use among white-collar workers of 46% and 77% across two manufacturing industries, but only produced maintenance (300% over baseline) among blue-collar workers when the incentives were combined with safety-belt education sessions. In all of these studies, the winners of the prizes were publicized through various media.

Thus, consistent with the successful commitment strategies, the incentive programs producing the most maintenance were highly visible.

All of these incentive/reward programs provided a low probability of being rewarded and a delay until the delivery of the consequence. Thus, occasions of reinforcement from occasions of non-reinforcement were essentially indiscriminable. Yet, for a chance to receive a reward, the behavior had to be observed. Interestingly, the programs successful at maintaining the target behavior were implemented on a much larger scale (n = 3000 to 25,000) than were programs reporting transient intervention effects (n = 475 to 3000).

Without asking participants if they calculated the costs versus the benefits of an incentive program, objective evidence for successful maintenance by large-scale incentive/reward interventions appears to rest on one common factor–the use of on-site workers to help administer program contingencies and deliver rewarding consequences. Using regular workers may indeed interact with and change the natural contingencies controlling behavior, including verbal behavior.

In sum, the characteristics common to the successful incentive/reward programs included: (a) the programs were well publicized in advance and during the intervention, and thus highly visible; (b) the intervention period was shorter in duration; (c) incentive/reward programs emphasized that workers had "choice" in participating; (d) on-site workers helped to administer the program contingencies and deliver the rewards; and (e) the interventions were implemented in larger populations, resulting presumably in the "perception" of a lower probability of any one individual winning the prize or monetary raffle.

Methods of maintenance. From the perspective of the technologies of generalization, the incentive/reward interventions reviewed gain their maintenance effectiveness from the methods of indiscriminable contingencies *and* common stimuli. These techniques rely on individuals' failure to discriminate occasions of reinforcement from occasions of non-reinforcement. The experimental literature is replete with examples of how, once acquired, behaviors reinforced on intermittent schedules are highly resistant to extinction. That the physical presence of workers themselves can exert stimulus control adds to the likelihood of obtaining long-term impact.

In many incentive programs, the highly variable probability of receiving contingent rewards provides just enough incentive to get the behavior started. This is sometimes accomplished by providing fre-

quent secondary reinforcers on an intermittent schedule, that can then be exchanged for a primary reinforcer at a later time if certain conditions are met (Geller, 1983; Geller & Hahn, 1984; Geller, Paterson, & Talbott, 1982). Finally, the various on-site workers delivering the consequences may condition stimuli that set the occasion for the target behavior. This is operationally the method of common stimuli.

Behavioral Feedback

Behavior-based feedback was the primary intervention component in eight of the industry-based safety studies reviewed for this paper. Such feedback can be based on individual or group performance. It can be public or private, and is sometimes combined with social reinforcement for success or corrective comments for less than desired performance.

Of the eight studies reviewed, only three did not report successful maintenance of behavior change. The interventions that did not maintain behavior change combined feedback with: (a) education to increase the use of safety belts among 370 workers at a training institute (Grant, 1990); (b) education and contingent praise for each successive improvement to increase the number and quality of injury reports completed by 12 foremen at a papermill (Fox & Sulzer-Azaroff, 1987); and (c) goal setting and periodic response-contingent praise to improve housekeeping and increase the safe performance of 162 workers at a wholesale bakery (Komaki, Barwick, & Scott, 1978). Among these feedback programs, interventions ranged in duration from three to 11 weeks, and follow-up data were collected at times ranging from two weeks following the termination of the formal intervention to one year later.

Interventions in the five studies finding maintenance lasted a minimum of three weeks to a maximum of one year. Follow-up data were collected at times ranging from six weeks to one year following the withdrawal of performance feedback contingencies. A particular weakness of the feedback research using MBDs was the failure of investigators to report detailed information about the duration of each research phase.

Successful maintenance was achieved by combining feedback with: (a) education regarding the benefits of using hearing protection and the costs of hearing loss to increase the use of earplugs among 2000 workers at a metal fabrication plant (Zohar, Cohen, & Azar, 1980);

(b) positive and corrective feedback to reduce environmental hazards and improve housekeeping behaviors among 128 operators at a custom-product manufacturing facility (Sulzer-Azaroff & de Santamaria, 1980); and (c) training/education and participative goal setting to eliminate hazards at a commercial shipyard with 2,400 employees (Saarela, 1990) and improve fork-truck driving among 96 workers at two retail distribution warehouses (Cohen & Jensen, 1984). Finally, Alavosius and Sulzer-Azaroff (1986) achieved maintenance using individual spoken and written behavioral feedback to improve the lifting and transfer behaviors of six direct care staff members at a state residential school for the developmentally and physically disabled.

The level of maintenance obtained using behavior-based feedback ranged from a low of 10% over baseline levels of safety-belt use and a 40% reduction in injuries to a high of more than 100% over baseline levels of overall safety-related behaviors in the workplace. The average level of maintenance obtained was not calculated because of the disparity of dependent measures used in these studies and the lack of consistency across studies in how the results were reported.

Successful versus unsuccessful maintenance with feedback. The unsuccessful programs appeared not to be very visible or especially customized for the culture in which they were implemented. It is noteworthy that the feedback in programs unsuccessful at promoting behavioral maintenance was delivered by the researchers. In contrast, all feedback programs that obtained some maintenance used on-site workers to provide feedback, praise, or corrective action in various combinations.

The feedback packages that successfully produced maintenance shared the following characteristics. First, each successful industrial study targeted multiple behaviors, with the exception of Zohar et al. (1980) who provided daily audiometric feedback to a random sample of employees. Thus, the intervention evaluated by Zohar et al. (1980) probably produced its effects through rule-governance as relatively few employees received daily hearing tests (7.5%). The results of these tests were posted prominently in the plant, and labeled conspicuously with the name of the employee and the extent to which he/she used hearing protection that day.

Second, feedback was provided as a global representation of overall safe performance, instead of being presented specifically on individual

behaviors. Third, on-site workers provided all performance feedback and contingent praise or correction. Finally, the most successful feedback packages emphasized safe performance, and thus could be complemented with support and recognition from others. In fact, the authors of the interventions producing maintenance reported that supportive verbal praise for improvements was a common side-effect of their behavior-based feedback interventions, and thus reflected a change in the verbal behavior related to safety.

Methods of maintenance. Providing general feedback, as opposed to specific feedback, and targeting multiple behaviors decrease the likelihood that only one behavior will be emitted because of the consequences provided for performance improvement. Thus, specific response-consequence contingencies are less discriminable than if feedback were provided explicitly for each behavior targeted. Indeed, Baer (1982) discussed the use of indiscriminable contingencies as a method of facilitating generalization.

Another characteristic shared by the successful studies was that they all emphasized safe performance and recognition for success, with the single exception of Cohen and Jensen (1984) who provided group feedback in the form of mean error-rate of fork-truck driving behaviors. When emphasizing safe performance, achievements are highlighted and are likely to be paired with social reinforcement resulting from successive goal attainment, whether external or self-selected. Although Baer (1982) argued against programming social approval (natural communities of reinforcement), social reinforcement can be considered a positive outcome of performance accomplishment (Gilbert, 1978) which could increase the probability of the safety-related response.

Finally, without exception, all of the feedback interventions which produced some maintenance of behavior change used on-site workers to provide the feedback and or contingent praise. As noted above, this characteristic was conspicuously absent from the feedback packages that did not produce maintenance. From the perspective of the technologies of generalization, it could be argued that the deliverers of contingencies assumed stimulus control of continued safe performance. This notion is also supported by the observation that recipients of an industry-based incentive/reward program to increase safety-belt use were more likely to be buckled-up *when leaving* work than when arriving (Geller, 1983).

Token Economies

Only one token economy met the criteria for inclusion in the present review. Zohar and Fussfeld (1981) reported maintenance of earplug use after tokens were contingent on earplug use for six weeks. Tokens were dispensed daily and their monetary value varied as a function of the proportion of workers using earplugs at the time they were dispensed. That is, the worth of the tokens received on a given day increased or decreased as the number of earplug users increased or decreased on a daily basis. This program was developed by the workers, and supervisors delivered the tokens.

It could be speculated that the success of participative development may be partly due to the fact that program contingencies are kept within a range described as acceptable by the program recipients. These "rules" are influenced by individual histories of reinforcement. People will emit behaviors that are successive approximations of desired behaviors because of their history of reinforcement for similar behaviors, or because it is "rule-governed."

The maintenance function served by using on-site workers to deliver consequences (tokens) was discussed above. Consider, however, that in this program the tokens (secondary reinforcers) gained their value according to group performance and were exchanged for items the workers selected (from a menu of available rewards). Thus, individual behavior was reinforced by contingencies for the selection of prizes, and group performance was reinforced by contingencies determining the value of a token. Up to six months following termination of their token economy, Zohar and Fussfeld (1981) reported maintenance of earplug use 50% over baseline.

Another safety program using token economies was evaluated by Fox, Hopkins, and Anger (1987). The token economies were still in place at two coal mines at the time of their report. As such, this study, although often cited as a seminal large-scale injury reduction program, did not meet the criteria set for inclusion in this review. This study is only mentioned here because it is referenced so often, and to highlight the distinction between the present definition of maintenance and the concept of "institutionalization." Fox et al. (1987) discussed the success of their token economies in terms of behavioral maintenance, even when actual behaviors were not observed nor reported. Rather, their contrived program was institutionalized in these mines for over

ten years. A behavior analysis of institutionalization is beyond the scope of the present paper.

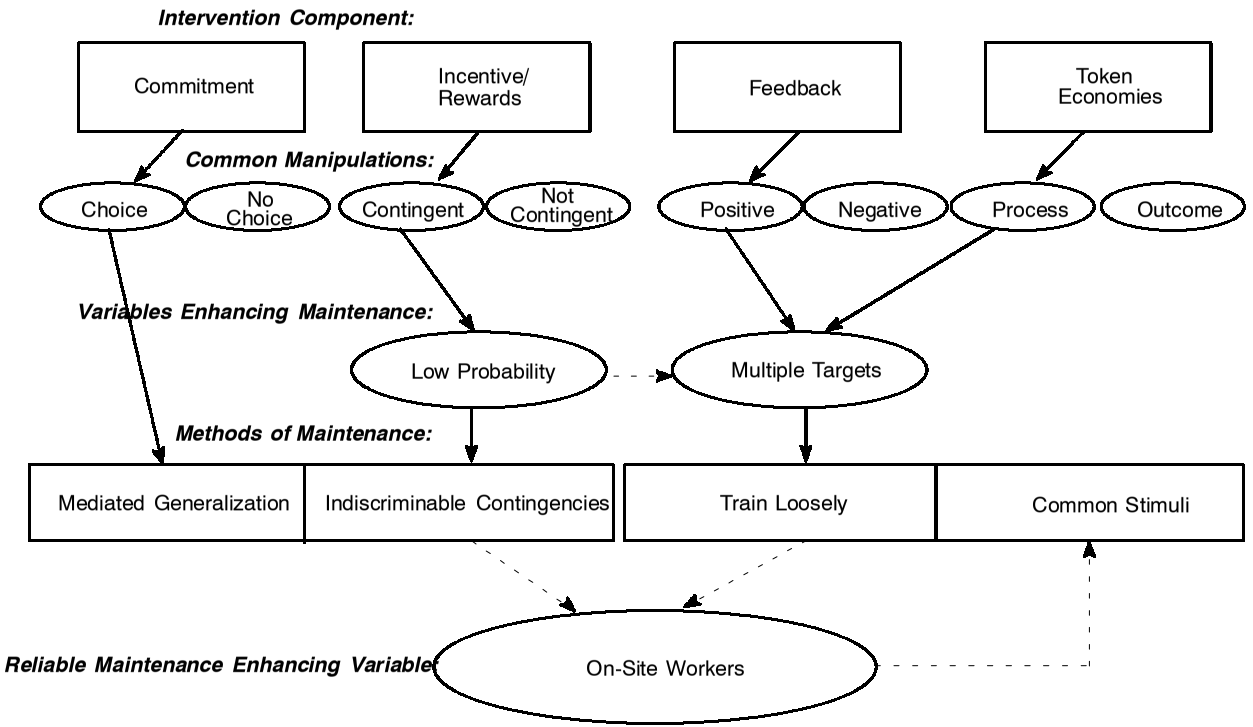
WHAT ARE THE NECESSARY AND SUFFICIENT CONDITIONS FOR MAINTENANCE?

Maintenance is a characteristic of numerical data that describes some relevance beyond the *time* of their origin (Johnston & Pennypacker, 1980). Put differently, maintenance is the occurrence of a target behavior at a time when the original contingencies that reinforced behavior during acquisition are no longer relevant. Through our systematic review of the literature, it is apparent that at a minimum, regardless of intervention type, large-scale behavior change programs need to include certain characteristics for maintenance to occur. To operationalize and apply these program characteristics, it is useful to explore some general predictions regarding maintenance that emanate from the current review. Subsequently, these characteristics and predictions are related to the specific intervention strategies reviewed above.

Several general predictions follow from Baer's (1982) discussion of generalization as a model. Specifically, maintenance should be facilitated when: (a) multiple behaviors are targeted and consequences are contingent on a global representation of those behaviors; (b) intervening on a target behavior takes place in the presence of environmental stimuli or events that occur naturally in the environment where the behavior is desired; (c) low probability, delayed, or intermittent consequences are provided contingent on the occurrence of a target behavior; (d) interventions generate a three-term contingency that can be stated as a "memorable rule"; and (e) verbal behavior about safety changes in conjunction with the target behavior(s). It should be emphasized that these predictions are based on the assumption that the interventions to which they pertain produced initial behavior change.

Figure 1 shows the relationship between the interventions reviewed and Baer's methods of generalization. This is a flow chart indicating the primary variables necessary to produce maintenance as a function of the techniques discussed by Baer (1982). The top row indicates the primary intervention components reviewed. The second row indicates common manipulations within these strategies. The third row indi-

FIGURE 1. The Relationship Between Intervention Strategies and Maintenance Facilitating Methods



cates variables that enhance the probability maintenance will be obtained. Finally, on-site workers are indicated as a reliable producer of maintenance because they are stimuli naturally occurring in the work environment that can exert stimulus control over desired behavior.

More specifically, Figure 1 indicates that the use of on-site workers will strengthen the effect of interventions on maintenance by adding the element of ''common stimuli'' to other maintenance-producing strategies. For example, incentive/reward programs with a low probability of individual pay-off impact behavioral maintenance as a result of the method of indiscriminable contingencies. If on-site workers are used to deliver the contingent rewards, then added to the effects of indiscriminable contingencies are the effects of the method of common stimuli. Furthermore, if the low probability of rewards is contingent on any of several target behaviors (multiple targets), then the effects of training loosely could be added to the effects described above. Thus, it is predicted that the probability of obtaining maintenance is enhanced as a function of combining the effects of multiple methods of generalization. Presumed additional effects are indicated by the dotted lines in Figure 1.

Table 1 summarizes the primary characteristics presumed to enhance behavioral maintenance for each intervention strategy reviewed, and the method of generalization to which each characteristic corresponds.

CONCLUSIONS

Many researchers do not set out to demonstrate long-term intervention impact, and thus omit some detail in their reports that might allow for more specific conclusions or patterns to emerge. Based on the current review, however, evidence is accumulating that reward schedules which are just sufficient to initiate behavior change are more likely to produce maintenance in occupational settings than more powerful rewards. This is demonstrated by the relative effectiveness of low probability incentives. Global representations of behavior will more likely produce maintenance than references to specific behavior, provided that just enough information is provided. This was demonstrated by the effectiveness of providing global performance feedback regarding multiple target behaviors (e.g., Sulzer-Azaroff & de Santamaria, 1980). Also, commitment strategies accompanied by informa-

TABLE 1. Summary of Maintenance Producing Characteristics and Generalization Facilitating Methods

Intervention	Characteristic	Method
Commitment Strategies	Choice-generates self-rules, provide own discriminative stimuli and consequences	**Mediated Generalization**
	On-site workers-naturally occurring discriminative stimuli	**Method of Common Stimuli**
Incentive/Reward Strategies	Low probability contingent rewards	**Method of Indiscriminable Contingencies**
	On-site workers-naturally occurring discriminative stimuli	**Method of Common Stimuli**
Performance Feedback	Multiple target behaviors	**Method of Training Loosely**
	Feedback provided as global representation of overall behaviors	**Method of Indiscriminable Contingencies**
	On-site workers-naturally occurring discriminative stimuli	**Method of Common Stimuli**
Token Economies	Rewards contingent on multiple behaviors in multiple settings	**Method of Training Loosely**
	Low probability contingent rewards	**Method of Indiscriminable Contingencies**
	On-site workers-naturally occurring discriminative stimuli	**Method of Common Stimuli**
	Participative development and choice generates self-rules, provide own discriminative stimuli and consequences	**Mediated Generalization**

tion regarding the rationale for performing a target response can produce maintenance by specifying rules describing response-consequences contingencies.

Future research needs to evaluate, systematically, the effects on maintenance of using on-site workers in large-scale safety interventions. In addition, maintenance of a target behavior needs to be evaluated as a function of the frequency and magnitude of rewards used in incentive/reward programs and token economies. A systematic evaluation of providing specific versus global performance feedback when

targeting multiple behaviors to produce maintenance is also needed. Additionally, researchers should decide at the onset of their study whether they want to demonstrate functional control of their intervention package or to demonstrate long-term impact. As a result, they should describe their methods with enough detail to allow critical evaluations of the contingencies used to produce acquisition and/or maintenance of behavior change.

REFERENCES

Alavosius, M. P., & Sulzer-Azaroff, B. (1986). The effects of performance feedback on the safety of client lifting and transfer. *Journal of Applied Behavior Analysis*, 19, 261-267.

Baer, D. M. (1982). The role of current pragmatics in the future analysis of generalization technology. In R. B. Stuart (Ed.), *Adherence, compliance, and generalization in behavioral medicine* (pp. 192-212). New York: Brunner/Mazel Inc.

Baer, D. M., Wolf, M. M., & Risley, T. (1968). Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis*, 1, 91-97.

Bandura, A. (1986). *Social foundation of thought and action: A social cognitive theory.* Englewood Cliffs, NJ: Prentice Hall.

Biglan, A. (1987). A behavior-analytic critique of Bandura's self-efficacy theory. *The Behavior Analyst*, 10, 1-15.

Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research and teaching. In N. L. Gage (Ed.), *Handbook of research on teaching* (pp. 171-246). Chicago: Rand McNally.

Cialdini, R. B. (1993). *Influence: Science and practice.* New York: Harper Collins.

Cohen, A., & Margolis, B. (1973). Initial psychological research related to the occupational safety and health act of 1970. *American Psychologist*, 27, 600-606.

Cohen, H. H., & Jensen, R. C. (1984). Measuring the effectiveness of an industrial lift truck safety training program. *Journal of Safety Research*, 15, 125-135.

Cooper, J. O., Heron, T. E., & Heward, W. L. (1987). *Applied behavior analysis*. Columbus, OH: Merrill Publishing.

Cope, J. G., Smith, G. A., & Grossnickle, W. F. (1986). The effect of variable-rate cash incentives on safety belt use. *Journal of Safety Research*, 17, 95-99.

Fox, C. J., & Sulzer-Azaroff, B. (1987). Increasing completion of accident reports. *Journal of Safety Research*, 18, 65-71.

Fox, D. K., Hopkins, B. L., & Anger, W. K. (1987). The long-term effects of a token economy on safety performance in open-pit mining. *Journal of Applied Behavior Analysis*, 20, 215-224.

Geller, E. S. (1983). Rewarding safety-belt usage at an industrial setting: Tests of treatment generality and response maintenance. *Journal of Applied Behavior Analysis*, 16, 189-202.

Geller, E. S. (1984). A delayed reward strategy for large-scale motivation of safety-belt use: A test of long-term impact. *Accident Analysis & Prevention*, 10(5,6), 457-463.

Geller, E. S. (2001). *The psychology of safety handbook*. Boca Raton, FL: CRC Press.

Geller, E. S., Davis, L., & Spicer, K. (1983). Industry based incentives for promoting seat-belt use: Differential impact on white-collar versus blue-collar employees. *Journal of Organizational Behavior Management*, 5(1), 17-29.

Geller, E. S., & Hahn, H. A. (1984). Promoting safety belt use at industrial sites: An effective program for blue collar employees. *Professional Psychology*: *Research and Practice*, 15, 553-563.

Geller, E. S., Kalsher, M. J., Rudd, J. R., & Lehman, G. R. (1989). Promoting safety-belt use on a university campus: An integration of commitment and incentive strategies. *Journal of Applied Social Psychology*, 19, 3-19.

Geller, E. S., & Lehman, G. R. (1991). The buckle-up promise card: A versatile intervention for large-scale behavior change. *Journal of Applied Behavior Analysis*, 24, 91-94.

Geller, E. S., Paterson, L., & Talbott, E. (1982). A behavioral analysis of incentive prompts for motivating seat-belt use. *Journal of Applied Behavior Analysis*, 15, 403-415.

Gilbert, T. F. (1978). *Human competence*: *Engineering performance*. New York: McGraw-Hill.

Grant, B. A. (1990). Effectiveness of feedback and education in an employment based seat-belt program. *Health Education and Research*: *Theory and Practice*, 5(2), 197-205.

Hayes, S. C. (1989). *Rule-governed behavior*: *Cognition*, *contingencies*, *and instructional control*. New York: Plenum Press.

Johnston, J. M., & Pennypacker, H. S. (1980). *Strategies and tactics of human behavioral research*. Hillsdale, NJ: Lawrence Erlbaum.

Jones, R. R., Vaught, R. S., & Weinrott, M. (1977). Time-series analysis in operant research. *Journal of Applied Behavior Analysis*, 10, 151-166.

Kalsher, M. J., Geller, E. S., & Clarke, S. W., & Lehman, G. R. (1989). Safety-belt promotion on a naval base: A comparison of incentives vs. disincentives. *Journal of Safety Research*, 20, 103-113.

Katzev, R. D., & Pardini, A. U. (1987). The comparative effectiveness of reward and commitment approaches in motivating community recycling. *Journal of Environmental Systems*, 13, 245-254.

Kazdin, A. E. (1973). Covert modeling and the reduction of avoidance behavior. *Journal of Abnormal Psychology*, 81, 87-95.

Kazdin, A. E. (1977). Assessing the clinical or applied importance of behavior change through social validation. *Behavior Modification*, 1(4), 427-452.

Kazdin, A. E. (1994). *Behavior modification in applied settings*. Pacific Grove, CA: Brooks/Cole.

Kazdin, A. E., & Polster, R. (1973). Intermittent token reinforcement and response maintenance in extinction. *Behavior Therapy*, 4, 386-391.

Kello, J. E., Geller, E. S., Rice, J. C., & Bryant, S. L. (1988). Motivating auto safety belt wearing in industrial settings: From awareness to behavior change. *Journal of Organizational Behavior Management*, 9, 7-21.

Komaki, J., Barwick, K. D., & Scott, L. R. (1978). A behavioral approach to occupa-

tional safety: Pinpointing and reinforcing safe performance in a food manufacturing plant. *Journal of Applied Behavior Analysis*, 11, 434-445.

Ludwig, T. D., & Geller, E. S. (1991). Improving the driving practices of pizza deliverers: Response generalization and moderating effects of driving history. *Journal of Applied Behavior Analysis*, 24, 31-44.

Malott, R. W. (1992). A theory of rule-governed behavior and organizational behavior management. *Journal of Organizational Behavior Management*, 12, 45-65.

Nimmer, J. G., & Geller, E. S. (1988). Motivating safety belt use at a community hospital: An effective integration of incentive and commitment strategies. *American Journal of Community Psychology*, 16, 381-394.

Premack, D. (1962). Reversibility of the reinforcement relation. *Science*, 136, 255-257.

Rachlin, H. (1991). *Introduction to modern behaviorism*. New York: W.H. Freeman and Company.

Reber, R. A., & Wallin, J. A. (1983). Validation of a behavioral measure of occupational safety. *Journal of Organizational Behavior Management*, 5, 69-77.

Rogers-Warren, A., & Baer, D. M. (1976). Correspondence between saying and doing: Teaching children to share and praise. *Journal of Applied Behavior Analysis*, 9, 335-354.

Rudd, J. R., & Geller, E. S. (1985). A university-based incentive program to increase safety-belt use: Toward cost-effective institutionalization. *Journal of Applied Behavior Analysis*, 18, 215-226.

Rusch, F. R., & Kazdin, A. E. (1981). Toward a methodology of withdrawal designs for the assessment of response maintenance. *Journal of Applied Behavior Analysis*, 14, 131-140.

Saarela, K. L. (1990). An intervention program utilizing small groups: A comparative study. *Journal of Safety Research*, 21, 149-156.

Skinner, B. F. (1953). *Science and human behavior*. New York: MacMillan.

Stokes, T. F., & Baer, D. M. (1977). An implicit technology of generalization. *Journal of Applied Behavior Analysis*, 10, 349-367.

Streff, F. M., Kalsher, M. J., & Geller, E. S. (1993). Developing efficient workplace safety programs: Observations of response covariation. *Journal of Organizational Behavior Management*, 13, 3-14.

Sulzer-Azaroff, B., & de Santamaria, M. C. (1980). Industrial safety hazard reduction through performance feedback. *Journal of Applied Behavior Analysis*, 13, 287-295.

Zohar, D., & Fussfeld, N. (1981). Modifying earplug wearing behavior by behavior modification techniques: An empirical evaluation. *Journal of Organizational Behavior Management*, 3(2), 41-52.

Zohar, D., Cohen, A., & Azar, N. (1980). Promoting increased use of ear protectors in noise through information feedback. *Human Factors*, 22, 69-79

APPENDIX A. Characteristics of Large-Scale Interventions Producing Long-Term Maintenance

Authors	Intervention Package	Design	Key Variables	Level and Duration of Effect
Alavosius & Sulzer-Azaroff (1986)	Performance feedback; contingent praise Target: 18 components of client lifting and transfer procedures	Multiple Baseline across Subjects	Written and verbal individual feedback; feedback delivered by on-site workers; natural S^{R+} (less disruptive client behavior); multiple target behaviors	Follow-up observations initiated one week after Tx and probed at 2 weeks, 1, 2, and 7 months; behavior change maintained throughout Level: Phase means not provided, only visual inspection
Cohen & Jensen (1984)	Feedback; training; and goal-setting Target: 14 fork-lift driving behaviors	A-B between groups comparison (training vs. training plus feedback) with no treatment control group	Highly visible; choice (participative goal-setting by department); feedback (emphasized mean error rate) delivered by on-site workers and posted for group; sometimes accompanied by one-on-one rewarding or corrective feedback; multiple target behaviors	Follow-up observations initiated 3 months after Tx; both training alone, and training plus feedback demonstrated maintenance; overall plant demonstrated maintenance after control group trained Level: Phase means not provided; approximately 70% net reduction in errors associated with fork-lift driving
Fox, Hopkins, & Anger (1987)	Token Economy Target: hazard reduction; operations; PPE use	Multiple Baseline across settings	Tokens contingent on avoiding injury; tokens lost for having injury	Intervention never withdrawn (10 years) **See text for explanation of inclusion**

APPENDIX A (continued)

Authors	Intervention Package	Design	Key Variables	Level and Duration of Effect
Geller (1984)	Incentive prompt with lottery entry contingent on target behavior Target: safety-belt use	A-B with extended withdrawal and no treatment control group	Choice; customized by posting winners names; highly visible; contingencies administered by on-site workers	Follow-up observations initiated 25 days after Tx and data collected as probes for 10 months; Level: 150% above baseline after 3 months; 75% above baseline after 10 months
Geller & Hahn (1984)	Incentive and feedback vs. incentive and performance feedback plus awareness Target: safety-belt use	A-B-A reversal with multiple comparison groups	Choice; customized; visible; contingencies administered by on-site workers	Follow-up initiated as immediate withdrawal; maintenance obtained for 3 weeks was mediated by awareness sessions; feedback not withdrawn Level: Site1: 46% above baseline among white-collar workers only; Site 2: 77% above baseline among white-collar workers, 300% above baseline among blue-collar workers (who received education through awareness sessions)
Geller, Kalsher, Rudd, Lehman (1989)	Written commitment; raffle incentive (contingent on pledge and target behavior) Target: safety-belt use	A-B-A reversal (multiple treatments)	Choice; customization; high visibility; on-site workers administered contingent rewards	Follow-up observations initiated 2 months after final withdrawal; maintenance for up to 6 days Level: Faculty and Staff: 23% above baseline; Students: 16% above Baseline
Geller, Paterson, & Talbott (1982)	Incentive prompt and contingent vs. non-contingent reward Target: safety-belt use	A-B-A reversal with comparison group	Choice; moderately visible; secondary reinforcers; multiple reward opportunities; rewards associated with institution	Follow-up observations occurred for 11 days immediately after Tx; maintenance for contingent rewards only and increased as a function of reward experiences; non-contingent reward group served as control Level: 40% above baseline

Grant (1990)	Performance feedback; education; reminder prompts Target: safety-belt use	Successive treatments with return to baseline and no treatment control group	Highly visible; positive group feedback provided daily and posted prominently Level:	Follow-up observations initiated 2, 4, and 6 weeks after Tx; maintenance occurred up to four weeks (no phase means provided)-approximate level of maintenance: 10% above baseline
Kalsher, Geller, Clarke, & Lehman (1989)	Incentives vs. disincentives and written commitment Target: safety-belt use	Multiple baseline across sites	Choice; customization; high visibility; on-site workers administered contingent rewards/penalties	Follow-up observations initiated 6 months after withdrawal of contingencies Level: Site 1: 21% above baseline; Site 2: 20% above baseline
Kello, Geller, Rice, & Bryant (1988)	Awareness sessions vs. awareness sessions and written commitment Target: safety-belt use	A-B-A reversal with no treatment control	Choice (duration assigned); participation	Follow-up observations initiated 1 week after longest pledge (3 months); maintenance obtained for 7 weeks regardless of pledge vs. no pledge and pledge length Level: 187% above baseline
Ludwig & Geller (1991)	Written commitment; education and awareness sessions; reminder prompts Target: safety-belt use	Multiple baseline across sites	Choice; participation; customization; high visibility	Follow-up observations initiated 2 weeks after Tx; maintenance for up to 11 weeks Level: Site 1: 68% above baseline; Site 2: 200% above baseline
Nimmer & Geller (1988)	Education; written commitment; and raffle incentive (contingent on pledge and target behavior) Target: safety-belt use	A-B-A reversal	Choice; customization; high visibility; on-site workers administered contingent rewards, distributed pledge cards, and posted the names of winners	Follow-up observations initiated 4 months after withdrawal; maintenance (below Tx; but above baseline levels for pledge-signers) and pledge non-signers (whose baseline rates were significantly lower) Level: Pledge-signers: 53% above baseline

APPENDIX A (continued)

Authors	Intervention Package	Design	Key Variables	Level and Duration of Effect
Rudd & Geller (1985)	Incentive-raffle entry contingent on target behavior Target: safety-belt use	A-B-A reversal (multiple treatments)	Choice; customization; highly visible; on-site workers administered contingent rewards	Each successive withdrawal initiated the first weekday following the termination of Tx; maintenance occurred as a result of less return to baseline levels after each succesive treatment; final withdrawal occurred approximately 5 months after first Tx Level: 46% above first baseline
Saarela (1990)	Feedback; goal-setting; training and education Target: housekeeping; operations; general safety	A-B with follow-up after 1 year	Highly visible; customized; promoted choice through participation; on-site workers made and posted daily observations; natural S^{R+} (cleaner work environment) multiple target behaviors	Follow-up initiated the year after Tx; demonstrated maintenance through reduction of accidents for year following Tx Level: 40% reduction in accidents
Sulzer-Azaroff & de Santamaria (1980)	Performance feedback and contingent praise or correction Target: 18 environmental hazards (housekeeping)	Multiple baseline across work departments	Highly visible; positive group feedback posted prominently by on-site workers and visible management support; multiple target behaviors	Follow-up initiated 3 days after formal feedback Tx; and probes occurred 2 and 6 weeks, and 4 months later; behavior maintained throughout follow-up and management continued program upon termination of study Level: Phase means not discussed, only provided visual inspection
Zohar, Cohen, & Azar (1980)	Feedback (audiometric hearing tests); and education Target: hearing protection	A-B with extended withdrawal and no treatment control	Visible; demonstrated actual negative outcome of at-risk behavior; promoted discussion and vicarious experience; feedback delivered by on-site workers and posted prominently	Follow-up observations initiated immediately after Tx; continuous improvement of target demonstrated for 5 months Level: 119% above baseline
Zohar & Fussfeld (1981)	Token Economy Target: hearing protection	Multiple baseline across work shifts	Highly visible; customized and participative (developed by employees); tokens dispensed daily by supervisor; tokens' worth determined by proportion of employees using earplugs at time of dispensation; choice, choice, choice	Follow-up observations occurred immediately after Tx and then again 6 months later; behavior change maintained at intervention levels on all workshifts Level: Dept. 1: 36% above baseline; Dept. 2: 63% above baseline