JOURNAL OF APPLIED BEHAVIOR ANALYSIS 2002, **35,** 431–464 NUMBER 4 (WINTER 2002)

ON THE STATUS OF KNOWLEDGE FOR USING PUNISHMENT: IMPLICATIONS FOR TREATING BEHAVIOR DISORDERS

DOROTHEA C. LERMAN AND CHRISTINA M. VORNDRAN

LOUISIANA STATE UNIVERSITY AND
THE LOUISIANA CENTER FOR EXCELLENCE IN AUTISM

In this paper, we review basic and applied findings on punishment and discuss the importance of conducting further research in this area. The characteristics of responding during punishment and numerous factors that interact with basic processes are delineated in conjunction with implications for the treatment of behavior disorders in clinical populations. We conclude that further understanding of punishment processes is needed to develop a highly systematic, effective technology of behavior change, including strategies for improving the efficacy of less intrusive procedures and for successfully fading treatment.

DESCRIPTORS: behavior disorders, functional analysis, punishment, treatment

Punishment is generally defined as an environmental change contingent on behavior that produces a decrease in responding over time (Michael, 1993). Numerous procedural variations of punishment have been developed for clinical use. Results of research conducted over the past four decades have shown that punishment is effective in reducing problem behavior in clinical populations, and in some cases, may be an essential component of treatment (see Kazdin, 2001, and O'Brien, 1989, for reviews of this literature). However, more knowledge is needed about factors that may influence the effects of punishment on problem behavior. Few strategies have been identified for enhancing the effectiveness of less intrusive punishment procedures, for attenuating undesirable aspects of punishment, or for successfully fading treatment with punishment.

The direct and indirect effects of punishment have been studied extensively in the laboratory. Nevertheless, basic research on punishment has been declining rapidly despite substantial gaps in knowledge (Baron, 1991; Crosbie, 1998). The generality of basic findings to clinical populations and problems also is questionable (Hayes & McCurry, 1990). Most studies evaluated the effects of intense, unconditioned punishers (e.g., electric shock), and a number of important relations have not yet been replicated with humans or clinically relevant punishers.

The purpose of this paper is to review basic and applied research findings on punishment, identify gaps in the literature, and discuss the implications of these findings for the use of punishment in clinical settings. Basic findings that contradict common assumptions about punishment effects found in textbooks and review papers and that help elucidate inconsistent results on punishment in the applied literature are highlighted. The main premise of this paper is that further understanding of punishment processes may lead to an improved technology of behavior change. We extend the most recent review papers on punishment (Matson & DiLorenzo, 1984; Van Houten, 1983) by (a)

We thank Don Baer, Alan Baron, Linda LeBlanc, Tony Nevin, and the anonymous reviewers for their comments on earlier versions of this paper.

Reprints may be obtained by contacting Dorothea C. Lerman, 236 Audubon Hall, Louisiana State University, Baton Rouge, Louisiana 70803 (e-mail: dlerman@lsu.edu).

providing a broader overview of the direct and indirect effects of punishment and factors that influence basic processes (e.g., history), (b) identifying areas in need of further research from both basic and applied literatures, and (c) discussing recent research findings on punishment within the context of advances in the functional analysis of behavior.[1]

Some authors have suggested that additional applied research on punishment is unnecessary in light of refinements to the functional analysis methodology and treatment with reinforcement (Donnellan & LaVigna, 1990; Guess, Helmstetter, Turnbull, & Knowlton, 1987). Results of numerous studies conducted over the past 15 years have shown that the function of problem behavior often can be determined and that this information can be used to develop treatments based on extinction, reinforcement, and other processes such as establishing operations (e.g., Iwata, Pace, Dorsey, et al., 1994). Nevertheless, punishment may be critical to treatment success when the variables maintaining problem behavior cannot be identified or controlled (for further discussion, see Axelrod, 1990; Iwata, Vollmer, & Zarcone, 1990; Vollmer & Iwata, 1993). Punishment also may be preferable to reinforcement-based treatments when problem behavior must be suppressed rapidly to prevent serious physical harm (Dura, 1991; see also Iwata et al.; Vollmer & Iwata). More important, results of several studies indicate that treatments derived from functional analyses (e.g., differential reinforcement of alternative behavior [DRA]) may not always reduce behavior to clinically acceptable levels without a punishment component (e.g., Grace, Kahng, & Fisher, 1994; Hagopian, Fisher, Sullivan, Acquisto, & LeBlanc, 1998; Wacker et al., 1990).

[1] Much of the applied research on functional analysis and treatment of behavior disorders has been conducted with individuals diagnosed with developmental disabilities. Thus, this review reflects this emphasis.

Knowledge about punishment also is important because common treatments that are associated with other processes may in fact reduce problem behavior through the mechanism of punishment. For example, procedures such as response blocking, guided compliance, and the application of protective equipment are often presumed to reduce problem behavior by terminating the reinforcement contingency that maintains the response (i.e., through extinction; e.g., Reid, Parsons, Phillips, & Green, 1993; Rincover, 1978). Research findings suggest that these procedural variations of extinction may function as punishment instead of, or in combination with, extinction (e.g., Lerman & Iwata, 1996b; Mazaleski, Iwata, Rodgers, Vollmer, & Zarcone, 1994). Some authors also have suggested that the contingent loss of reinforcement associated with differential reinforcement procedures (e.g., differential reinforcement of other behavior [DRO]) may constitute a form of punishment (e.g., Rolider & Van Houten, 1990). Thus, the process of punishment may underlie a number of popular function-based treatments.

Functional analysis methodology now permits more precise investigations of important environment–behavior relations in the area of punishment. Basic findings indicate that various parameters of reinforcement influence the direct and indirect effects of punishment and interact with nearly every other factor that has been found to influence responding during punishment (e.g., punishment schedule and magnitude). Such interactions have important clinical implications because punishment is more likely to be used when the response–reinforcer relation cannot be terminated completely. In applied studies, important reinforcement variables identified via functional analysis could be manipulated prior to and during punishment, even when the behavior is maintained by nonsocial consequences (i.e., the behavior is maintained in the absence of socially mediated reinforcers, such as attention, tangible items, and escape from instructions; see, e.g., Lerman & Iwata, 1996b).

Important issues related to the ethics and acceptability of using punishment to treat problem behavior in individuals with developmental disabilities have been the subject of numerous articles over the past 30 years. An overview of these issues is beyond the scope of this paper but can be found in a variety of sources (see Donnellan & LaVigna, 1990; Emerson, 1992; Guess et al., 1987; Jacob-Timm, 1996; Sidman, 1989; Van Houten et al., 1988). Suggestions for further applied studies on punishment are made throughout this paper, with the assumption that pertinent guidelines and cautions about the application of punishment will accompany published research findings (see Alberto & Troutman, 1999; Lovaas & Favell, 1987; Matson & DiLorenzo, 1984).

Ultimately, the consumers of behavioral technologies (e.g., clinicians, caregivers) will determine which treatments are used with individuals with developmental disabilities (see Iwata, 1988, for a cogent discussion of this issue). These decisions are at least partly guided by information generated by the scientific community. Safe, acceptable, and highly effective technologies of behavior change should be available to consumers who request them, including procedures that are based on punishment.

Basic and applied research findings on clinically relevant factors that influence the direct effects of punishment will be discussed in the first half of the paper. Other characteristics of punished responding, including maintenance, generalization, and side effects, will be discussed in the second half.

FACTORS THAT INFLUENCE THE DIRECT EFFECTS OF PUNISHMENT

Much of the basic research on the direct effects of punishment was conducted more than 30 years ago with nonhumans (see Azrin & Holz, 1966, for a review of this literature). Procedural variations of punishment examined in the laboratory have included the delivery of stimuli, often called *positive punishment,* and the removal of stimuli, often called *negative punishment* (i.e., response cost, time-out from positive reinforcement).[2] The majority of studies, however, employed contingent electric shock. In the earliest basic studies, the effects of punishment were evaluated while the punished response was undergoing extinction (e.g., Estes, 1944; Skinner, 1938; Thorndike, 1932). However, in most subsequent studies, punishment contingencies were introduced with no change in the prevailing reinforcement schedule. This latter arrangement assured a certain level of responding by which to examine the effects of punishment independently of those produced by extinction (Azrin & Holz, 1966). Complex interactions between reinforcement and punishment processes also could be evaluated. A number of authors suggested that this laboratory arrangement may be more pertinent to application because punishment is most likely to be used when the reinforcer that maintains problem behavior cannot be identified or controlled (Azrin & Holz, 1966; Dinsmoor, 1952).

In fact, the function of problem behavior was not determined prior to treatment in most clinical studies on punishment. Punishment thus was superimposed on an unknown schedule of reinforcement that likely took the form of extinction when the behavior was maintained by social consequenc-

[2] Consistent with previous articles and chapters on punishment, few distinctions will be drawn between positive and negative punishment in this review. Current research findings suggest that the procedures are associated with similar direct and indirect effects on responding. Nevertheless, the literature on positive punishment far exceeds that on negative punishment. Basic processes may differ in important ways under these two forms of punishment.

es (Iwata, Pace, Cowdery, & Miltenberger, 1994). That is, social consequences that may have maintained responding during baseline (e.g., verbal reprimands, escape from instructions) often were removed with the introduction of punishment. A substantial portion of applied findings thus may have little generality to contemporary treatment approaches because punishment is most likely to be used when problem behavior continues to produce reinforcement. Basic findings suggest that reinforcement parameters can influence the effects of punishment in important ways.

The Direct Effects of Punishment

Basic research findings have shown that response-contingent shock, noise, blasts of air, response cost, and time-out can produce a rapid decrease in the frequency of behavior and, in some cases, may lead to complete response suppression in rats, pigeons, monkeys, and humans (e.g., college students, psychiatric patients; Azrin, 1960; Crosbie, Williams, Lattal, Anderson, & Brown, 1997). Several studies with humans and nonhumans also found that the initial reductive effects of punishment with shock or point loss occurred more rapidly, or to a greater extent, than those produced by extinction, satiation, and differential reinforcement (e.g., Holz & Azrin, 1963; Johnson, McGlynn, & Topping, 1973; Rawson & Leitenberg, 1973).

The potential benefit of using punishment to treat intractable behavior problems led to the development of numerous punishment procedures for clinical use. Results of research have shown that treatment with a wide variety of punishers (e.g., verbal reprimands, restraint, water mist, lemon juice, shock, removal of reinforcing activities or conditioned reinforcers) can produce an immediate, substantial suppression in problem behavior (see Kazdin, 2001, and Matson & DiLorenzo, 1984, for reviews of this literature). Applied findings also indicate that the effects of punishment are superior to those obtained with less intrusive procedures alone, such as differential reinforcement (e.g., Barrett, Matson, Shapiro, & Ollendick, 1981; Favell et al., 1982; Scotti, Evans, Meyer, & Walker, 1991). Although results of such comparisons are consistent with those obtained in the laboratory, the findings are difficult to interpret because numerous parameters likely influence the effects of these behavior-reduction procedures. For example, a dense schedule of differential reinforcement may reduce behavior more effectively than a mild punisher.

More important, the relative efficacy of treatment with reinforcement versus punishment likely depends on a variety of factors (e.g., history; use of extinction; type, amount, and schedule of the consequence). These complex interactions need to be evaluated to generate more definitive findings about the suppressive effects of punishment relative to other procedures. Further research on strategies to improve the efficacy of punishment would be more pragmatic over the long run than additional, complex comparative studies of reinforcement versus punishment.

Several authors have suggested that treatment with punishment is so effective because punishment usually can compete successfully with reinforcement contingencies that maintain problem behavior (e.g., Van Houten, 1983). Although punishment often was confounded inadvertently with extinction in applied research, recent studies have demonstrated that common punishment procedures (e.g., time-out, brief manual restraint) can be effective in the absence of extinction (Fisher, Piazza, Bowman, Hagopian, & Langdon, 1994; Lerman, Iwata, Shore, & DeLeon, 1997; Thompson, Iwata, Conners, & Roscoe, 1999). Reductions in behavior were obtained even after unsuccessful attempts to treat the behavior with less intrusive procedures (e.g., Fisher et al., 1993; Lindberg, Iwata, & Kahng, 1999).

Nevertheless, the generality of these findings may be limited because data on effective treatments are more likely to be published than those that show unsuccessful outcomes. Potentially important reinforcement parameters also were unspecified in these studies. Results of pretreatment functional analyses indicated that the behavior was maintained independent of social consequences but did not isolate the precise reinforcer. Methods to identify the type of nonsocial reinforcement (often called *automatic reinforcement*) that is functionally related to problem behavior have been examined in a number of studies (e.g., Goh et al., 1995; Kennedy & Souza, 1995; Patel, Carr, Kim, Robles, & Eastridge, 2000; Piazza, Adelinis, Hanley, Goh, & Delia, 2000). Although further methodological refinements are needed, these strategies may be useful for identifying and manipulating various reinforcement parameters (e.g., reinforcer schedule and magnitude) while treating automatically reinforced problem behavior with punishment (e.g., Lerman & Iwata, 1996b). As described in more detail below, results of studies employing these types of manipulations with behavior maintained by either social or nonsocial consequences could lead to a greater understanding of punishment processes and improved treatments.

As discussed in the following sections, a number of factors directly relevant to the development of an applied technology have been found to influence the direct effects of punishment. These factors include historical variables (e.g., prior experience with the punishing stimulus or intermittent reinforcement); the use of conditioned punishers; reinforcement variables (e.g., schedule, availability of alternative sources of reinforcement); and punishment variables (e.g., magnitude, immediacy, schedule). However, much of the research on these factors has been conducted in the basic laboratory, and our knowledge of some important complex relations is relatively incomplete (Baron, 1991).

History

Basic findings indicate that previous exposure to certain factors can alter responding during punishment, a phenomenon that is especially relevant to the application of punishment because clinical populations typically have diverse learning histories. Results of numerous basic studies have shown that prior experience with the punishing stimulus either contingently or noncontingently can decrease a behavior's sensitivity to punishment (e.g., Capaldi, Sheffer, Viveiros, Davidson, & Campbell, 1985; Halevy, Feldon, & Weiner, 1987). For example, research findings with rats indicate that exposure to intermittent punishment with shock decreases the efficacy of continuous punishment with shock, even when several days or weeks lapse between intermittent and continuous punishment (Banks, 1967; Halevy et al.; Shemer & Feldon, 1984). Deur and Parke (1970) replicated this effect with normally developing children and a loud buzzer as the punishing stimulus.

Although adaptation to the punishing stimulus may account for these findings (Capaldi et al., 1985), a similar relation has been obtained with intermittent reinforcement. That is, rats and college students with a history of intermittent reinforcement also showed less response suppression under either continuous or intermittent punishment with shock than participants with a history of continuous reinforcement (e.g., Brown & Wagner, 1964; Estes, 1944; Halevy et al., 1987; Vogel-Sprott, 1967). Moreover, Eisenberger, Weier, Masterson, and Theis (1989) found that resistance to shock punishment increased for one response (lever pressing) in rats even after a topographically different behavior (running) was exposed to intermittent reinforcement.

In clinical settings, an individual is likely to experience common punishers (e.g., verbal reprimands, time-out) before these consequences are specifically arranged to treat a particular inappropriate behavior. Furthermore, exposure to intermittent schedules of reinforcement and punishment is typical in the natural environment. Consequences often are delivered on intermittent schedules because it is difficult for parents and teachers to reinforce or punish every occurrence of behavior. When initial attempts to treat a behavior with intermittent punishment fail, caregivers may switch to a continuous schedule of punishment in an attempt to improve the efficacy of the treatment. Basic findings indicate that a history with intermittent punishment may complicate treatment success in these cases, such that more intense punishers will be required to suppress behavior effectively (Halevy et al., 1987; Shemer & Feldon, 1984). However, the relevance of these findings to the types of punishers that are more commonly used in clinical settings is unknown because nearly all basic studies in this area evaluated the effects of shock, and no applied studies have replicated and extended these findings to problem behavior.

Nevertheless, basic findings in this area suggest some important guidelines for clinical research and practice. First, it may be beneficial for caregivers to identify novel punishers when designing treatments and to avoid using common consequences, such as verbal reprimands and time-out, in unsystematic or unplanned ways. Second, intermittent schedules of punishment should not be implemented prior to continuous schedules. Third, if adaptation to the punishing stimulus accounts for the decreased sensitivity of behavior (Capaldi et al., 1985), brief hiatus from punishment may be useful, as described in more detail below (Rachlin, 1966). Alternating among several effective punishment procedures in lieu of using a single procedure is another potentially useful strategy for minimizing exposure to any single punisher (e.g., Charlop, Burgio, Iwata, & Ivancic, 1988).

Conditioned Stimuli

Neutral stimuli that are paired with punishing stimuli eventually may acquire properties of the punishing stimuli. Results of basic studies indicate that these conditioned stimuli can function as punishers when delivered contingent on behavior in the absence of the primary, or unconditioned, stimulus (e.g., Hake & Azrin, 1965). Conditioned punishers may be useful for increasing both the efficacy and acceptability of punishment in clinical settings. Suppose that a relatively nonintrusive but ineffective consequence (e.g., a brief verbal cue) was established and maintained as a potent conditioned punisher via intermittent pairings with a more restrictive, time-consuming intervention (e.g., overcorrection, time-out). Application of the conditioned punisher would reduce the individual's exposure to the intrusive intervention and the degree of effort required by caregivers to implement treatment, factors that might circumvent problems with program inconsistency, habituation to the unconditioned punisher, and ethical issues associated with the use of restrictive procedures.

Various stimuli have been established as conditioned punishers in the basic laboratory, including tones, lights, and low-voltage shock (e.g., Crowell, 1974; Davidson, 1970; Hake & Azrin, 1965). With a few exceptions (e.g., Trenholme & Baron, 1975), the unconditioned stimulus was electric shock and the subjects were rats or pigeons. Conditioned punishers were established via one of two primary methods. Under one method, the neutral stimulus was presented before the onset of an inescapable stimulus (e.g., shock delivered independent of responding) and then was either removed with the onset of the unconditioned stimulus (e.g., Hake & Azrin, 1965; Mowrer & Solomon, 1954) or remained in the environment while the unconditioned stimulus was delivered periodically (e.g., Orme-Johnson & Yarczower, 1974). Results of several studies on this method indicated that more conditioning occurred if the neutral stimulus was presented prior to the onset of the unconditioned stimulus rather than simultaneously with or after its onset (e.g., Evans, 1962; Mowrer & Aiken, 1954).

Under the other method, the neutral stimulus was established as a discriminative stimulus for punishment. That is, the presence of the neutral stimulus was correlated with delivery of the unconditioned stimulus contingent on responding (e.g., Davidson, 1970). Although the discriminative stimulus then was shown to suppress responding when delivered contingent on behavior, results of Orme-Johnson and Yarczower (1974) indicated that stimuli established as discriminative stimuli were much less effective as conditioned punishers than stimuli established via the former method. Regardless of the conditioning method, research findings have shown that the effects of conditioned punishers on behavior are temporary unless the conditioned stimulus and the unconditioned punisher continue to be paired in some manner (Davidson, 1970; Hake & Azrin, 1965).

A few basic studies have evaluated factors that appear to influence the conditioning process, such as the magnitude of the unconditioned stimulus (Mowrer & Solomon, 1954) and the duration of the conditioned stimulus (Hake & Azrin, 1965). However, other clinically relevant parameters of conditioning, such as the number of pairings between the conditioned and unconditioned stimuli, the type of unconditioned punisher used, and characteristics of the neutral stimulus (e.g., intensity or saliency), should be examined in further research. In addition, conditioned stimuli typically were established and maintained independent of responding, a method that likely would invoke ethical concerns if extended to clinical populations. Although it may be more acceptable to pair conditioned and unconditioned stimuli contingent on problem behavior, opportunities to condition the stimulus would be severely restricted if the unconditioned punisher suppressed problem behavior to low levels.

The use of conditioned punishers in treating problem behavior has been reported in surprisingly few applied studies. More important, no applied studies have focused exclusively on methods to develop and maintain stimuli as conditioned punishers in clinical settings. Lovaas and Simmons (1969) paired the word "no" with shock contingent on severe self-injury with 1 participant. The brief verbal reprimand then was presented for self-injury in the absence of shock during a limited number of sessions, and results suggested that the stimulus had acquired the suppressive properties of the original punisher. In a more thorough evaluation, Dorsey, Iwata, Ong, and McSween (1980) paired the word "no" with contingent water mist for 2 participants who engaged in self-injury. Results showed that contingent presentation of the verbal stimulus maintained low levels of self-injury when water mist was withdrawn from the original treatment setting. Furthermore, the suppressive effects of the verbal reprimand generalized to a setting that had not been previously associated with the water mist procedure, as well as to other therapists who had never delivered the water mist. Finally, Dixon, Helsel, Rojahn, Cipollone, and Lubetsky (1989) paired a mild, less effective punisher (visual screen) with a more effective punisher (the odor of ammonia) while treating aggression and disruption exhibited by a young boy with developmental disabilities. Problem behavior remained suppressed for a short time when the visual screen was used alone.

Although results of these three studies indicated that conditioned punishers were established successfully for a clinical problem, the efficacy of treatment was evaluated across a limited number of brief sessions. Basic findings have shown that the effects of conditioned punishers on behavior are temporary unless the conditioned stimulus and the unconditioned punisher continue to be paired in some manner (Davidson, 1970; Hake & Azrin, 1965). In addition, details necessary to replicate the conditioning procedure (e.g., method of pairing, total number of pairings, rules for determining when to test the conditioned effect) were not delineated. The generality of these findings and those obtained in the basic laboratory also may be limited by the use of relatively intrusive unconditioned punishers (i.e., shock, water mist, ammonia).

Thus, current knowledge about conditioned punishment is fairly incomplete, and prescriptions for the application of conditioned punishers should await further research. The efficacy of pairing various types of auditory, tactile, and visual stimuli should be evaluated with more common and socially acceptable forms of punishment (e.g., time-out). The number of pairings necessary to produce conditioning and factors that might alter the outcome of conditioning (e.g., intensity of the conditioned stimulus) could be evaluated by periodically testing the suppressive effects of the paired stimulus in the absence of the unconditioned punisher. The durability of conditioning could be determined by presenting the conditioned stimulus without the unconditioned punisher until the effects on responding dissipate. This strategy also may be useful when developing a schedule for pairing the conditioned and unconditioned stimuli to maintain conditioning over time. For example, clinicians could determine the maximum number of times that the conditioned stimulus could be presented before the conditioning effect begins to be extinguished. The conditioned and unconditioned stimulus then could be paired prior to that number on a regular basis.

Reinforcement Schedule

Basic findings indicate that the characteristics of responding during punishment may depend on the reinforcement schedule that maintains the behavior, a factor that is especially relevant to application because problem behavior is likely to be maintained by some form of reinforcement during treatment. Behavior may be concurrently exposed to schedules of reinforcement and punishment when caregivers do not completely withhold social consequences during treatment or when the behavior is maintained by automatic reinforcement. Results of basic studies generally showed that the amount of response suppression under punishment was negatively related to the density of the reinforcement schedule, with extinction producing the greatest decrease in responding (Azrin & Holz, 1966). Various parameters of punishment (e.g., schedule, intensity) also appear to interact with the relation between response suppression and reinforcement density (Bouzas, 1978; Bradshaw, Szabadi, & Bevan, 1977, 1978). For example, Bradshaw and his colleagues found that the negative relation between reinforcement density and responding was much more pronounced when human subjects were exposed to a variable-ratio (VR) punishment schedule of monetary loss than to a variable-interval (VI) punishment schedule.

Intermittent reinforcement schedules examined in the laboratory have included fixed-interval (FI), fixed-ratio (FR), VI, and VR schedules. Although these reinforcement schedules have been found to interact differentially with the effects of punishment, this interaction has not been well studied and likely depends on various factors, such as the reinforcement density, punishment schedule, and amount of reinforcement lost due to a reduction in responding (e.g., Powell, 1970; Scobie & Kaufman, 1969; see also Baron, 1991, for further discussion). The various ways in which these schedules influence punishment effects are relevant to an applied technology because social contingencies for problem behavior often approximate these laboratory arrangements in the natural environment (e.g., Lalli & Goh, 1993; Vollmer, Borrero, Wright, Van Camp, & Lalli, 2001). Such complex interactions between reinforcement and punishment schedules also are likely responsible for some inconsistent findings reported in both the basic and applied literatures on punishment (see further discussion below). Additional basic research in this area is needed to clarify these relations.

Knowledge about basic processes and prescriptions for best practices when using punishment in clinical settings will be incomplete without further evaluation of potential interactions between reinforcement schedules and parameters of punishment. Nevertheless, no applied studies have examined the effects of reinforcement schedule or density on the outcome of treatment with punishment. Further research should determine if reducing the density of the reinforcement schedule operating in the natural environment would substantially enhance the efficacy of commonly used punishment procedures. If so, strategies are needed to thin the schedule of reinforcement for problem behavior during treatment with punishment. The parameters under which reinforcement schedule is and is not an important factor when treating problem behavior with punishment also should be evaluated. Potentially relevant parameters include the schedule, type, and intensity of the punisher.

Although nonsocial sources of reinforcement may be difficult to modify (but see Lerman & Iwata, 1996b, for one approach), results of this research may lead to useful guidelines for designing effective yet practical treatments when caregivers will be unlikely or unable to withhold social reinforcement for problem behavior. Current knowledge indicates that reinforcement for problem behavior should be withheld or diminished if possible. Thus, for example, when caregivers cannot withhold reinforcement completely during punishment (e.g., ignore all instances of self-injury), the density and magnitude of reinforcement for problem behavior should be reduced and punishment should be delivered on a continuous schedule (see further discussion below).

Availability of Alternative Reinforcement

Most textbooks and literature reviews on application highlight the benefits of combining punishment with some type of reinforcement procedure, such as DRA (e.g., Cooper, Heron, & Heward, 1987; Matson & DiLorenzo, 1984). Basic studies with rats, pigeons, and psychiatric patients have shown that the suppressive effects of contingent shock, noise, or time-out were enhanced when reinforcement could be obtained in some manner other than, or in addition to, engaging in the punished response (e.g., Boe, 1964; Herman & Azrin, 1964; Holz, Azrin, & Ayllon, 1963; Rawson & Leitenberg, 1973). Although these findings suggest that increasing the density of alternative reinforcement might enhance the efficacy of mild punishers in clinical settings, few basic studies have evaluated clinically relevant punishers or the parameters under which reinforcement may provide optimal benefits during punishment.

In a study with pigeons reported by Azrin and Holz (1966), for example, an FR 25 schedule of alternative reinforcement did not increase sensitivity to punishment when a response maintained by the same reinforcement schedule was punished with low-voltage shock (less than 50 V). It is possible, however, that a denser schedule of alternative reinforcement would have enhanced the efficacy of this mild punisher. Results of a study by Fantino (1973) indicated that the beneficial effects of alternative reinforcement were compromised when the reinforcement rate provided by a concurrent VI schedule prior to punishment could not be obtained via exclusive responding on the unpunished alternative. Punishment parameters such as schedule and delay and various reinforcement parameters also likely modify the effects of alternative reinforcement. Thus, basic findings suggest that various factors (e.g., type or intensity of the punisher, density of available reinforcement prior to punishment) must be considered when combining reinforcement with punishment in clinical settings.

However, other commonly used reinforcement arrangements, such as noncontingent reinforcement (NCR) and differential reinforcement of low response rates (DRL), have not been evaluated in the context of a concurrent punishment contingency in the laboratory. More important, basic findings on alternative reinforcement may have limited generality to application because clinically relevant factors (e.g., response topography; reinforcement quality, schedule, delay, and magnitude) usually were held constant across available response options. Treatment with punishment and differential reinforcement typically will incorporate different response, reinforcement, and punishment parameters across the targeted behaviors, especially when problem behavior is maintained by unknown or uncontrolled sources of reinforcement. In such cases, alternative reinforcement may neither suppress the punished response nor increase adaptive behavior. Results of basic studies in which schedules of reinforcement and punishment were arranged for both response options also suggest that the amount of suppression produced by punishment for a given behavior can be influenced by contingencies that operate on other behavior, including the relative schedule, delay, and magnitude of reinforcement and punishment (e.g., Deluty, 1976, 1978; Farley, 1980).

Despite obvious clinical implications, only one applied study has evaluated the relation between punishment effects and the availability of alternative reinforcement. Thompson et al. (1999) examined the separate and combined effects of punishment and reinforcement on self-injury after results of a functional analysis indicated that the behavior was maintained by automatic reinforcement. Reinforcement was arranged for an alternative behavior (toy manipulation) by giving the participants access to preferred toys (thereby establishing automatically reinforced toy play) or by delivering food contingent on toy manipulation. Results for the 4 participants indicated that alternative reinforcement enhanced the efficacy of relatively mild punishers (e.g., brief manual restraint). Furthermore, reinforcement alone was fairly ineffective for all participants, and punishment alone was ineffective for 1 participant. As noted by the authors, however, the combined treatment was differentially confounded with an additional contingency (i.e., time-out from positive reinforcement—access to programmed reinforcement was withheld during punishment delivery) that may constrain the generality of the findings.

Further research is needed on clinical strategies to enhance the efficacy of mild punishers through the use of DRA, DRO, DRL, and NCR procedures. Thus far, basic findings in this area suggest a number of tentative prescriptions for application. Naturalistic reinforcement schedules for targeted problem and alternative behavior should be considered first when developing treatments that combine reinforcement and punishment. The type, schedule, and magnitude of reinforcement maintaining problem behavior should be identified, so that a larger amount of the same reinforcer could be provided independent of undesirable responding or contingent on alternative behavior (Fantino, 1973). When the functional reinforcer cannot be identified or delivered by others, selecting reinforcers that compete with or substitute for maintaining reinforcers may be critical to effective treatment (e.g., Shore, Iwata, DeLeon, Kahng, & Smith, 1997).

As many of the available sources of reinforcement as possible also should be determined prior to treatment, so that steps can be taken to ensure that the total amount of obtainable reinforcement can be sustained or exceeded despite a reduction in the punished behavior. To this end, differential reinforcement procedures should target simple free-operant responses or adaptive behavior that is already in the individual's repertoire. Procedures such as NCR, DRO, and DRL, which do not require an alternative response for reinforcement delivery, may be preferable during the initial stages of treatment with punishment to insure a sufficient density of available reinforcement. In fact, multiple responses, reinforcers, and reinforcement contingencies (e.g., DRO plus NCR) should be incorporated into treatment such that punishment for a restricted number of responses is implemented within the context of a richly reinforcing environment.

Punishment Magnitude

The relation between the effects of punishment and the magnitude or amount of punishment delivered for responding is especially germane to the efficacy and acceptability of punishment in clinical settings. From an ethical and practical standpoint, the least amount of punishment that is effective (i.e., lowest intensity, shortest duration) should be used to treat behavior problems, and strategies that increase the effectiveness of mild punishers should be incorporated into treatment. Basic studies on magnitude have shown that response suppression is positively related to the intensity and duration of shock with rats, pigeons, monkeys, and college students (Church, 1969; Deluty, 1978; Scobie & Kaufman, 1969), the duration of time-out with normal humans (Kaufman & Baron, 1968; N. B. Miller & Zimmerman, 1966), and the number of points lost as part of response cost with normal humans (Weiner, 1964). In fact, recovery during punishment and following the termination of the punishment contingency was most likely to occur with mild punishers, such as a bar slap (Skinner, 1938) and low-voltage shock (Hake, Azrin, & Oxford, 1967). Furthermore, punishment with high-intensity shock or forceful air blasts was found to be ineffective with rats if the intensity of the punisher was initially low and then gradually increased over time (Cohen, 1968; N. E. Miller, 1960; Terris & Barnes, 1969).

On the basis of these findings, numerous authors have recommended using moderate or high-intensity punishers to treat problem behavior and cautioned against increasing the intensity of punishment gradually over time (e.g., Cooper et al., 1987; Martin & Pear, 1996; O'Brien, 1989). These guidelines may be difficult to reconcile with ethical mandates to identify the least restrictive procedure that is effective. Moreover, close examination of basic findings in this area indicates that the relation between responding and punishment magnitude is more complex than frequently assumed. As a result, strategies that are based on recommendations delineated in applied textbooks (e.g., Cooper et al.; Martin & Pear) and literature reviews may not influence behavior as expected.

As previously noted, the basic relation between response suppression and punishment magnitude may be influenced by the availability of alternative reinforcement (e.g., Holz et al., 1963) and the type of reinforcement schedule that maintains behavior (e.g., Powell, 1970; Scobie & Kaufman, 1969). Other variables (e.g., immediacy; Cohen, 1968) also may alter the relation between punishment magnitude and responding. Such complex interactions may be responsible for some contradictory findings on punishment magnitude that have been reported in the applied literature. The extent to which basic findings are directly comparable to applied findings also is limited because most basic studies examined the magnitude of shock, whereas applied studies in this area have focused on more clinically acceptable forms of punishment (e.g., overcorrection, time-out).

In one of the few applied studies to evaluate the magnitude of electric shock, D. E. Williams, Kirkpatrick-Sanchez, and Iwata (1993) compared the efficacy of treatment for self-injury under two shock-intensity levels (3.5 mA vs. 18.5 mA). Results were consistent with those obtained in basic studies. The high-intensity shock produced larger, more immediate decreases in behavior than the low-intensity shock. The generality of this finding, however, is somewhat limited because punishment was combined with extinction. In addition, sequence effects could have influenced the outcome because the participant was exposed to the lower intensity shock prior to the higher intensity, and a reversal to the low-intensity condition was not implemented.

Results of studies examining the relation between punishment magnitude and treatment efficacy using other types of punishers (e.g., physical restraint, unpleasant smells, time-out) have been inconsistent and often appeared to be confounded with other variables (e.g., Altman, Haavik, & Cook, 1978; Cole, Montgomery, Wilson, & Milan, 2000; Marholin & Townsend, 1978; Singh, Dawson, & Manning, 1981). For example, Cole et al. found that treatment with overcorrection produced similar decreases in stereotypic behavior, regardless of whether the intervention lasted 30 s, 2 min, or 8 min. The effects of overcorrection, however, may have been confounded with those of extinction and verbal reprimands. Results of studies on the duration of time-out have shown a positive relation (e.g., Burchard & Barrera, 1972; Hobbs, Forehand, & Murray, 1978), a negative relation (e.g., Kendall, Nay, & Jeffers, 1975), and no relation (e.g., White, Nielsen, & Johnson, 1972) between duration length and treatment effects. These findings are difficult to interpret because the function of problem behavior was not identified (thus, time-out may have been contraindicated for some participants), and sequence effects may have confounded the results (Matson & DiLorenzo, 1984).

Further research on the relation between punishment magnitude and response to treatment, as well as on factors that can alter this relation (e.g., reinforcement schedule, punishment delay), may be useful for reconciling inconsistent findings in the literature and for developing more comprehensive prescriptions for application. The common assumption that a larger magnitude of a given punisher will be more effective than a smaller magnitude is not strongly supported in the current literature, with the exception of findings on contingent shock. Magnitude should be manipulated with a variety of punishers and in a variety of ways that have not been examined in basic research. For example, the amount of reinforcement available during "time-in" is another potentially important dimension of magnitude when treatment with time-out is implemented (e.g., Solnick, Rincover, & Peterson, 1977).

Strategies to enhance response suppression and maintenance under less effective values of punishment magnitude also should be explored. Basic findings indicate that smaller magnitudes of punishment may be more effective if the punisher is delivered immediately following the behavior (e.g., Cohen, 1968) and if reinforcement is available for an alternative response (e.g., Holz et al., 1963). Basic studies with shock also indicate that a less intense punisher may be effective, at least temporarily, if a high intensity level is decreased gradually over time (e.g., Azrin, 1960; Cohen, 1968; Hake et al., 1967). An approach that involves periodically interspersing less intense punishers with more intense punishers may be useful for maintaining treatment effects while the magnitude of punishment is gradually reduced. As discussed above, further research on conditioned punishers and treatments combining reinforcement with punishment also may lead to methods for increasing the effectiveness of mild punishers.

Finally, further studies should evaluate methods for identifying the most appropriate magnitude of a given punishment procedure prior to treatment implementation in the natural environment. The typical trial-and-error approach to punishment selection is inefficient and may be counterproductive if an individual receives prolonged exposure to ineffective procedures (e.g., N. E. Miller, 1960; Terris & Barnes, 1969). Efficient strategies for identifying the least restrictive, effective treatment are surprisingly absent from the applied literature. In two studies conducted by Fisher and colleagues (Fisher, Piazza, Bowman, Hagopian, & Langdon, 1994; Fisher, Piazza, Bowman, Kurtz, et al., 1994), the potential suppressive effects of various procedures (e.g., time-out, facial screen, contingent demands) were rapidly assessed by exposing participants to the putative punishers while negative vocalizations (e.g., yelling, crying) and avoidance or escape responses (e.g., dropping to the floor) were measured. Each punishment procedure was delivered noncontingently across five different durations, ranging from 15 s to 180 s. Results of subsequent treatment analyses indicated that the procedure associated with the highest levels of negative vocalizations, avoidance, or escape responses was the most effective punisher for problem behavior. However, results of the initial assessment did not differentiate among the various duration lengths for any participant, possibly because the procedures were alternated rapidly across a limited number of trials. The utility of such assessments should be evaluated in further studies. For example, a separate assessment of punishment magnitude, similar in design to that conducted by Fisher et al., might be useful after an initial assessment has identified a potent punisher.

Until further applied research on magnitude is conducted, practitioners should select magnitudes that have been shown to be safe and effective in clinical studies, as long as the magnitude is considered acceptable and practical by those who will be implementing treatment. Punishment should be combined with some type of reinforcement procedure, and the punisher should be delivered as immediately as possible following occurrences of problem behavior (see further discussion below). If the punisher fails to suppress behavior over time, alternative procedures probably should be considered instead of increasing the magnitude of the punisher under the presumption that this strategy will improve the efficacy of treatment.

Immediacy of the Punisher

Consequences for problem behavior are frequently delayed in the natural environment. Caregivers and teachers often are unable to monitor behavior closely or to deliver lengthy punishers (e.g., 15-min contingent work) immediately following instances of problem behavior (Azrin & Powers, 1975). Punishment also may be delayed when the individual actively resists application of the programmed consequences by struggling with the punishing agent or running away.

In some cases, problem behavior occurs primarily in the absence of the punishing agent, necessarily delaying programmed consequences until the behavior is detected (Grace, Thompson, & Fisher, 1996; Van Houten & Rolider, 1988).

For these reasons, research on delayed punishment is especially pertinent. Laboratory findings with rats indicate that the lengthier the delay between the occurrence of the response and delivery of contingent shock, the smaller the amount of response suppression under punishment (e.g., Baron, Kaufman, & Fazzini, 1969; Camp, Raymond, & Church, 1967). Even brief delays of 10 s or 20 s have been found to seriously compromise the effects of contingent shock with rats and college students (e.g., Banks & Vogel-Sprott, 1965; Goodall, 1984) and of reinforcement loss with college students (Trenholme & Baron, 1975).

Stimuli that might bridge the interval between a response and its consequence have been notably absent from laboratory arrangements involving delayed punishment. Results of at least one study suggest that factors such as the presence of a conditioned punisher and delivery of instructions can alter the efficacy of delayed punishment. In Trenholme and Baron (1975), delays of 10 s, 20 s, and 40 s were equally effective with college students when a brief noise that was paired with reinforcement loss also occurred immediately following the behavior. A subsequent experiment showed that delayed punishment was just as effective as immediate punishment when the participants received instructions about the delay. The generality of these findings to clinical populations, such as individuals with developmental disabilities, has not been determined. In addition, no basic studies have evaluated the effects of numerous other potentially important factors on delayed punishment (e.g., history, reinforcement schedule, availability of alternative reinforcement).

Surprisingly few applied studies have evaluated the efficacy of delayed punishment or strategies to improve treatment effects when consequences do not occur contiguous to the behavior. In one of the few studies to compare immediate and delayed punishment, Abramowitz and O'Leary (1990) found that immediate verbal reprimands were much more effective in decreasing off-task behavior in school children than were reprimands that were delayed by 2 min. These results are somewhat difficult to interpret, however, because delayed reprimands were delivered only if off-task behavior had occurred continuously for 2 min, during which time the students had varied opportunities to interact with other students and non-task-related objects. Thus, the effects of punishment delay were not separated from those of reinforcement schedule and punishment schedule.

Results of just two studies have delineated conditions under which delayed punishment may produce effective outcomes. Rolider and Van Houten (1985) and Van Houten and Rolider (1988) demonstrated the efficacy of delayed punishment using various mediated consequences with children with emotional and developmental disabilities. One form of mediation involved playing audiotape recordings of the child's disruptive behavior that were collected earlier in the day. The punishing consequence (physical restraint, verbal reprimands) then was delivered. In some cases, the tape recorder was clearly visible to the child while the recordings were being collected, and a verbal explanation of its role in the delivery of delayed punishment was provided. These factors may have served to bridge the temporal gap between inappropriate behavior and its consequence (e.g., by functioning as discriminative stimuli for punishment; Trenholme & Baron, 1975). However, for 1 participant, neither instructions about delayed punishment nor an immediate consequence designed to signal that punishment was forthcoming (a mark placed on the child's hand) was as effective as the audiotape procedure.

In Van Houten and Rolider (1988), caregivers physically guided 2 participants to engage in the problem behavior (aggression or theft) after occurrences of the behavior were detected or reported. Caregivers then delivered the punishing consequence (physical restraint) immediately following the guided response. Although treatment was effective, the efficacy of delayed punishment without the guided response component was not examined. The length of the delay and caregivers' immediate response to the problem behavior also were not specified.

Current knowledge indicates that the mild punishers typically used in clinical settings will be ineffective unless the consequence immediately follows problem behavior. Thus, further research is needed on factors that might enhance treatment effects under delayed punishment, especially procedures or stimuli that would bridge the temporal gap between a response and its consequence. Research on the utility of delivering conditioned punishers, instructions, and other types of stimuli associated with delayed consequences is needed with clinical populations.

Until further applied research is conducted, teachers and caregivers should be concerned with selecting punishers that can be readily delivered as soon as the behavior occurs. Consequences that do not require the close proximity of the caregiver (i.e., stimuli that can be delivered or removed from a distance) and technology to increase the practicality of immediate punishment may be especially useful in this regard. Electronic devices that detect occurrences of problem behavior and either alert caregivers or deliver consequences automatically might circumvent the problems of delayed punishment (e.g., Linscheid, Iwata, Ricketts, Williams, & Griffin, 1990).

However, the timing of punishment in relation to reinforcement delivery also should be considered, because some basic studies have found that immediate punishment was less effective than delayed punishment if the immediate punisher preceded reinforcement delivery but the delayed punisher followed it (e.g., Epstein, 1984; Rodriguez & Logan, 1980). It is conceivable that diligent caregivers may respond to problem behavior by first delivering the prescribed punisher (e.g., contingent work, time-out), followed (inadvertently) by the maintaining social reinforcer (e.g., access to materials). In a similar manner, automated punishment may be delivered immediately prior to social or nonsocial consequences for problem behavior. Results of other basic studies, in which the availability of reinforcement for one response was perfectly correlated with the delivery of mild punishment for an immediately preceding response, indicated that pairing punishment and reinforcement in this manner established the mild punisher as a conditioned positive reinforcer (e.g., Murray & Nevin, 1967; D. R. Williams & Barry, 1966). Thus, the timing of punishment and reinforcement in the natural environment should be evaluated carefully as part of treatment.

Schedule of Punishment

The effects of punishment schedules on responding have important implications for the efficacy and acceptability of treatment with punishment. Intermittently delivered consequences that successfully reduce problem behavior are easier to use, less time consuming, and less intrusive than consequences that must follow each occurrence of behavior. Results of basic research with pigeons and rats suggest that punishment with shock or time-out will not produce acceptable results unless the punisher follows nearly every occurrence of the behavior in situations in which no alternative is available or when the density of reinforcement is not reduced (Appel, 1968; Azrin, Holz, & Hake, 1963; Deluty, 1976; Farley, 1980; Thomas, 1968).

Although a number of applied studies have examined the efficacy of intermittent punishment for treating problem behavior, results have been inconsistent, and the conditions under which intermittent punishment might be effective remain unclear. In some studies, for example, intermittent punishment schedules were associated with a socially significant reduction in behavior, particularly if responding was already suppressed to low levels via continuous punishment (e.g., Clark, Rowbury, Baer, & Baer, 1973; Rollings & Baumeister, 1981; Romanczyk, 1977). Conversely, even dense intermittent punishment schedules were ineffective for some individuals in other studies (e.g., Calhoun & Matherne, 1975; Lerman et al., 1997). Basic findings on factors that interact with the effects of punishment schedules (e.g., reinforcement schedule) may explain why clinical applications have produced inconsistent results.

First, punishment appeared to be confounded with extinction and other potential punishers (e.g., verbal reprimands) in applied studies that showed significant treatment effects under thin punishment schedules (e.g., Barton, Brulle, & Repp, 1987; Clark et al., 1973; Romanczyk, 1977). Second, important parameters of punishment (i.e., type, intensity, and schedule) varied considerably among these studies. Basic findings indicate that these variables alter the relation between intermittent punishment and response suppression. Increasing the intensity of a punisher, for example, can either enhance or degrade the efficacy of intermittent schedules, depending on other factors (e.g., Appel, 1968; Lande, 1981). Certain punishment schedules (e.g., VI) also have been associated with greater decreases in responding than other schedules (e.g., FR or FI; Azrin, 1956; Camp, Raymond, & Church, 1966), although the nature of this relation is complex (e.g., Arbuckle & Lattal, 1992) and appears to be influenced by the schedule of reinforcement that maintains the behavior (e.g., Bradshaw et al., 1977, 1978; Powell, 1970; Scobie & Kaufman, 1969).

In the only applied study that examined the interaction between type or intensity of punishment and intermittent punishment schedules, Cipani, Brendlinger, McDowell, and Usher (1991) found that a VR 4 schedule of punishment with contingent application of lemon juice was just as effective as a continuous schedule in reducing a child's stereotypic behavior. A "manual guidance overcorrection" procedure (i.e., physically guiding the child's arms over the head and to the sides 10 times) also was effective when the procedure was delivered under a continuous schedule. Unlike the lemon juice, however, a VR 4 schedule with the overcorrection procedure did not produce clinically significant reductions in behavior.

Finally, interactions between schedules of reinforcement and punishment may account for the idiosyncratic effects of intermittent punishment on problem behavior. For example, Lerman et al. (1997) treated 5 participants' self-injurious behavior (SIB) with a continuous schedule of punishment after results of a functional analysis indicated that the behavior was maintained by automatic reinforcement. Initial application of intermittent punishment (FI 2 min or FI 5 min) was ineffective for 4 of the 5 participants. The continuous punishment schedule then was successfully thinned to FI 5 min for 2 of these participants. In contrast, continuous punishment was necessary to suppress SIB for the other 2 participants, despite repeated attempts to thin the schedule. Although the function of SIB had been identified prior to treatment, important parameters of the maintaining reinforcers were unknown (e.g., schedule, density, magnitude). These parameters, which likely varied across participants, may have been responsible for the inconsistent success of the schedule-thinning procedure.

Further research on interactions between punishment schedule (e.g., VR vs. VI) and other potentially important parameters of punishment and reinforcement is needed to clarify the conditions under which intermittent punishment would and would not be effective. Few studies have directly evaluated strategies to systematically thin punishment schedules or to utilize highly variable (and thus unpredictable) schedules. Combining a thin schedule of punishment with a rich schedule of conditioned punishment is another potential approach for increasing the efficacy of intermittent punishment. Depending on the nature of the conditioned and unconditioned punishers, this arrangement may be more practical than using a rich schedule of unconditioned punishment alone. A dense schedule of alternative reinforcement also may promote the efficacy of intermittent punishment.

Other types of punishment schedules examined in the basic laboratory, such as the differential punishment of high (DPH) or low (DPL) response rates, also may be useful in clinical settings. These schedules do not specify a direct contingency between the delivery of the punisher and the occurrence of a response. For example, under DPH or DPL, punishment is delivered contingent on the pause length that immediately preceded a response (i.e., the selective punishment of certain lengths of interresponse times). Results of basic studies on DPH and DPL schedules showed that overall responding increased when relatively long interresponse times (DPL) were punished and decreased when short interresponse times (DPH) were punished (e.g., Galbicka & Branch, 1981; Laurence, Hineline, & Bersh, 1994). DPH schedules may be more beneficial than continuous punishment when treating behavior that is considered problematic only because it occurs at high rates or in bursts. Further understanding of these schedules also is important because they may commonly operate in the natural environment. For example, caregivers may be more likely to deliver punishment when problem behavior occurs infrequently (i.e., is characterized by long interresponse times) than when behavior occurs at high rates (see Arbuckle & Lattal, 1992, for a discussion of this issue). Such an arrangement could compromise the efficacy of treatment by increasing the frequency of short interresponse times.

Current knowledge about punishment schedules suggests that parents and teachers should punish each occurrence of problem behavior unless the behavior is simultaneously exposed to extinction. Until further research is conducted, clinicians should be extremely cautious when attempting to thin the punishment schedule, utilize DRH schedules, or evaluate other strategies for improving the effects of intermittent punishment (e.g., employing variable schedules or conditioned punishers). A continuous schedule of punishment always should be implemented initially, and intermittent schedules should be considered only if the continuous schedule remains effective in suppressing problem behavior to low levels over a considerable amount of time.

FACTORS RELATED TO MAINTENANCE, GENERALIZATION, AND INDIRECT EFFECTS

A much smaller proportion of basic and applied studies on punishment have evaluated the long-term maintenance, generalization, and side effects of punishment relative to those on direct effects. The extent to which punishment effects are maintained over time, transfer across settings and contexts, and produce changes in other behavior has significant implications for treatments involving punishment.

Maintenance

The durability of treatment with punishment is one of the most important considerations for practitioners, teachers, and caregivers of individuals with behavior disorders. A number of authors, however, have suggested that the clinical effects of punishment are relatively short-lived, even when the treatment remains unchanged over time (e.g., Parsons, Hinson, & Sardo-Brown, 2001; Walker & Shea, 1999). In basic studies with both humans and nonhumans, various punishers have been associated with continued response suppression under punishment, including relatively intense levels of electric shock with pigeons and rats (Azrin, 1960; Crosbie et al., 1997), point or monetary loss with normal humans (Crosbie et al.; Weiner, 1962), and time-out from positive reinforcement with squirrel monkeys (McMillan, 1967). Response recovery, however, has been associated with less intense punishers, such as low-voltage shock with pigeons (Rachlin, 1966), a bar slap with rats (Skinner, 1938), and noise with pigeons (Holz & Azrin, 1962). Basic findings on the maintenance of response suppression following the termination of the punishment contingency also generally showed that response rates immediately returned to prepunishment levels—sometimes even temporarily exceeding baseline—unless intense punishers were used (e.g., high-voltage shock; Azrin, 1960).

These results suggest that sufficiently intense punishers, including some commonly used clinical procedures (e.g., time-out), may produce lasting reductions in problem behavior as long as the punishment contingency remains in effect. Nevertheless, basic findings may not be applicable to treatment outcomes in clinical settings because the time periods evaluated in the laboratory (e.g., 30-min to 60-min punishment sessions across 10 to 20 days) may have little relation to the numerous months (and sometimes years) over which problem behavior requires treatment. Moreover, few studies have examined factors that may influence the durability of punishment effects.

Although brief treatment evaluations are predominant in the applied literature on punishment, an increasing number of studies have examined the long-term efficacy of punishment over the past 10 years. Treatment effects have been examined for 1 to 60 months after punishment was initiated and continued with minor changes to the procedure (Duker & Seys, 1996; Ricketts, Goza, & Matese, 1993; D. E. Williams, Kirkpatrick-Sanchez, & Crocker, 1994), and after the original punishment component was withdrawn (Arntzen & Werner, 1999; Foxx, Bittle, & Faw, 1989; Rolider, Williams, Cummings, & Van Houten, 1991). Results have shown varying success in maintaining the reduction in behavior, yet potential reasons for the inconsistent outcomes have not yet been identified.

For example, D. E. Williams et al. (1993) observed a relapse in treatment with contingent electric shock 6 months after punishment was initiated. Conversely, Linscheid, Hartel, and Cooley (1993) found that contingent electric shock continued to suppress 2 individuals' self-injurious behavior for 5 years. Duker and Seys (1996) examined the long-term efficacy of contingent shock with 12 individuals by obtaining information on the degree of physical restraint each required from 2 to 47 months after the initiation of punishment. Results at follow-up suggested that treatment remained effective for 7 participants, including 1 individual who was evaluated at 36 months and another who was evaluated at 47 months.

Conclusions about applied findings on maintenance are difficult to draw for a number of reasons. First, the majority of studies examined the long-term effectiveness of contingent electric shock, so results may not be applicable to other (or more mild) punishers (Azrin & Holz, 1966). Second, the reinforcing consequences of problem behavior were not identified prior to treatment in most cases. Long-term maintenance may have been more likely to occur if the maintaining reinforcer was withheld contingent on problem behavior or readily available for engaging in more appropriate behavior (Estes, 1944). Third, other factors potentially responsible for both successful and unsuccessful cases of treatment maintenance may have varied widely across the studies (e.g., punishment schedule, availability of reinforcers that competed with or substituted for the maintaining reinforcer). In fact, components of the original intervention were modified over time in some studies (e.g., additional behavioral procedures or drugs were introduced; Duker & Seys, 1996), and it is difficult to determine which, if any, treatment modifications may have been responsible for the outcomes. Furthermore, the lengthy time period required to conduct these studies increased the likelihood that unplanned changes or other uncontrolled factors interacted with the efficacy of the original treatment in either desirable or undesirable ways. Finally, the number of treatment relapse cases reported in the literature may not accurately reflect the prevalence of this problem in applied settings because such cases are less likely to be submitted or accepted for publication than successful cases of treatment maintenance.

Identifying factors or processes associated with long-term maintenance is key to the design of a systematic technology for preventing and remediating treatment relapse. Several authors have suggested that adaptation, or habituation, to the punishing stimulus accounts for instances of recovery (i.e., repeated exposure decreases the aversiveness of the punisher; Goodall, 1984). Moreover, adaptation is more likely to occur with mild punishers, which are typically employed in clinical settings. One strategy that may decrease the likelihood of habituation is the use of hiatus from punishment. In several basic studies with pigeons, response suppression under shock punishment was enhanced following brief time periods during which the subject was removed from the punishing situation or exposed to reinforcement only (e.g., Rachlin, 1966). Further research is needed, however, because the beneficial effect of this procedure was found to wane across repeated punishment–hiatus cycles (e.g., Orme-Johnson, 1967). Other strategies that may prevent or attenuate habituation, such as using intermittent, varied, or brief punishers (e.g., Charlop et al., 1988), should be evaluated in further studies.

Research also is needed on strategies to maintain punishment effects while the intervention is systematically faded. Basic findings with pigeons and monkeys have shown that responding will remain suppressed under low-intensity shock if an initially intense shock is reduced very gradually (e.g., Hake et al., 1967). Further applied research is needed to determine if treatment effects will be maintained while the intensity or duration of a punishment procedure is altered very gradually or less intrusive procedures are simultaneously introduced. For example, it may be possible to reduce a 5-min time-out to a 1-min time-out over time. The use of conditioned punishers may enhance the likelihood of fading certain dimensions of intrusive punishers while treatment effects are maintained over the long run. Moreover, basic studies have found that response recovery is more gradual following the withdrawal of intermittent shock punishment than following the removal of other punishment schedules (e.g., Azrin et al., 1963; Camp et al., 1966). Thus, strategies to increase the utility of conditioned punishers and intermittent punishment for routine clinical practice also may promote the long-term efficacy of punishment.

Several authors have suggested that combining punishment with differential reinforcement may increase the likelihood that punishment can be faded successfully (e.g., Kazdin, 2001). Although this clinical strategy has not been evaluated directly, one study found that differential reinforcement was more effective in reducing problem behavior after a participant had been exposed to a period of punishment (contingent work) than when differential reinforcement preceded punishment (Fisher et al., 1993). Research findings on the indirect effects of punishment suggest that punishment may increase responsivity to reinforcement (see below for further discussion). Thus, punishment may enhance the efficacy of reinforcement for establishing appropriate behavior that competes with or replaces inappropriate behavior, an outcome that in turn may increase the likelihood that punishment can be withdrawn.

Until additional research on long-term maintenance is conducted, practitioners and caregivers should not assume that punishment will remain effective over the long run. Strategies for increasing the likelihood of maintenance should be employed from the outset of treatment. Although basic findings suggest that relatively intense punishers may be associated with successful long-term outcomes, the use of analogous procedures to treat problem behavior probably would raise ethical concerns for all but the most serious cases. Caregivers instead should focus on the use of reinforcement to insure that alternative behavior is at high strength in the repertoire of individuals who are exposed to punishment. Systematic reinforcer assessments and functional analyses of problem behavior always should precede treatment implementation (Fisher et al., 1992). The effects of punishment may last longer if appropriate behavior is maintained by the same reinforcers that maintain problem behavior or by reinforcers that are effective substitutes for maintaining reinforcers. The reinforcing consequences of problem behavior also should be minimized or withheld if possible.

Potential problems with habituation may be curtailed by limiting exposure to the punisher in various ways. For example, caregivers could schedule brief vacations from punishment on a regular basis (Rachlin, 1966) or restrict the use of specific procedures to one or two problem behaviors (e.g., those of greatest concern) instead of applying the same treatment for a variety of responses. Comprehensive punisher assessments also should be employed to identify clinically acceptable procedures that produce the greatest reduction in behavior and, hence, would lead to the least amount of exposure to the punisher (see Fisher, Piazza, Bowman, Hagopian, & Langdon, 1994). If the assessment identifies more than one effective form of punishment, caregivers could alternate among several procedures to minimize exposure to any single punisher.

Finally, practitioners and caregivers should have a plan for dealing with treatment relapse when it occurs during punishment or following the removal of punishment. The first step is to identify and rectify other factors that may be responsible for treatment failure. Many instances of relapse likely are attributable to problems with treatment integrity (D. E. Williams et al., 1993), especially when the punishment procedure is complex or time consuming (e.g., Foxx & Livesay, 1984). The next step is to reassess a wide range of stimuli and activities that may function as potent reinforcers for appropriate behavior and to target additional, multiple responses that might compete with the punished behavior. Alternative forms of punishment should be considered only after determining that the current punisher is ineffective within the context of a richly reinforcing environment. At this point, another comprehensive punisher assessment should be conducted to identify other effective punishers. Increasing the intensity or magnitude of the ineffective punisher is not recommended, as discussed above (see *Reinforcement Magnitude*).

Stimulus Generalization

The transfer of treatment effects across different settings and contexts (i.e., stimulus generalization) is another critically important outcome for individuals with behavior disorders. (Response generalization under punishment, or a concomitant reduction in unpunished behavior when a punished response decreases, will be discussed under the heading *Indirect Effects of Punishment.*) Integration into the community may be restricted even when problem behavior is responsive to treatment if the procedure cannot be implemented everywhere the behavior occurs (e.g., during transitions at school or in public places such as stores and buses). Basic findings on stimulus generalization, however, suggest that the effects of punishment on problem behavior may transfer to untreated settings and contexts. Results of several basic studies with pigeons showed that the suppression in responding produced by shock punishment occurred in the presence of antecedent stimuli that were not used in the training situation, even though punishment was withheld during tests for generalization (e.g, Hoffman & Fleshler, 1965; Honig & Slivka, 1964). The amount of response suppression (i.e., level of stimulus control) was a function of the physical similarity between the generalization stimuli and the stimuli present during training with punishment, a finding that is analogous to basic findings on stimulus generalization and reinforcement effects (Guttman & Kalish, 1956).

Research findings with pigeons and shock punishment, however, may not be directly applicable to humans in clinical settings. In the only two studies to examine stimulus generalization with humans in the laboratory, generalization was relatively difficult to obtain with college students and a more clinically relevant form of punishment (point loss; O'Donnell & Crosbie, 1998; O'Donnell, Crosbie, Williams, & Saunders, 2000). Further basic research is needed with both humans and nonhumans to identify factors that influence the degree and durability of stimulus generalization during punishment. Such factors may include parameters of punishment or reinforcement (e.g., intensity, amount, schedule) and features of the generalization stimuli (e.g., saliency).

In fact, results of numerous applied studies indicate that punishment effects rarely transfer to settings or contexts that are unassociated with punishment delivery (e.g., Corte, Wolf, & Locke, 1971; Doke & Epstein, 1975; Marholin & Townsend, 1978; Rollings, Baumeister, & Baumeister, 1977). Surprisingly few studies have evaluated strategies to promote generalization since Matson and Taras (1989) lamented this gap in a 20-year review of the applied literature on punishment. In early studies, factors such as the presence of the therapist (Risley, 1968) and proximity of the individual to the therapist or treatment setting (Lovaas & Simmons, 1969; Rollings et al.) were found to influence generalization. These findings are consistent with those of basic studies showing a positive relation between amount of response suppression and the degree of similarity between punishment and generalization contexts. This relation may in fact explain why generalization has rarely been observed in applied studies. In most cases, the phenomenon was tested via abrupt alteration of both the stimulus context and the punishment contingency. Participants who continued to exhibit at least some instances of the target behavior would readily detect the transition from a continuous schedule of punishment to the removal of punishment (Azrin & Holz, 1966). In basic studies, responding eventually recovered when the generalization stimuli were repeatedly presented in the absence of punishment.

Thus, it is not surprising that punishment effects typically failed to generalize in the absence of procedures designed to promote transfer (Stokes & Baer, 1977). Several authors have suggested that techniques found to enhance generalization under reinforcement may be similarly effective under punishment (e.g., Matson & DiLorenzo, 1984; Miltenberger, 2001). A few applied studies on punishment have evaluated generalization strategies analogous to those used to promote reinforcement effects. For example, common stimuli were introduced into treatment and generalization settings (e.g., a discriminative stimulus for punishment was presented in the generalization setting; Birnbrauer, 1968), stimuli that might acquire discriminative control over the behavior were removed from the treatment setting (e.g., the therapist was hidden from view; Corte et al., 1971; Tate & Baroff, 1966), and training was conducted with multiple stimulus exemplars (e.g., several different therapists delivered shock; Lovaas & Simmons, 1969). In nearly all cases, however, these strategies were ineffective unless punishment was delivered in the generalization context.

Other potential tactics drawn from the literature on reinforcement include pairing the punisher with naturalistic consequences (e.g., verbal reprimands), varying the stimulus conditions during initial treatment with punishment, providing instruction on self-management, and using delayed or intermittent punishment (i.e., indiscriminable contingencies; Stokes & Baer, 1977; see also O'Donnell & Crosbie, 1998, Experiments 3 and 4). Generalization also may be achieved by implementing a modified form of the treatment in generalization contexts (e.g., delivering a smaller amount of the punisher or a single component of a multicomponent treatment procedure).

Current knowledge about punishment, however, is insufficient to guide the application of such strategies. For example, many factors that are useful for promoting generalization under reinforcement, such as delayed or intermittent contingencies, have been found to undermine the efficacy of punishment (Azrin et al., 1963; Goodall, 1984; Trenholme & Baron, 1975). The suppressive effects of naturalistic consequences and component derivatives of complex interventions likely depend on the process of conditioned punishment, an area that requires further study. Finally, the development of stimulus control under punishment has been evaluated in few basic or applied studies (see Rollings & Baumeister, 1981, for a notable exception). Successful generalization may hinge on the presence of stimuli that have acquired tight control over responding, such that few responses occur in the absence of the punishment contingency. Relative to reinforced responding, it may be difficult to establish control over punished responding with stimuli that are not perfectly correlated with the delivery of punishment. Thus, further research in the areas of conditioned punishment, stimulus control, and intermittent or delayed punishment appears to be critical for developing a technology of generalization.

The current literature indicates that punishment must be delivered consistently in all relevant contexts. Nevertheless, various generalization strategies described by Stokes and Baer (1977) may be useful for promoting treatment generality when the procedure is extended beyond the initial treatment setting. For example, a variety of stimulus conditions could be arranged in the initial treatment setting (e.g., different caregivers and peers could be present, diverse activities could be scheduled, physical features of the environment could vary). Stimuli common to other settings and contexts in which punishment will be applied could be introduced in the initial treatment setting before the intervention is widely implemented. Treatment generality also may be enhanced by ensuring that reinforcement is implemented consistently across settings, incorporating certain aspects of self-management into treatment (e.g., self-monitoring), and establishing salient discriminative stimuli for punishment in all settings and contexts (see Stokes & Baer for further discussion of generalization procedures).

Indirect Effects of Punishment

The effects of punishment on responses that can occur concurrently with the punished behavior or in a different context as the punished behavior also have been studied in basic and applied research. Among these side effects, collateral increases in aggression, escape behavior, and emotional reactions are most commonly described in basic textbooks and literature reviews (e.g., Azrin & Holz, 1966; Mazur, 1998) and by authors who recommend against using punishment in clinical settings (e.g., LaVigna & Donnellan, 1986; McGee, Menolascino, Hobbs, & Menousek, 1987; Parsons et al., 2001).

Aggression (i.e., attacking nearby subjects, biting inanimate objects) in rats, pigeons, and monkeys has been associated with noncontingent delivery of unavoidable stimuli, including shock and intense heat (e.g., Hutchinson, 1977; Ulrich & Azrin, 1962). Although this phenomenon is often called *punishment-elicited aggression,* few studies have examined this side effect of punishment. Basic findings on the effects of inescapable, intense punishers probably have limited generality to the application of punishment (see Linscheid & Meinhold, 1990, for further discussion). Furthermore, elicited aggression in monkeys and rats has been observed to decrease when the subject could exhibit a response (e.g., lever press) to escape from the situation in which the stimulus was delivered (e.g., Azrin, Hutchinson, & Hake, 1966). This finding suggests that elicited aggression may be less problematic during punishment than is commonly assumed because the contingency itself provides an escape response (i.e., delivery of the punisher can be avoided by refraining from the punished behavior). In fact, results of several studies with rats showed that emotional responses in the form of crouching and defecation were more pronounced and persistent when subjects were exposed to unavoidable shock than to response-contingent stimulation (Hearst, 1965; Hunt & Brady, 1955).

On the other hand, numerous basic studies indicate that other forms of unpunished behavior, including responses that occur in the absence of programmed consequences (e.g., species-specific behavior) and those that are maintained by experimenter-delivered reinforcement, may increase, decrease, or remain unchanged during punishment. Factors that determine whether unpunished behavior will increase or decrease (called *contrast* and *induction,* respectively) have not been thoroughly studied. The function of the behavior, schedule and intensity of the punisher, and prior exposure to the punisher may be important (see Crosbie et al., 1997, for a discussion).

In a series of studies, Dunham and colleagues examined the effects of punishment on multiple responses in gerbils by delivering shock contingent on one response (e.g., eating) while changes in alternative responses (e.g., digging, grooming, running) were measured. Results indicated that the most probable of the unpunished responses increased during punishment, whereas responses that tended to follow the punished response decreased (Dunham, 1977, 1978; Dunham & Grantmyre, 1982). A subsequent study suggested that the function of the unpunished behavior also may determine these side effects. Baker, Woods, Tait, and Gardiner (1986) found that when eating in gerbils was punished by shock or noise, digging increased even though running was the most probable response during baseline. The authors suggested that digging was a species-specific response to food deprivation (i.e., digging was in the same response class as eating). Results of studies on behavior that is maintained in the absence of experimenter-arranged reinforcement may have some relevance to problem behavior that produces its own reinforcing consequences (e.g., sensory stimulation).

In other basic studies with humans and pigeons, the same reinforcer was used to establish and maintain two or more functionally equivalent responses (e.g., key pecking, lever pulling) under concurrent or multiple schedules. Punishment in the form of shock, timeout, and point or monetary loss was then delivered for one member of the response class while the effects on the other responses were observed. The most common finding for both humans and nonhumans was an increase in unpunished behavior (i.e., contrast; Bennett & Cherek, 1990; Bradshaw, Szabadi, & Bevan, 1979; Brethower & Reynolds, 1962; Crosbie, 1991; Powell, 1971; Thomas, 1968). However, results were inconsistent both within and across subjects, and the effects often were short-lived. The main findings of Dunham and colleagues (i.e., an increase in the most probable behavior and a decrease in behavior that often followed the punished behavior; Dunham, 1977, 1978; Dunham & Grantmyre, 1982) were not replicated with college students when up to 10 responses were reinforced simultaneously and one response was exposed to contingent point loss (Crosbie, 1990, 1991).

From a clinical standpoint, collateral increases in appropriate behavior and collateral decreases in unpunished inappropriate behavior would be desirable. In fact, applied research findings suggest that a variety of desirable and undesirable side effects can occur within and across individuals. Punishment of problem behavior has been associated with increases in appropriate behavior, such as compliance and toy play (e.g., Koegel, Firestone, Kramme, & Dunlap, 1974; Rolider, Cummings, & Van Houten, 1991) and with decreases in unpunished inappropriate behavior, including aggression and crying (e.g., Bitgood, Crowe, Suarez, & Peters, 1980; Lovaas & Simmons, 1969; Linscheid et al., 1990; Ricketts et al., 1993; Singh, Watson, & Winton, 1986). On the other hand, decreases in appropriate behavior (e.g., toy play, speaking) and increases in unpunished problem behavior (e.g., aggression, emotional reactions, stereotypic behavior) also have been reported (e.g., Bitgood et al., 1980; Duker & Seys, 1996; Foxx & Azrin, 1973; Harris & Wolchick, 1979; Pendergrass, 1971; Singh, Manning, & Angell, 1982; Thompson et al., 1999). A number of authors have suggested that desirable side effects are more likely to occur than undesirable side effects during treatment with punishment (see Lundervold & Bourland, 1988, and Matson & Taras, 1989, for reviews). The prevalence of these side effects is unknown, however, because relatively few studies have directly examined the effects of punishment on unpunished behavior in clinical settings. More important, data collection often was limited to one or two collateral behaviors (see Sisson, Hersen, & Van Hasselt, 1993, for a notable exception). Positive side effects also may be reported more frequently than undesirable side effects in clinical studies on punishment, regardless of the actual prevalence.

Although relatively incomplete, basic findings in this area suggest some possible explanations for the inconsistent outcomes obtained in applied studies. Punishment-elicited aggression and emotional reactions are less probable when exposure to the punishing stimulus can be reduced or avoided (e.g., Azrin et al., 1966; Hunt & Brady, 1955). Thus, these undesirable side effects may have been less likely to occur in clinical studies when a relatively brief punisher suppressed behavior to low levels or when an alternative source of reinforcement was highly effective in strengthening an incompatible response. Conversely, aggression and other inappropriate behavior may have increased in some applied studies because the responses were in the same response class as the punished behavior (Baker et al., 1984). Results of several basic studies with rats and pigeons also indicate that the likelihood of contrast is positively related to the density of the punishment schedule (e.g., Deluty, 1976; Thomas, 1968). Thus, increases in unpunished behavior may have been more probable when the punishment procedure was implemented with a high degree of integrity (i.e., consistently followed each occurrence of the response).

Some basic findings suggest that context may be an important factor in determining which responses increase or decrease during punishment. Bolles, Holtz, Dunn, and Hill (1980) found that induction in rats was more likely to occur when the unpunished response was performed under the same stimulus condition as the punished response (e.g., pushing vs. lifting the same lever) than under different stimulus conditions (e.g., manipulating separate levers). Crosbie et al. (1997) found that induction in both humans and nonhumans was more likely to occur under mixed schedules during the initial exposure to punishment, whereas contrast was more likely to occur under multiple schedules. Together, these findings suggest that functionally equivalent unpunished responses may have been likely to decrease in applied studies if the behavior tended to be exhibited in the same context as the punished behavior.

Further research on factors that determine whether a particular unpunished behavior will increase, decrease, or remain unchanged during punishment is needed so that the direction and nature of side effects can be predicted and controlled. Further research also is needed to determine if these side effects are specific to punishment because similar results have been obtained when a response was suppressed via satiation or response blocking (e.g., Dunham & Grantmyre, 1982), indicating that such effects may be associated with response suppression per se. In fact, several undesirable side effects of punishment (e.g., increases in aggression, escape, and emotional reactions) have been associated with extinction procedures (Lerman & Iwata, 1996a) that often were confounded with punishment in applied research.

Nevertheless, basic findings suggest a number of potentially useful clinical strategies. Punishment-elicited aggression or emotional responses (e.g., crying) may be attenuated by implementing procedures that minimize exposure to the punishing stimulus (e.g., using brief stimuli that produce near-complete suppression of the behavior; strengthening competing responses by delivering potent reinforcers on rich schedules). Comprehensive descriptive or functional analyses of appropriate and inappropriate behavior in an individual's repertoire may be useful for predicting the likelihood of undesirable collateral effects and for arranging conditions to increase desirable effects. Insuring that caregivers withhold reinforcement for unpunished problem behavior might prevent collateral increases in behavior that is in the same response class as the punished behavior. Alternative strategies for managing undesirable contrast effects (e.g., arranging punishment for other inappropriate responses) may be necessary if extinction cannot be used.

Identifying and punishing precursors to dangerous behavior (e.g., mild forms of self-injury that consistently precede more severe forms) may lead to collateral reductions in the severe behavior, augmenting the safety and efficacy of treatment (e.g., Dunham, 1977, 1978). Collateral reductions in functionally equivalent appropriate behavior might be avoided by arranging reinforcement for the behavior under stimulus conditions that are distinctly different from those present when the punished behavior occurs (e.g., by modifying features of the environment or introducing reinforcement in a new setting). To this end, caregivers should ensure that alternative reinforcement is available for multiple responses across a variety of contexts and settings.

Finally, the potential for undesirable side effects that involve caregiver behavior (e.g., overusing punishment, dehumanizing the punished individual) should be given more attention in the applied literature. Although these effects are frequently described in textbooks and reviews (e.g., Cooper et al., 1987; Guess et al., 1987; Kazdin, 2001), few studies have directly evaluated changes in the punishing agent's behavior. Research findings that are relevant to this phenomenon have not supported the hypothesis that using punishment negatively affects caregiver behavior or attitudes toward the punished individual (e.g., Bihm, Sigelman, & Westbrook, 1997; Goza, Ricketts, & Perkins, 1993; Harris, Handleman, Gill, & Fong, 1991; Propst & Nagle, 1981). Bihm et al., for example, found that college students' attitudes toward a fictitious client (i.e., ratings of client competence, adjustment, and learning potential) were related to treatment success rather than to the type of intervention used (i.e., reinforcement vs. mild or intense punishment).

CONCLUSIONS

The use of punishment to treat problem behavior in clinical populations has remained controversial for many years (see Iwata, 1988; Johnston, 1991). Results of basic and applied research indicate that current treatment approaches based on punishment have advantages (e.g., they are highly effective) and disadvantages (e.g., there are unpredictable side effects). Nevertheless, punishment is still sometimes needed to reduce destructive behavior to acceptable levels (e.g., Grace et al., 1994; Hagopian et al., 1998; Wacker et al., 1990); punishment may underlie the effects of certain common function-based treatments (e.g., Lerman & Iwata, 1996b; Mazaleski et al., 1994); and caregivers continue to use punishment to reduce problem behavior in the natural environment (e.g., Peterson & Martens, 1995).

Further understanding of punishment processes is needed to develop a systematic, effective technology of behavior change. A review of the applied literature indicates that a wide range of punishment procedures can successfully treat severe behavior disorders in clinical populations. Variables that contribute to the findings reported in this research, however, have not been delineated. The majority of studies focused on procedural variations of punishment rather than on factors that may influence the direct and indirect effects of punishment. A review of laboratory research on punishment suggests that numerous variables alter basic processes in complex ways. Factors such as history, reinforcement schedule, various punishment parameters, and alternative sources of reinforcement may influence the immediate effects of punishment and other clinically relevant outcomes, such as long-term maintenance, generalization, and the emergence of side effects. Basic findings also contradict some commonly held assumptions about punishment effects and provide possible explanations for the inconsistent findings that have been reported in applied studies.

Nevertheless, current knowledge about basic processes is insufficient for translation to application. The basic literature on some important relations remains incomplete (Baron, 1991; Crosbie, 1998). More important, the extent to which findings with nonhumans and response-contingent electric shock can be extrapolated to the treatment of behavior disorders in clinical populations may be substantially restricted. Although some basic studies have been conducted with human participants, the typical punisher (point or monetary loss) and population (normal adults) may limit the relevance of the findings for individuals with restricted verbal skills (see Hayes & McCurry, 1990, for further discussion).

Research on punishment only tentatively supports most prescriptions for application discussed in previous articles on punishment and in this updated review of the literature. Basic findings also suggest that the process of punishment may be more complex than frequently assumed. As such, clinicians, parents, and teachers should be cautioned about the need for further research on potential strategies for increasing the effectiveness of mild punishers, for attenuating undesirable aspects of punishment, and for successfully fading treatment with punishment.

The basic literature suggests a number of avenues for further research. In particular, knowledge about conditioned, intermittent, and delayed punishment and the interactive effects of reinforcement and punishment may lead to technological advances that increase the effectiveness and acceptability of punishment in clinical settings. Concurrent schedules of reinforcement and punishment are especially germane to application because multiple contingencies typically operate in the natural environment. Treatment also may have a greater likelihood of immediate, long-term, and generalized success when multiple sources of alternative reinforcement are arranged within the context of punishment.

The refinement of functional analysis methodologies has led to greater understanding of variables that can maintain problem behavior and an emphasis on function-based intervention. Yet, potentially important links between function and treatments based on punishment often may be overlooked. The utility of functional analysis for guiding decisions about potential treatments that include a punishment component should be emphasized in further research. Results of functional analyses may indicate whether treatment with reinforcement alone (e.g., differential reinforcement) or with common punishment procedures (e.g., time-out, verbal reprimands) will be successful. For example, time-out is more likely to be effective than contingent verbal reprimands if problem behavior is sensitive to attention (Iwata, Pace, Dorsey, et al., 1994). Problem behavior that occurs at high levels across a wide variety of conditions may be less responsive to treatment with reinforcement than behavior that is differentially low under certain conditions (e.g., the play or control condition of the functional analysis; see Paisey, Whitney, & Hislop, 1990). Functional analysis methodology also may be useful for identifying potent reinforcers to arrange as part of treatment (e.g., for use in differential reinforcement or noncontingent reinforcement procedures) and for identifying functionally equivalent appropriate behavior to strengthen as part of treatment.

Finally, a greater willingness to publish studies showing treatment failures and other undesirable outcomes during punishment may be helpful. For example, unsuccessful attempts to attenuate side effects, to increase the efficacy of delayed punishment, or to promote long-term maintenance and generalization may guide further research for delineating variables that are and are not relevant to clinical application. This knowledge ultimately may lead to more effective, empirically sound recommendations for treatment.

REFERENCES

Abramowitz, A. J., & O'Leary, S. G. (1990). Effectiveness of delayed punishment in an applied setting. *Behavior Therapy, 21,* 231–239.

Alberto, P. A., & Troutman, A. C. (1999). *Applied behavior analysis for teachers* (5th ed.). Upper Saddle River, NJ: Prentice Hall.

Altman, K., Haavik, S., & Cook, J. W. (1978). Punishment of self-injurious behavior in natural settings using contingent aromatic ammonia. *Behaviour Research and Therapy, 16,* 85–96.

Appel, J. B. (1968). Fixed-interval punishment. *Jour-*

nal of the Experimental Analysis of Behavior, 11, 803–808.

Arbuckle, J. L., & Lattal, K. A. (1992). Molecular contingencies in schedules of intermittent punishment. *Journal of the Experimental Analysis of Behavior, 58,* 361–375.

Arntzen, E., & Werner, S. B. (1999). Water mist punishment for two classes of problem behaviour. *Scandinavian Journal of Behaviour Therapy, 28,* 88–93.

Axelrod, S. A. (1990). Myths that (mis)guide our profession. In A. C. Repp & N. N. Singh (Eds.), *Perspectives on the use of nonaversive and aversive interventions for persons with developmental disabilities* (pp. 59–72). Sycamore, IL: Sycamore.

Azrin, N. H. (1956). Some effects of two intermittent schedules of immediate and non-immediate punishment. *Journal of Psychology, 42,* 3–21.

Azrin, N. H. (1960). Effects of punishment intensity during variable-interval reinforcement. *Journal of the Experimental Analysis of Behavior, 3,* 123–142.

Azrin, N. H., & Holz, W. C. (1966). Punishment. In W. K. Honig (Ed.), *Operant behavior: Areas of research and application* (pp. 380–447). New York: Appleton-Century-Crofts.

Azrin, N. H., Holz, W. C., & Hake, D. F. (1963). Fixed-ratio punishment. *Journal of the Experimental Analysis of Behavior, 6,* 141–148.

Azrin, N. H., Hutchinson, R. R., & Hake, D. F. (1966). Extinction induced aggression. *Journal of the Experimental Analysis of Behavior, 9,* 191–204.

Azrin, N. H., & Powers, M. A. (1975). Eliminating classroom disturbances of emotionally disturbed children by positive practice procedures. *Behavior Therapy, 6,* 525–534.

Baker, A. G., Woods, W., Tait, R., & Gardiner, K. (1986). Punishment suppression: Some effects on alternative behaviour. *The Quarterly Journal of Experimental Psychology: Comparative and Physiological Psychology, 38B,* 191–215.

Banks, R. K. (1967). Intermittent punishment effect (IPE) sustained through changed stimulus conditions and through blocks of nonpunished trials. *Journal of Experimental Psychology, 73,* 456–460.

Banks, R. K., & Vogel-Sprott, M. (1965). Effect of delayed punishment on an immediately rewarded response in humans. *Journal of Experimental Psychology, 70,* 357–359.

Baron, A. (1991). Avoidance and punishment. In I. H. Iversen & K. A. Lattal (Eds.), *Experimental analysis of behavior* (Part 1, pp. 173–217). Amsterdam: Elsevier Science.

Baron, A., Kaufman, A., & Fazzini, D. (1969). Density and delay of punishment of free-operant avoidance. *Journal of the Experimental Analysis of Behavior, 12,* 1029–1037.

Barrett, R. P., Matson, J. L., Shapiro, E. S., & Ollendick, T. H. (1981). A comparison of punishment and DRO procedures for treating stereotypic behavior of mentally retarded children. *Applied Research in Mental Retardation, 2,* 247–256.

Barton, L. E., Brulle, A. R., & Repp, A. C. (1987). Effects of differential schedule of time-out to reduce maladaptive responding. *Exceptional Children, 53,* 351–356.

Bennett, R. H., & Cherek, D. R. (1990). Punished and nonpunished responding in a multiple schedule in humans: A brief report. *The Psychological Record, 40,* 187–196.

Bihm, E. M., Sigelman, C. K., & Westbrook, J. P. (1997). Social implications of behavioral interventions for persons with mental retardation. *American Journal on Mental Retardation, 101,* 567–578.

Birnbrauer, J. S. (1968). Generalization of punishment effects: A case study. *Journal of Applied Behavior Analysis, 1,* 201–211.

Bitgood, S. C., Crowe, M. J., Suarez, Y., & Peters, R. D. (1980). Imbolization: Effects and side effects on stereotyped behavior in children. *Behavior Modification, 4,* 187–208.

Boe, E. E. (1964). Extinction as a function of intensity of punishment, amount of training, and reinforcement of a competing response. *Canadian Journal of Psychology, 18,* 328–342.

Bolles, R. C., Holtz, R., Dunn, T., & Hill, W. (1980). Comparisons of stimulus learning and response learning in a punishment situation. *Learning and Motivation, 11,* 78–96.

Bouzas, A. (1978). The relative law of effect: Effects of shock intensity on response strength in multiple schedules. *Journal of the Experimental Analysis of Behavior, 30,* 307–314.

Bradshaw, C. M., Szabadi, E., & Bevan, P. (1977). Effect of punishment on human variable-interval performance. *Journal of the Experimental Analysis of Behavior, 27,* 275–279.

Bradshaw, C. M., Szabadi, E., & Bevan, P. (1978). Effect of variable-interval punishment on the behavior of humans in variable-interval schedules of monetary reinforcement. *Journal of the Experimental Analysis of Behavior, 29,* 161–166.

Bradshaw, C. M., Szabadi, E., & Bevan, P. (1979). The effect of punishment on free-operant choice behavior in humans. *Journal of the Experimental Analysis of Behavior, 31,* 71–81.

Brethower, D. M., & Reynolds, G. S. (1962). A facilitative effect of punishment on unpunished behavior. *Journal of the Experimental Analysis of Behavior, 5,* 191–199.

Brown, R. T., & Wagner, A. R. (1964). Resistance to punishment and extinction following training with shock or nonreinforcement. *Journal of Experimental Psychology, 68,* 503–507.

Burchard, J. D., & Barrera, F. (1972). An analysis of timeout and response cost in a programmed environment. *Journal of Applied Behavior Analysis, 5,* 271–282.

Calhoun, K. S., & Matherne, P. (1975). The effects of varying schedules of time-out on aggressive behavior of a retarded girl. *Journal of Behavior Therapy and Experimental Psychiatry, 6,* 139–143.

Camp, D. S., Raymond, G. A., & Church, R. M. (1966). Response suppression as a function of the schedule of punishment. *Psychonomic Science, 5,* 23–24.

Camp, D. S., Raymond, G. A., & Church, R. M. (1967). Temporal relationship between response and punishment. *Journal of Experimental Psychology, 74,* 114–123.

Capaldi, E. D., Sheffer, J. D., Viveiros, D. M., Davidson, T. L., & Campbell, D. H. (1985). Shock preexposure and the reduced effectiveness of shock. *Learning and Motivation, 16,* 357–380.

Charlop, M. H., Burgio, L. D., Iwata, B. A., & Ivancic, M. T. (1988). Stimulus variation as a means of enhancing punishment effects. *Journal of Applied Behavior Analysis, 21,* 89–95.

Church, R. M. (1969). Response suppression. In B. A. Campbell & R. M. Church (Eds.), *Punishment and aversive behavior* (pp. 111–156). New York: Appleton-Century-Crofts.

Cipani, E., Brendlinger, J., McDowell, L., & Usher, S. (1991). Continuous vs. intermittent punishment: A case study. *Journal of Developmental and Physical Disabilities, 3,* 147–156.

Clark, H. B., Rowbury, T., Baer, A. M., & Baer, D. M. (1973). Timeout as a punishing stimulus in continuous and intermittent schedules. *Journal of Applied Behavior Analysis, 6,* 443–455.

Cohen, P. S. (1968). Punishment: The interactive effects of delay and intensity of shock. *Journal of the Experimental Analysis of Behavior, 11,* 789–799.

Cole, G. A., Montgomery, R. W., Wilson, K. M., & Milan, M. A. (2000). Parametric analysis of overcorrection duration effects: Is longer really better than shorter? *Behavior Modification, 24,* 359–378.

Cooper, J. O., Heron, T. E., & Heward, W. L. (1987). *Applied behavior analysis.* New York: Macmillan.

Corte, H. E., Wolf, M. M., & Locke, B. J. (1971). A comparison of procedures for eliminating self-injurious behavior of retarded adolescents. *Journal of Applied Behavior Analysis, 4,* 201–213.

Crosbie, J. (1990). Some effects of response cost on a free-operant multiple-response repertoire with humans. *The Psychological Record, 40,* 517–539.

Crosbie, J. (1991). The effects of punishment on unpunished reinforced free-operant responses. *Australian Journal of Psychology, 43,* 1–5.

Crosbie, J. (1998). Negative reinforcement and punishment. In K. A. Lattal & M. Perone (Eds.), *Handbook of research methods in human operant behavior* (pp. 163–189). New York: Plenum.

Crosbie, J., Williams, A. M., Lattal, K. A., Anderson, M. M., & Brown, S. M. (1997). Schedule interactions involving punishment with pigeons and humans. *Journal of the Experimental Analysis of Behavior, 68,* 161–175.

Crowell, C. R. (1974). Conditioned aversive aspects of electric shock. *Learning and Motivation, 5,* 209–220.

Davidson, R. S. (1970). Conditioned punishment and conditioned negative reinforcement on a multiple schedule. *Psychonomic Science, 20,* 163–165.

Deluty, M. Z. (1976). Choice and the rate of punishment in concurrent schedules. *Journal of the Experimental Analysis of Behavior, 25,* 75–80.

Deluty, M. Z. (1978). Self-control and impulsiveness involving aversive events. *Journal of Experimental Psychology, 4,* 250–266.

Deur, J. L., & Parke, R. D. (1970). Effects of inconsistent punishment on aggression in children. *Developmental Psychology, 2,* 403–411.

Dinsmoor, J. A. (1952). A discrimination based on punishment. *Quarterly Journal of Experimental Psychology, 4,* 27–45.

Dixon, M. J., Helsel, W. J., Rojahn, J., Cipollone, R., & Lubetsky, M. J. (1989). Aversive conditioning of visual screening with aromatic ammonia for treating aggressive and disruptive behavior in a developmentally disabled child. *Behavior Modification, 13,* 91–107.

Doke, L. A., & Epstein, L. H. (1975). Oral overcorrection: Side effects and extended applications. *Journal of Experimental Child Psychology, 20,* 496–511.

Donnellan, A. M., & LaVigna, G. W. (1990). Myths about punishment. In A. C. Repp & N. N. Singh (Eds.), *Perspectives on the use of nonaversive and aversive interventions for persons with developmental disabilities* (pp. 33–57). Sycamore, IL: Sycamore.

Dorsey, M. F., Iwata, B. A., Ong, P., & McSween, T. E. (1980). Treatment of self-injurious behavior using a water mist: Initial response suppression and generalization. *Journal of Applied Behavior Analysis, 13,* 343–353.

Duker, P. C., & Seys, D. M. (1996). Long-term use of electrical aversion treatment with self-injurious behavior. *Research in Developmental Disabilities, 17,* 293–301.

Dunham, P. J. (1977). The nature of reinforcing stimuli. In W. K. Honig & J. E. R. Staddon (Eds.), *Handbook of operant behavior* (pp. 98–124). Englewood Cliffs, NJ: Prentice Hall.

Dunham, P. J. (1978). Changes in unpunished responding during response-contingent punishment. *Animal Learning & Behavior, 6,* 174–180.

Dunham, P. J., & Grantmyre, J. (1982). Changes in a multiple-response repertoire during response-contingent punishment and response restriction: Sequential relationships. *Journal of the Experimental Analysis of Behavior, 37,* 123–133.

Dura, J. R. (1991). Controlling extremely dangerous aggressive outbursts when functional analysis fails. *Psychological Reports, 69,* 451–459.

Eisenberger, R., Weier, F., Masterson, F. A., & Theis,

L. Y. (1989). Fixed-ratio schedules increase generalized self-control: Preference for large rewards despite high effort or punishment. *Journal of Experimental Psychology: Animal Behavior Processes, 15,* 383–392.

Emerson, E. (1992). Self-injurious behaviour: An overview of recent trends in epidemiological and behavioural research. *Mental Handicap Research, 5,* 49–81.

Epstein, R. (1984). An effect of immediate reinforcement and delayed punishment, with possible implications for self-control. *Journal of Behavior Therapy and Experimental Psychiatry, 15,* 291–298.

Estes, W. K. (1944). An experimental study of punishment. *Psychological Monographs, 57* (Serial No. 3).

Evans, W. O. (1962). Producing either positive or negative tendencies to a stimulus associated with shock. *Journal of the Experimental Analysis of Behavior, 5,* 335–337.

Fantino, E. (1973). Aversive control. In J. A. Nevin & G. S. Reynolds (Eds.), *The study of behavior: Learning, motivation, emotion, and instinct* (pp. 239–279). Glenview, IL: Scott, Foresman.

Farley, J. (1980). Reinforcement and punishment effects in concurrent schedules: A test of two models. *Journal of the Experimental Analysis of Behavior, 33,* 311–326.

Favell, J. E., Azrin, N. H., Baumeister, A. A., Carr, E. G., Dorsey, M. F., Forehand, R., et al. (1982). The treatment of self-injurious behavior. *Behavior Therapy, 13,* 529–554.

Fisher, W. W., Piazza, C. C., Bowman, L. G., Hagopian, L. P., & Langdon, N. A. (1994). Empirically-derived consequences: A data-based method for prescribing treatments for destructive behavior. *Research in Developmental Disabilities, 15,* 133–149.

Fisher, W. W., Piazza, C. C., Bowman, L. G., Hagopian, L. P., Owens, J. C., & Slevin, I. (1992). A comparison of two approaches for identifying reinforcers for persons with severe and profound disabilities. *Journal of Applied Behavior Analysis, 25,* 491–498.

Fisher, W. W., Piazza, C. C., Bowman, L. G., Kurtz, P. F., Sherer, M. R., & Lachman, S. R. (1994). A preliminary evaluation of empirically derived consequences for the treatment of pica. *Journal of Applied Behavior Analysis, 27,* 447–457.

Fisher, W. W., Piazza, C. C., Cataldo, M. F., Harrell, R., Jefferson, G., & Conner, R. (1993). Functional communication training with and without extinction and punishment. *Journal of Applied Behavior Analysis, 26,* 23–36.

Foxx, R. M., & Azrin, N. H. (1973). The elimination of autistic self-stimulatory behavior by overcorrection. *Journal of Applied Behavior Analysis, 6,* 1–14.

Foxx, R. M., Bittle, R. G., & Faw, G. D. (1989). A maintenance strategy for discontinuing aversive procedures: A 32-month follow-up of the treatment of aggression. *American Journal on Mental Retardation, 94,* 27–36.

Foxx, R. M., & Livesay, J. (1984). Maintenance of response suppression following overcorrection. *Analysis and Intervention in Developmental Disabilities, 1,* 65–79.

Galbicka, G., & Branch, M. N. (1981). Selective punishment of interresponse times. *Journal of the Experimental Analysis of Behavior, 35,* 311–322.

Goh, H., Iwata, B. A., Shore, B. A., DeLeon, I. G., Lerman, D. C., Ulrich, S. M., et al. (1995). An analysis of the reinforcing properties of hand mouthing. *Journal of Applied Behavior Analysis, 28,* 269–283.

Goodall, G. (1984). Learning due to the response-shock contingency in signaled punishment. *Quarterly Journal of Experimental Psychology, 36B,* 259–279.

Goza, A. B., Ricketts, R. W., & Perkins, T. S. (1993). The social validity of an argument supporting a ban on aversive procedures. *Journal of Intellectual Disability Research, 37,* 449–458.

Grace, N. C., Kahng, S. W., & Fisher, W. W. (1994). Balancing social acceptability with treatment effectiveness of an intrusive procedure: A case report. *Journal of Applied Behavior Analysis, 27,* 171–172.

Grace, N. C., Thompson, R. H., & Fisher, W. W. (1996). The treatment of covert self-injury through contingencies on response products. *Journal of Applied Behavior Analysis, 29,* 239–242.

Guess, D., Helmstetter, E., Turnbull, H. R., III, & Knowlton, S. (1987). *Use of aversive procedures with persons who are disabled: An historical review and critical analysis.* Seattle: The Association for Persons with Severe Handicaps.

Guttman, N., & Kalish, H. I. (1956). Discriminability and stimulus generalization. *Journal of Experimental Psychology, 51,* 79–88.

Hagopian, L. P., Fisher, W. W., Sullivan, M. T., Acquisto, J., & LeBlanc, L. A. (1998). Effectiveness of functional communication training with and without extinction and punishment. *Journal of Applied Behavior Analysis, 31,* 211–235.

Hake, D. F., & Azrin, N. H. (1965). Conditioned punishment. *Journal of the Experimental Analysis of Behavior, 8,* 279–293.

Hake, D. F., Azrin, N. H., & Oxford, R. (1967). The effects of punishment intensity on squirrel monkeys. *Journal of the Experimental Analysis of Behavior, 10,* 95–107.

Halevy, G., Feldon, J., & Weiner, I. (1987). Resistance to extinction and punishment following training with shock and non-reinforcement: Failure to obtain cross-tolerance. *Quarterly Journal of Experimental Psychology, 39B,* 147–160.

Harris, S. L., Handleman, J. S., Gill, M. J., & Fong, P. L. (1991). Does punishment hurt? The impact of aversives on the clinician. *Research in Developmental Disabilities, 12,* 17–24.

Harris, S. L., & Wolchick, S. A. (1979). Suppression of self-stimulation: Three alternative strategies. *Journal of Applied Behavior Analysis, 12,* 185–198.

Hayes, L. J., & McCurry, C. (1990). Moral and scientific aspects of the punishment controversy. In A. C. Repp & N. N. Singh (Eds.), *Perspectives on the use of nonaversive and aversive interventions for persons with developmental disabilities* (pp. 87–101). Sycamore, IL: Sycamore.

Hearst, E. (1965). Stress induced breakdown of an appetitive discrimination. *Journal of the Experimental Analysis of Behavior, 8,* 135–146.

Herman, R. L., & Azrin, N. H. (1964). Punishment by noise in an alternative response situation. *Journal of the Experimental Analysis of Behavior, 7,* 185–188.

Hobbs, S. A., Forehand, R., & Murray, R. G. (1978). Effects of various durations of timeout on the non-compliant behavior of children. *Behavior Therapy, 9,* 652–656.

Hoffman, H. S., & Fleshler, M. (1965). Stimulus aspects of aversive controls: The effects of response contingent shock. *Journal of the Experimental Analysis of Behavior, 8,* 89–96.

Holz, W. C., & Azrin, N. H. (1962). Recovery during punishment by intense noise. *Psychological Reports, 11,* 655–657.

Holz, W. C., & Azrin, N. H. (1963). A comparison of several procedures for eliminating behavior. *Journal of the Experimental Analysis of Behavior, 6,* 399–406.

Holz, W. C., Azrin, N. H., & Ayllon, T. (1963). Elimination of behavior of mental patients by response-produced extinction. *Journal of the Experimental Analysis of Behavior, 6,* 407–412.

Honig, W. K., & Slivka, R. M. (1964). Stimulus generalization of the effects of punishment. *Journal of the Experimental Analysis of Behavior, 7,* 21–25.

Hunt, H. F., & Brady, J. V. (1955). Some effects of punishment and intercurrent "anxiety" on a simple operant. *Journal of Comparative and Physiological Psychology, 48,* 305–310.

Hutchinson, R. R. (1977). By-products of aversive control. In W. K. Honig & J. E. R. Staddon (Eds.), *Handbook of operant behavior* (pp. 415–431). Englewood Cliffs, NJ: Prentice Hall.

Iwata, B. A. (1988). The development and adoption of controversial default technologies. *The Behavior Analyst, 11,* 149–157.

Iwata, B. A., Pace, G. M., Cowdery, G. E., & Miltenberger, R. G. (1994). What makes extinction work: An analysis of procedural form and function. *Journal of Applied Behavior Analysis, 27,* 131–144.

Iwata, B. A., Pace, G. M., Dorsey, M. F., Zarcone, J. R., Vollmer, T. R., Smith, R. G., et al. (1994). The functions of self-injurious behavior: An experimental-epidemiological analysis. *Journal of Applied Behavior Analysis, 27,* 215–240.

Iwata, B. A., Vollmer, T. R., & Zarcone, J. R. (1990). The experimental (functional) analysis of behavior disorders: Methodology, applications, and limitations. In A. C. Repp & N. N. Singh (Eds.), *Perspectives on the use of nonaversive and aversive interventions for persons with developmental disabilities* (pp. 301–330). Sycamore, IL: Sycamore.

Jacob-Timm, S. (1996). Ethical and legal issues associated with the use of aversives in the public schools: The SIBIS controversy. *School Psychology Review, 25,* 184–198.

Johnson, D. L., McGlynn, F. D., & Topping, J. S. (1973). The relative efficiency of four response-elimination techniques following variable-ratio reinforcement training. *The Psychological Record, 23,* 203–208.

Johnston, J. M. (1991). What can behavior analysis learn from the aversives controversy? *The Behavior Analyst, 14,* 187–196.

Kaufman, A., & Baron, A. (1968). Suppression of behavior by timeout punishment when suppression results in loss of positive reinforcement. *Journal of the Experimental Analysis of Behavior, 11,* 595–607.

Kazdin, A. E. (2001). *Behavior modification in applied settings* (6th ed.). Belmont, CA: Wadsworth/Thompson Learning.

Kendall, P. C., Nay, W. R., & Jeffers, J. (1975). Timeout duration and contrast effects: A systematic evaluation of a successive treatments design. *Behavior Therapy, 6,* 609–615.

Kennedy, C. H., & Souza, G. (1995). Functional analysis and treatment of eye poking. *Journal of Applied Behavior Analysis, 28,* 27–37.

Koegel, R. L., Firestone, P. B., Kramme, K. W., & Dunlap, G. (1974). Increasing spontaneous play by suppressing self-stimulation in autistic children. *Journal of Applied Behavior Analysis, 7,* 521–528.

Lalli, J. S., & Goh, H. (1993). Naturalistic observations in community settings. In R. Reichle & D. P. Wacker (Eds.), *Communicative alternatives to challenging behavior* (pp. 11–39). Paul H. Brookes.

Lande, S. D. (1981). An interresponse time analysis of variable-ratio punishment. *Journal of the Experimental Analysis of Behavior, 35,* 55–67.

Laurence, M. T., Hineline, P. N., & Bersh, P. J. (1994). The puzzle of responding maintained by response-contingent shock. *Journal of the Experimental Analysis of Behavior, 61,* 135–153.

LaVigna, G. W., & Donnellan, A. M. (1986). *Alternatives to punishment: Solving behavior problems with non-aversive strategies.* New York: Irvington.

Lerman, D. C., & Iwata, B. A. (1996a). Developing a technology for the use of operant extinction in clinical settings: An examination of basic and applied research. *Journal of Applied Behavior Analysis, 29,* 345–382.

Lerman, D. C., & Iwata, B. A. (1996b). A methodology for distinguishing between extinction and punishment effects associated with response blocking. *Journal of Applied Behavior Analysis, 29,* 231–233.

Lerman, D. C., Iwata, B. A., Shore, B. A., & DeLeon, I. G. (1997). Effects of intermittent punishment on self-injurious behavior: An evaluation of schedule thinning. *Journal of Applied Behavior Analysis, 30,* 187–201.

Lindberg, J. S., Iwata, B. A., & Kahng, S. W. (1999). On the relation between object manipulation and stereotypic self-injurious behavior. *Journal of Applied Behavior Analysis, 32,* 51–62.

Linscheid, T. R., Hartel, F., & Cooley, N. (1993). Are aversive procedures durable? A five year follow-up of three individuals treated with contingent electric shock. *Child and Adolescent Mental Health Care, 3,* 67–76.

Linscheid, T. R., Iwata, B. A., Ricketts, R. W., Williams, D. E., & Griffin, J. C. (1990). Clinical evaluation of the self-injurious behavior inhibiting system (SIBIS). *Journal of Applied Behavior Analysis, 23,* 53–78.

Linscheid, T. R., & Meinhold, P. (1990). The controversy over aversives: Basic operant research and side effects of punishment. In A. C. Repp & N. N. Singh (Eds.), *Perspectives on the use of nonaversive and aversive interventions for persons with developmental disabilities* (pp. 435–450). Sycamore, IL: Sycamore.

Lovaas, O. I., & Favell, J. E. (1987). Protection for clients undergoing aversive/restrictive interventions. *Education and Treatment of Children, 10,* 311–325.

Lovaas, O. I., & Simmons, J. Q. (1969). Manipulation of self-destruction in three retarded children. *Journal of Applied Behavior Analysis, 2,* 143–157.

Lundervold, D., & Bourland, G. (1988). Quantitative analysis of treatment of aggression, self-injury, and property destruction. *Behavior Modification, 12,* 590–617.

Marholin, D., & Townsend, N. M. (1978). An experimental analysis of side effects and response maintenance of a modified overcorrection procedure. *Behavior Therapy, 9,* 383–390.

Martin, G., & Pear, J. (1996). *Behavior modification: What it is and how to do it* (5th ed.). Upper Saddle River, NJ: Prentice Hall.

Matson, J. L., & DiLorenzo, T. M. (1984). *Punishment and its alternatives: New perspectives for behavior modification.* New York: Springer.

Matson, J. L., & Taras, M. E. (1989). A 20 year review of punishment and alternative methods to treat problem behaviors in developmentally delayed persons. *Research in Developmental Disabilities, 10,* 85–104.

Mazaleski, J. L., Iwata, B. A., Rodgers, T. A., Vollmer, T. R., & Zarcone, J. R. (1994). Protective equipment as treatment for stereotypic hand mouthing: Sensory extinction or punishment effects? *Journal of Applied Behavior Analysis, 27,* 345–355.

Mazur, J. E. (1998). *Learning and behavior* (4th ed.). Upper Saddle River, NJ: Prentice Hall.

McGee, J. J., Menolascino, F. J., Hobbs, D. C., & Menousek, P. E. (1987). *Gentle teaching: A nonaversive approach to helping persons with mental retardation.* New York: Human Sciences Press.

McMillan, D. E. (1967). A comparison of the punishing effects of response-produced shock and response-produced time out. *Journal of the Experimental Analysis of Behavior, 10,* 439–449.

Michael, J. L. (1993). *Concepts and principles of behavior analysis.* Kalamazoo, MI: Society for the Advancement of Behavior Analysis.

Miller, N. B., & Zimmerman, J. (1966). The effects of a pre-time-out stimulus on matching-to-sample of humans. *Journal of the Experimental Analysis of Behavior, 9,* 487–499.

Miller, N. E. (1960). Learning resistance to pain and fear effects over learning, exposure, and rewarded exposure in context. *Journal of Experimental Psychology, 60,* 137–145.

Miltenberger, R. G. (2001). *Behavior modification: Principles and procedures* (2nd ed.). Belmont, CA: Wadsworth.

Mowrer, O. H., & Aiken, E. G. (1954). Contiguity vs. drive-reduction in conditioned fear: Temporal variations in conditioned and unconditioned stimulus. *American Journal of Psychology, 67,* 26–38.

Mowrer, O. H., & Solomon, L. N. (1954). Contiguity vs. drive-reduction in conditioned fear: The proximity and abruptness of drive-reduction. *American Journal of Psychology, 67,* 15–25.

Murray, M., & Nevin, J. A. (1967). Some effects of correlation between response-contingent shock and reinforcement. *Journal of the Experimental Analysis of Behavior, 10,* 301–309.

O'Brien, F. (1989). Punishment for people with developmental disabilities. In E. Cipani (Ed.), *The treatment of severe behavior disorders: Behavior analysis approaches* (pp. 37–58). Washington, DC: American Association of Mental Retardation.

O'Donnell, J., & Crosbie, J. (1998). Punishment generalization gradients with humans. *The Psychological Record, 48,* 211–232.

O'Donnell, J., Crosbie, J., Williams, D. C., & Saunders, K. J. (2000). Stimulus control and generalization of point-loss punishment with humans. *Journal of the Experimental Analysis of Behavior, 73,* 261–274.

Orme-Johnson, D. W. (1967). Response suppression as a function of a vacation from punishment. *Psychonomic Science, 8,* 277–278.

Orme-Johnson, D. W., & Yarczower, M. (1974). Conditioned suppression. *Journal of the Experimental Analysis of Behavior, 21,* 57–74.

Paisey, T. J. H., Whitney, R. B., & Hislop, P. M. (1990). Client characteristics and treatment selection: Legitimate influences and misleading inferences. In A. C. Repp & N. N. Singh (Eds.), *Perspectives on the use of nonaversive and aversive interventions for persons with developmental disabilities* (pp. 175–197). Sycamore, IL: Sycamore.

Parsons, R. D., Hinson, S. L., & Sardo-Brown, D. (2001). *Educational psychology: A practitioner-researcher model of teaching.* Belmont, CA: Wadsworth/Thomson Learning.

Patel, M. R., Carr, J. E., Kim, C., Robles, A., & Eastridge, D. (2000). Functional analysis of aberrant behavior maintained by automatic reinforcement: Assessments of specific sensory reinforcers. *Research in Developmental Disabilities, 21,* 393–407.

Pendergrass, V. E. (1971). Effects of length of timeout from positive reinforcement and schedule of application reinforcement and schedule of application in suppression of aggressive behavior. *The Psychological Record, 21,* 75–80.

Peterson, F. M., & Martens, B. K. (1995). A comparison of behavioral interventions reported in treatment studies and programs for adults with developmental disabilities. *Research in Developmental Disabilities, 16,* 27–41.

Piazza, C. C., Adelinis, J. D., Hanley, G. P., Goh, H., & Delia, M. D. (2000). An evaluation of the effects of matched stimuli on behaviors maintained by automatic reinforcement. *Journal of Applied Behavior Analysis, 33,* 13–27.

Powell, R. W. (1970). The effect of punishment shock intensity upon responding under multiple schedules. *Journal of the Experimental Analysis of Behavior, 14,* 201–211.

Powell, R. W. (1971). Some effects of punishment shock intensity upon discriminative responding. *Journal of the Experimental Analysis of Behavior, 15,* 109–116.

Propst, L. B., & Nagle, R. J. (1981). Effects of labeling and a child's reaction to punishment on subsequent disciplinary practices of adults and peers. *American Journal of Mental Deficiency, 86,* 287–294.

Rachlin, H. (1966). Recovery of responses during mild punishment. *Journal of the Experimental Analysis of Behavior, 9,* 251–263.

Rawson, R. A., & Leitenberg, H. (1973). Reinforced alternative behavior during punishment and extinction with rats. *Journal of Comparative and Physiological Psychology, 85,* 593–600.

Reid, D. J., Parsons, M. B., Phillips, J. F., & Green, C. W. (1993). Reduction of self-injurious hand mouthing using response blocking. *Journal of Applied Behavior Analysis, 26,* 139–140.

Ricketts, R. W., Goza, A. B., & Matese, M. (1993). A 4-year follow-up of treatment of self-injury. *Journal of Behavior Therapy and Experimental Psychiatry, 24,* 57–62.

Rincover, A. (1978). Sensory extinction: A procedure for eliminating self-stimulatory behavior in developmentally disabled children. *Journal of Abnormal Psychology, 6,* 299–310.

Risley, T. R. (1968). The effects and side effects of punishing the autistic behaviors of a deviant child. *Journal of Applied Behavior Analysis, 1,* 21–34.

Rodriguez, W. A., & Logan, F. A. (1980). Preference for punishment of the instrumental or the consummatory response. *Animal Learning & Behavior, 8,* 116–119.

Rolider, A., Cummings, A., & Van Houten, R. (1991). Side effects of therapeutic punishment on academic performance and eye contact. *Journal of Applied Behavior Analysis, 24,* 763–773.

Rolider, A., & Van Houten, R. (1985). Suppressing tantrum behavior in public places through the use of delayed punishment mediated by audio recordings. *Behavior Therapy, 16,* 181–194.

Rolider, A., & Van Houten, R. (1990). The role of reinforcement in reducing inappropriate behavior: Some myths and misconceptions. In A. C. Repp & N. N. Singh (Eds.), *Perspectives on the use of nonaversive and aversive interventions for persons with developmental disabilities* (pp. 119–127). Sycamore, IL: Sycamore.

Rolider, A., Williams, L., Cummings, A., & Van Houten, R. (1991). The use of a brief movement restriction procedure to eliminate severe inappropriate behavior. *Journal of Behavior Therapy and Experimental Psychiatry, 22,* 23–30.

Rollings, J. P., & Baumeister, A. A. (1981). Stimulus control of stereotypic responding: Effects on target and collateral behavior. *American Journal of Mental Deficiency, 86,* 67–77.

Rollings, J. P., Baumeister, A. A., & Baumeister, A. A. (1977). The use of overcorrection procedures to eliminate the stereotyped behaviors of retarded individuals: An analysis of collateral behaviors and generalization of suppressive effects. *Behavior Modification, 1,* 29–46.

Romanczyk, R. G. (1977). Intermittent punishment of self-stimulation: Effectiveness during application and extinction. *Journal of Consulting and Clinical Psychology, 45,* 53–60.

Scobie, S. R., & Kaufman, A. (1969). Intermittent punishment of human responding maintained by intermittent reinforcement. *Journal of the Experimental Analysis of Behavior, 12,* 137–147.

Scotti, J. R., Evans, I. M., Meyer, L. H., & Walker, P. (1991). A meta-analysis of intervention research with problem behavior: Treatment validity and standards of practice. *American Journal on Mental Retardation, 96,* 233–256.

Shemer, A., & Feldon, J. (1984). Long-term partial reinforcement extinction effect and long-term partial punishment effect in a one-trial-a-day paradigm. *Bulletin of the Psychonomic Society, 22,* 221–224.

Shore, B. A., Iwata, B. A., DeLeon, I. G., Kahng, S., & Smith, R. G. (1997). An analysis of reinforcer substitutability using object manipulation and

self-injury as competing responses. *Journal of Applied Behavior Analysis, 30,* 21–40.
Sidman, M. (1989). *Coercion and its fallout.* Boston: Authors Cooperative.
Singh, N. N., Dawson, M. J., & Manning, P. J. (1981). The effects of physical restraint on self-injurious behaviour. *Journal of Mental Deficiency Research, 25,* 207–216.
Singh, N. N., Manning, P. J., & Angell, M. J. (1982). Effects of oral hygiene punishment procedure on chronic rumination and collateral behaviors in monozygous twins. *Journal of Applied Behavior Analysis, 15,* 309–314.
Singh, N. N., Watson, J. E., & Winton, A. S. (1986). Treating self-injury: Water mist spray versus facial screening or forced arm exercise. *Journal of Applied Behavior Analysis, 19,* 403–410.
Sisson, L. A., Hersen, M., & Van Hasselt, V. B. (1993). Improving the performance of youth with dual sensory impairment: Analyses and social validation of procedures to reduce maladaptive responding in vocational and leisure settings. *Behavior Therapy, 24,* 553–571.
Skinner, B. F. (1938). *The behavior of organisms: An experimental analysis.* Acton, MA: Copley.
Solnick, J. V., Rincover, A., & Peterson, C. R. (1977). Some determinants of the reinforcing and punishing effects of timeout. *Journal of Applied Behavior Analysis, 10,* 415–424.
Stokes, T. F., & Baer, D. M. (1977). An implicit technology of generalization. *Journal of Applied Behavior Analysis, 10,* 349–367.
Tate, B. G., & Baroff, G. S. (1966). Aversive control of self-injurious behavior in a psychotic boy. *Behaviour Research and Therapy, 4,* 281–287.
Terris, W., & Barnes, M. (1969). Learned resistance to punishment and subsequent responsiveness to the same and novel punishers. *Psychonomic Science, 15,* 49–50.
Thomas, J. R. (1968). Fixed-ratio punishment by timeout of concurrent variable-interval behavior. *Journal of the Experimental Analysis of Behavior, 11,* 609–616.
Thompson, R. H., Iwata, B. A., Conners, J., & Roscoe, E. M. (1999). Effects of reinforcement for alternative behavior during punishment of self-injury. *Journal of Applied Behavior Analysis, 32,* 317–328.
Thorndike, E. L. (1932). *Fundamentals of learning.* New York: Teachers College.
Trenholme, I. A., & Baron, A. (1975). Immediate and delayed punishment of human behavior by loss of reinforcement. *Learning and Motivation, 6,* 62–79.
Ulrich, R. E., & Azrin, N. H. (1962). Reflexive fighting in response to aversive stimulation. *Journal of the Experimental Analysis of Behavior, 5,* 511–520.
Van Houten, R. (1983). Punishment: From the animal laboratory to the applied setting. In S. Axelrod & J. Apsche (Eds.), *The effects of punishment on human behavior* (pp. 13–44). New York: Academic Press.
Van Houten, R., Axelrod, S., Bailey, J. S., Favell, J. E., Foxx, R. N., Iwata, B. A., et al. (1988). The right to effective behavioral treatment. *The Behavior Analyst, 11,* 111–114.
Van Houten, R., & Rolider, A. (1988). Recreating the scene: An effective way to provide delayed punishment for inappropriate motor behavior. *Journal of Applied Behavior Analysis, 21,* 187–192.
Vogel-Sprott, M. (1967). Partial-reward training for resistance to punishment and to subsequent extinction. *Journal of Experimental Psychology, 75,* 138–140.
Vollmer, T. R., Borrero, J. C., Wright, C. S., Van Camp, C., & Lalli, J. S. (2001). Identifying possible contingencies during descriptive analyses of severe behavior disorders. *Journal of Applied Behavior Analysis, 34,* 269–287.
Vollmer, T. R., & Iwata, B. A. (1993). Implications of a functional analysis technology for the use of restrictive behavioral interventions. *Child and Adolescent Mental Health Care, 3,* 95–113.
Wacker, D. P., Steege, M. W., Northup, J., Sasso, G. M., Berg, W., Reimers, T., et al. (1990). A component analysis of functional communication training across three topographies of severe behavior problems. *Journal of Applied Behavior Analysis, 23,* 417–429.
Walker, J. E., & Shea, T. M. (1999). *Behavior management: A practical approach for educators* (7th ed.). Upper Saddle River, NJ: Prentice Hall.
Weiner, H. (1962). Some effects of response cost upon human operant behavior. *Journal of the Experimental Analysis of Behavior, 5,* 201–208.
Weiner, H. (1964). Response cost and fixed-ratio performance. *Journal of the Experimental Analysis of Behavior, 7,* 79–81.
White, G. D., Nielsen, G., & Johnson, S. M. (1972). Timeout duration and the suppression of deviant behavior in children. *Journal of Applied Behavior Analysis, 5,* 111–120.
Williams, D. E., Kirkpatrick-Sanchez, S., & Crocker, W. T. (1994). A long-term follow-up of treatment for severe self-injury. *Research in Developmental Disabilities, 15,* 487–501.
Williams, D. E., Kirkpatrick-Sanchez, S., & Iwata, B. A. (1993). A comparison of shock intensity in the treatment of longstanding and severe self-injurious behavior. *Research in Developmental Disabilities, 14,* 207–219.
Williams, D. R., & Barry, H., III. (1966). Counter conditioning in an operant conflict situation. *Journal of Comparative and Physiological Psychology, 61,* 154–156.

Received September 1, 2001
Final acceptance August 12, 2002
Action Editor, Joseph Spradlin