

*EFFECTS OF TREATMENT INTEGRITY FAILURES DURING  
DIFFERENTIAL REINFORCEMENT OF ALTERNATIVE BEHAVIOR:  
A TRANSLATIONAL MODEL*

CLAIRE ST. PETER PIPKIN

WEST VIRGINIA UNIVERSITY

AND

TIMOTHY R. VOLLMER AND KIMBERLY N. SLOMAN

UNIVERSITY OF FLORIDA

Differential reinforcement of alternative behavior (DRA) is used frequently as a treatment for problem behavior. Previous studies on treatment integrity failures during DRA suggest that the intervention is robust, but research has not yet investigated the effects of different types of integrity failures. We examined the effects of two types of integrity failures on DRA, starting with a human operant procedure and extending the results to children with disabilities in a school setting. Human operant results (Experiment 1) showed that conditions involving reinforcement for problem behavior were more detrimental than failing to reinforce appropriate behavior alone, and that condition order affected the results. Experiments 2 and 3 replicated the effects of combined errors and sequence effects during actual treatment implementation.

*Key words:* autism, human operant, sequence effects, translational research, treatment integrity

Differential reinforcement of alternative behavior (DRA), a commonly used treatment for problem behavior, typically involves withholding reinforcers following problem behavior (extinction) and providing reinforcers contingent on some appropriate, alternative response. For example, a DRA treatment for attention-maintained screaming might involve refraining from talking to or making eye contact with the individual while he or she is screaming and providing attention following some appropriate behavior, such as saying hello. When implemented with high levels of treatment integrity, DRA produces substantial decreases in problem behavior and increases in appropriate behavior (e.g., Carr & Durand, 1985; Tiger, Hanley, & Bruzek, 2008; Vollmer & Iwata, 1992). However, caregivers may implement DRA with relatively low levels of treatment integrity

because those individuals may have a long history of reinforcing problem behavior (and may therefore find extinction procedures difficult), or they may have difficulty providing reinforcers consistently following alternative responses (Volkert, Lerman, Call, & Trosclair-Lasserre, 2009).

To date, the effects of treatment integrity failures on DRA procedures have not been examined extensively. Thus, the level of integrity adequate for achieving intervention effects during differential reinforcement procedures remains largely unknown. Studies examining reduced integrity on differential reinforcement have obtained mixed results (Mazaleski, Iwata, Vollmer, Zarcone, & Smith, 1993; Mueller et al., 2003; Worsdell, Iwata, Hanley, Thompson, & Kahng, 2000). These mixed results may be due to experimenters combining different types of integrity failures into a single measure of overall integrity. Some components of DRA procedures are implemented with better integrity than others (e.g., Coddington, Feinberg, Dunn, & Pace, 2005), which may have an

---

Correspondence concerning this article should be addressed to Claire St. Peter Pipkin, Box 6040, West Virginia University, Morgantown, West Virginia 26506 (e-mail: [claire.stpeterpipkin@mail.wvu.edu](mailto:claire.stpeterpipkin@mail.wvu.edu)).

doi: 10.1901/jaba.2010.43-47

impact on treatment outcomes. For example, providing reinforcers following problem behavior (one type of failure) may have different effects than failing to deliver reinforcers following appropriate behavior (another type of failure). Collapsing these failures into a single measure of treatment integrity would not capture the differences between types of failures. Thus, single measures of treatment integrity might suggest that less than optimal levels of integrity (e.g., 80%) are acceptable in some circumstances (when the majority of failures involve noncritical treatment components) but are unacceptable in others (when the failures occur on critical treatment components). In other words, poor integrity on noncritical components of the treatment may result in a low overall integrity score even though the effectiveness of the treatment is retained.

Two (of many) possible types of treatment integrity failures involve either the failure to deliver an earned reinforcer according to the treatment schedule (here termed an *omission error*) or the delivery of a reinforcer following problem behavior (here termed a *commission error*). These two types of errors reflect integrity failures during different components of DRA (the programmed reinforcement schedule component and the extinction component, respectively).

An error of omission occurs when the caregiver fails to deliver a reinforcer at the correct time. Northup, Fisher, Kahng, Harrell, and Kurtz (1997) examined the effects of omission errors on a treatment involving both differential reinforcement and punishment. Integrity failures involved intermittently omitting treatment components. For example, during evaluations of integrity failures, timeout was delivered on a variable-ratio (VR) schedule that successively doubled (e.g., fixed-ratio [FR] 1 to VR 2 to VR 4), and reinforcement was available according to variable-interval (VI) schedules that doubled in duration (e.g., VI 1 min and VI 2 min). The

treatment retained its efficacy for all 3 participants when therapists implemented both components with 50% integrity. Slight increases in problem behavior occurred when integrity dropped to 25% for 2 of the 3 participants. These results suggest that omission errors may be detrimental to treatment outcome if integrity decreases to low levels. However, Northup *et al.* did not examine the effects of integrity on differential reinforcement independent of the punishment procedure, so the effects of omission errors on DRA remain unknown.

Research on the effects of ongoing reinforcement for problem behavior during behavioral treatments may provide some evidence of the possible effects of commission errors. Often, the failure to implement the extinction component of the DRA causes the treatment to fail entirely, with rates of problem behavior remaining high and rates of appropriate behavior low or zero (e.g., Kelley, Lerman, & Van Camp, 2002; Shirley, Iwata, Kahng, Mazaleski, & Lerman, 1997; Worsdell *et al.*, 2000). For example, Worsdell *et al.* examined the effects of ongoing reinforcement for problem behavior (commission errors) by varying the reinforcement schedule for problem behavior while maintaining an FR 1 schedule for appropriate behavior. For all participants, response allocation eventually shifted toward appropriate behavior as the schedule for problem behavior became thinner. These results suggest that errors of commission may be detrimental to treatments like DRA.

Vollmer, Roane, Ringdahl, and Marcus (1999) examined the effects of combined commission and omission errors, which involved both periodic reinforcement of problem behavior (commission errors) and failures to reinforce appropriate behavior (omission errors). Overall, the effects of DRA were resistant to decrements in integrity level. However, rates of problem behavior increased somewhat when the experimenters reinforced both desirable and undesirable responses. These increases in response rate could be problematic when dealing

with severe problem behavior (e.g., self-injury, aggression, or property destruction) and again underscore the potential negative effects of treatment integrity failures. One limitation of the Vollmer et al. study was that almost all of the integrity failure phases followed a phase in which the treatment was implemented perfectly, raising the possibility that sequence effects played a role.

The Vollmer et al. (1999) investigation represents one of the few studies that have parametrically examined the types and levels of integrity failures on DRA. One potential reason for this lack of research may be due to difficulties associated with conducting parametric studies with clinical populations. In particular, these studies may require extensive periods of time to complete, which may delay successful treatment development. In addition, parametric treatment research may not yield differentiated results without preliminary data to guide the selection of parameter types and levels that are likely to affect treatment outcomes. One practical means of obtaining preliminary parametric data is through translational research.

Translational research typically begins with controlled laboratory studies that are later replicated with clinical populations (Lerman, 2003). Translational research may afford more control over variables than is usually available in application, such as precise delivery of antecedent stimuli and reinforcers. Isolation of particular variables may illuminate the most influential factors, which can later be examined in application. Translational research may also allow more rapid manipulation of variables than is typically afforded with clinical populations, for whom the influence of extraneous variables such as therapist differences, limited participant availability, or the participants' histories may affect the outcome of studies, particularly those that use relatively brief experimental phases.

The purpose of the current investigation was to study the effects of treatment integrity

failures on DRA, using a translational research model. Experiment 1 used a laboratory procedure to examine the effects of errors of omission alone, errors of commission alone, and combined errors of omission and commission during an analogue DRA treatment, at five levels of failure and with different sequences of exposure. Experiments 2 and 3 replicated the combined errors examined in Experiment 1 with children with developmental disabilities.

## EXPERIMENT 1: INITIAL EVALUATION OF INTEGRITY FAILURES

### *Method*

*Participants and setting.* Twenty-two undergraduate students enrolled in an introductory psychology course at the University of Florida participated. Students received course credit for completing the experiment, but this credit was not dependent on performance during the experimental sessions. Each session involved only 1 participant. We conducted all sessions in a laboratory room that was equipped with a computer desk, a computer, and a chair. Students participated for a total of 123 min, which included two 60-min blocks with a 3-min break between blocks.

*Procedure.* When the student arrived, we asked him or her to read and sign a consent form that stated that the experiment assessed the effects of different contingencies of reinforcement, but it did not specify the reinforcement schedules. The experimenter told the participant that he or she would be working at a computer and should use only the mouse to earn as many points as possible during the session.

During the sessions, the computer screen was blank except for one red circle, one black circle, and a cumulative point score. The circles were 127 mm in diameter and moved at a speed of 25 mm per second in random directions. Participants earned points according to the programmed schedules of reinforcement, which differed for both circles. We arbitrarily defined

clicking on the black circle as analogous to engaging in problem behavior and clicking on the red circle as analogous to engaging in appropriate behavior. The background of the screen was beige throughout all sessions; schedule-correlated stimuli were not used.

During baseline, clicking on the black circle was reinforced on an FR 1 schedule and clicking on the red circle was not reinforced (extinction). During the full-integrity DRA, clicking on the black circle was not reinforced (extinction) and clicking on the red circle was reinforced on an FR 1 schedule. The baseline and full-integrity DRA schedules were designed to replicate those published in the majority of studies that have evaluated DRA as a treatment for problem behavior (e.g., Shirley *et al.*, 1997; Vollmer, Iwata, Smith, & Rodgers, 1992; Vollmer *et al.*, 1999; Walsh, 1991).

During treatment integrity failures, we reinforced one or both of the available responses according to random-ratio (RR) schedules. An RR schedule is a type of VR schedule in which each response is associated with a particular probability of reinforcement. For example, in an RR 10 schedule, each response would have a .1 probability of resulting in reinforcer delivery. In the current experiments, each level of treatment integrity failure was associated with a particular probability. For example, 80% integrity with both omission and commission errors (hereafter, combined errors) was associated with a .8 probability of reinforcement for appropriate behavior (80% of responses being reinforced; an RR 1.25 schedule) and a .2 probability of reinforcement for problem behavior (80% of responses going unreinforced; an RR 5 schedule). Although the use of RR schedules is only one possible type of integrity failure, the use of ratio-based schedules was consistent with prior research on treatment integrity failures during DRA (Vollmer *et al.*, 1999; Worsdell *et al.*, 2000). The programmed levels of treatment integrity failures are described below, but in some cases, the pro-

grammed levels of treatment integrity differed slightly from the obtained probabilities of reinforcement as a function of participant response allocation. We calculated obtained probabilities of reinforcement for each of the programmed levels by dividing the total number of reinforcers delivered contingent on a response by the number of responses in that phase. Programmed and obtained probabilities differed by a mean of .01 (range, 0 to .03) across all subsets and phases of the experiment (obtained probability values for each phase are available from the first author).

At the beginning of the experiment, the experimenter instructed participants to earn as many points as possible. Points were not exchangeable for any backup reinforcers. Clear reinforcement effects and appropriate changes in rates of responding following contingency changes between at least one replication of baseline (FR 1 extinction schedule) and treatment (DRA schedule with perfect integrity, extinction FR 1) were required for a participant to be included in the experiment. Participants who did not show differentiation in response rates between baseline and DRA phases would have been excluded; however, all participants met the inclusion criteria.

Participants were assigned randomly to one of four subsets (described below) that varied in the type of integrity manipulation. For three of the subsets, condition sequence remained constant, regardless of subset assignment, to control for the influence of sequence effects. All participants were exposed to baseline, full-integrity DRA, and four phases of reduced treatment integrity (80%, 60%, 40%, and 20%), using a reversal design. The sequence of conditions for each subset is shown in Table 1.

Participants in Subset 1 were exposed only to omission errors; for these participants, reduced levels of treatment integrity meant that the schedule of reinforcement for appropriate behavior became thinner (e.g., probability

Table 1  
Condition Sequence for Experiment 1

Condition number	Condition type			
	Subset 1	Subset 2	Subset 3	Subset 4
1	BL	BL	BL	BL
2	DRA	DRA	DRA	DRA
3	80 omission	80 commission	80 combined	BL
4	60 omission	60 commission	60 combined	DRA
5	40 omission	40 commission	40 combined	50
6	20 omission	20 commission	20 combined	BL
7	BL	BL	BL	50
8	DRA	DRA	DRA	DRA
9	20 omission	20 commission	20 combined	50
10	40 omission	40 commission	40 combined	BL
11	60 omission	60 commission	60 combined	50
12	80 omission	80 commission	80 combined	DRA

decreased from 1.0 to .2). As treatment integrity levels decreased, these participants could earn fewer reinforcers for engaging in appropriate behavior, but could never earn points for engaging in problem behavior (extinction). The purpose of this subset was to examine the effects of missed reinforcer deliveries for appropriate behavior alone, without introduction of reinforcement for problem behavior.

Participants in Subset 2 were exposed to commission errors only. For these participants, reduced levels of treatment integrity meant that some reinforcers were introduced into the extinction schedule for problem behavior. That is, as treatment integrity levels became lower, the reinforcement schedule for problem behavior became richer, but appropriate behavior was always reinforced on an FR 1 schedule. The purpose of this subset was to examine the effects of introducing accidental reinforcement for problem behavior.

Participants in Subset 3 were exposed to combined errors of omission and commission. For these participants, as fewer reinforcers were available for problem behavior (fewer commission errors), more reinforcers were available for appropriate behavior (fewer omission errors). The purpose of this subset was to examine the combined effects of missed reinforcer deliveries for appropriate behavior and reinforcement for problem behavior.

Participants in Subset 4 experienced only one level of treatment integrity failure (50% integrity on both the reinforcement and extinction components), but the condition sequence was altered such that treatment integrity failures followed baseline twice and full-integrity DRA twice. We designed this manipulation to address the possibility that condition sequence affects the outcome of treatment integrity failures.

### *Results and Discussion*

Figure 1 shows the results for participants in Subset 1 (omission errors only). For all 3 participants in this group, rates of problem behavior were high and rates of appropriate behavior were low during baseline. They engaged in high rates of appropriate behavior and low rates of problem behavior when we implemented DRA with full integrity. During the first exposure to treatment integrity failures, participants usually engaged in appropriate behavior at somewhat lower rates when the treatment integrity decreased to 40% or 20%. In general, rates of appropriate responding remained high during the second exposure to the integrity failure phases, with the exception of the final failure phase (80% integrity). Regardless of the level of errors of omission in place, participants in Subset 1 did not engage in elevated rates of problem behavior. This finding

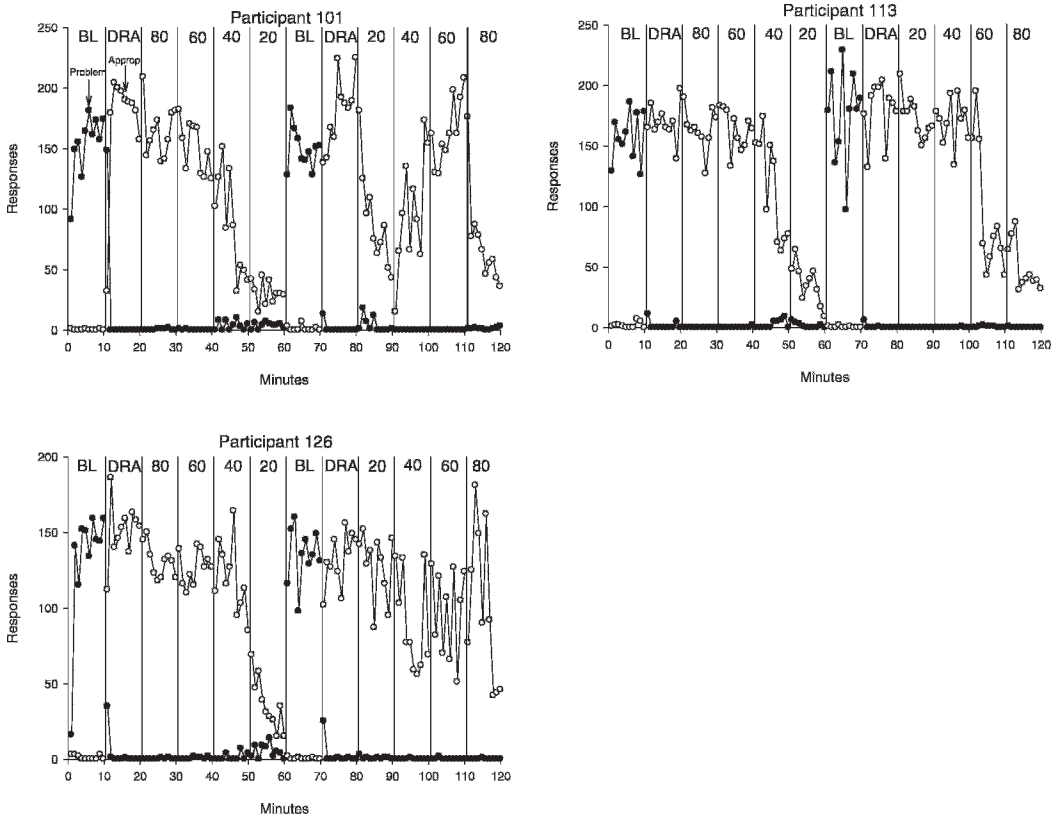


Figure 1. Results for participants in Subset 1, who were exposed to omission errors only. Each panel shows results for a participant. Filled circles denote rates of clicking on the black object (problem behavior), and open circles denote rates of clicking on the red object (appropriate behavior). Condition labels show baseline (BL), DRA, and treatment integrity failure phases. During integrity failures, condition labels show treatment integrity as omission integrity or commission integrity.

suggests that errors of omission in isolation may not be highly detrimental to DRA treatments. In fact, the low levels of appropriate behavior may be desirable in some cases (cf. Hanley, Iwata, & Thompson, 2001).

Figure 2 shows results for participants in Subset 2 (commission errors only). For these participants, commission errors did not become detrimental until the level of treatment integrity dropped to 40% or lower. For Participants 206 (upper right panel) and 222 (lower left panel), high rates of appropriate behavior and low rates of problem behavior were observed during DRA at full integrity and during 80% and 60% treatment integrity. However, rates of appro-

priate behavior decreased, with a corresponding increase in problem behavior at 40% and 20% integrity. The increase in problem behavior during 40% and 20% integrity was somewhat surprising. Because participants could consistently earn a point for each appropriate response, it seems counterintuitive that they would allocate responding to a thinner reinforcement schedule (i.e., the treatment integrity failures for problem behavior). Overall, however, the DRA treatment seemed relatively robust when treatment integrity was greater than 60% with commission errors.

Figure 3 shows results from participants in Subset 3. For these participants, errors of

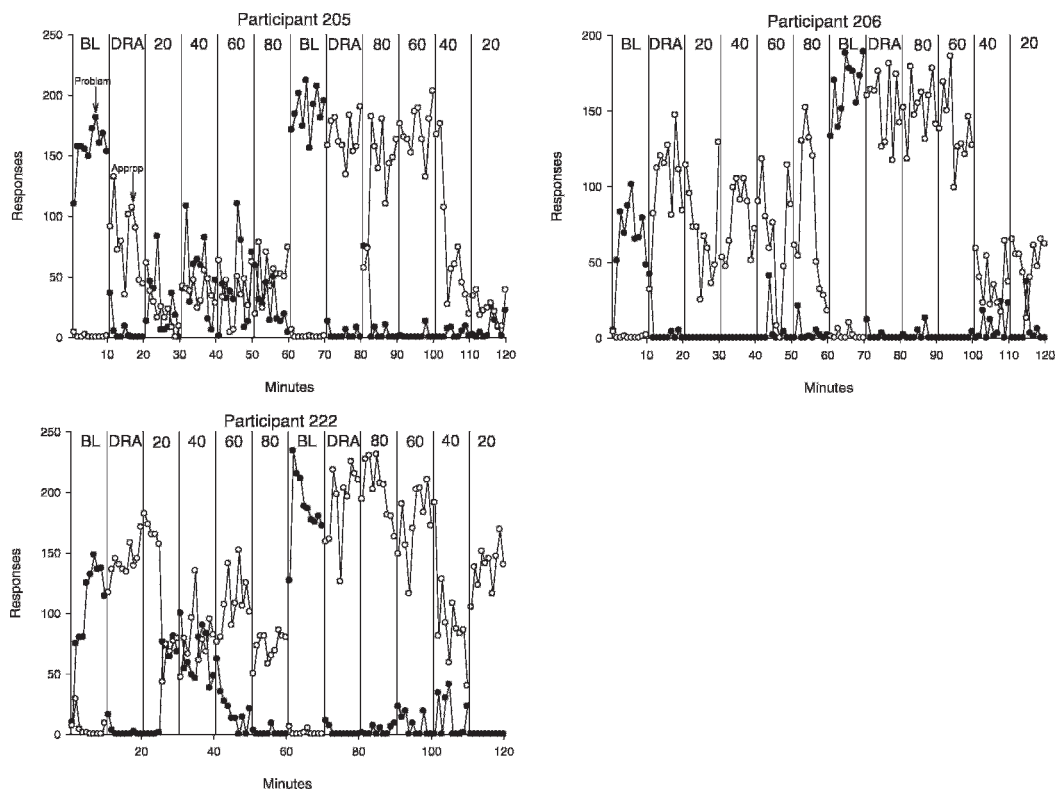


Figure 2. Results for participants in Subset 2, who were exposed to commission errors only. Each panel shows results for a participant. Filled circles denote rates of clicking on the black object (problem behavior), and open circles denote rates of clicking on the red object (appropriate behavior). Condition labels show baseline (BL), DRA, and treatment integrity failure phases. During integrity failures, condition labels show treatment integrity as omission integrity or commission integrity.

omission and commission covaried. For example, during the 20% integrity conditions, we degraded integrity on both the FR 1 and extinction components of the DRA; this resulted in a 20% chance that appropriate behavior would result in a point and an 80% chance that problem behavior would result in a point. Results were consistent across the 3 participants in Subset 3. They engaged in high rates of problem behavior and low rates of appropriate behavior when treatment integrity was 20% or 40%. However, once treatment integrity reached 60%, response allocations switched, with participants engaging in more appropriate behavior than problem behavior. Condition sequence did not affect these results.

The results obtained from Subset 3 could be explained based on the reinforcement rate available from each response type. When problem behavior was more likely to result in reinforcement than was appropriate behavior (during the 20% and 40% integrity phases), participants engaged in more problem behavior than appropriate behavior. Conversely, when appropriate behavior was more likely to result in reinforcement than was problem behavior (during the 60% and 80% integrity phases), they were more likely to engage in appropriate behavior than problem behavior.

Figures 4 and 5 show the results from participants in Subset 4. In general, two patterns of responding occurred across the 13 participants: Responding during the error phase

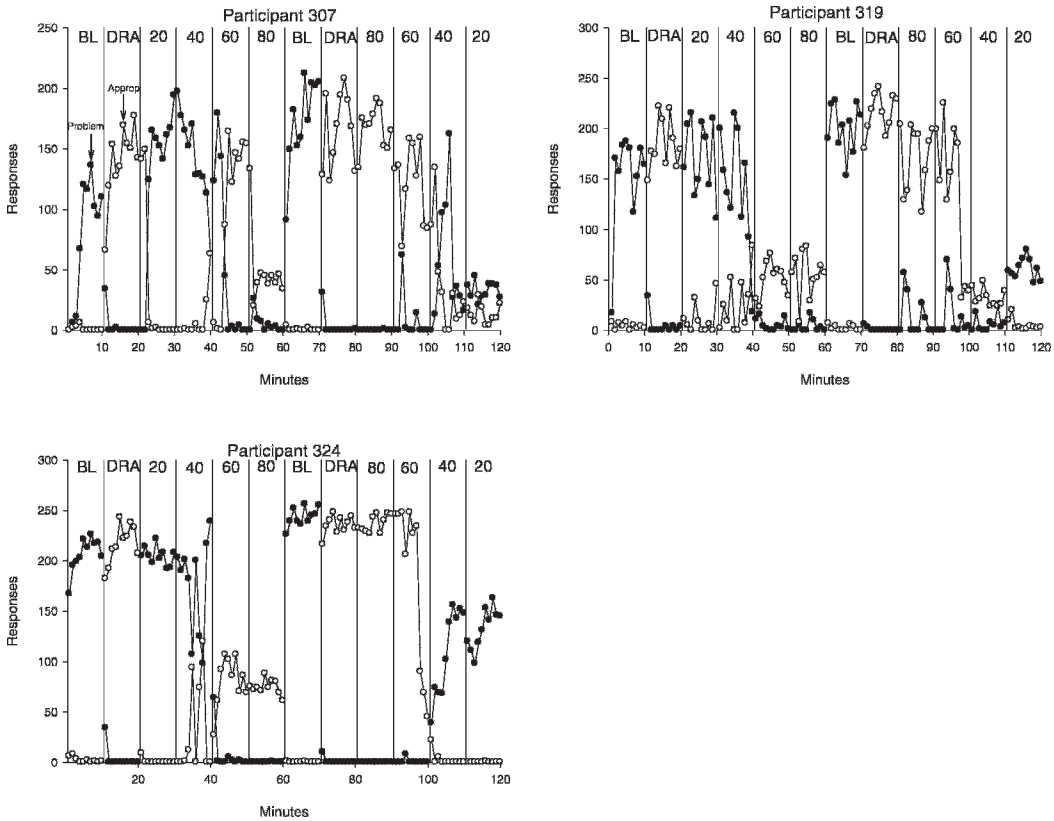


Figure 3. Results for participants in Subset 3, who were exposed to both omission and commission errors. Each panel shows results for a participant. Filled circles denote rates of clicking on the black object (problem behavior), and open circles denote rates of clicking on the red object (appropriate behavior). Condition labels show baseline (BL), DRA, and treatment integrity failure phases. During integrity failures, condition labels show treatment integrity as omission integrity or commission integrity.

carried over from the previous phase (i.e., participants continued to allocate responding toward the most recently reinforced response; Figure 4) or responding was almost completely allocated to the first reinforced response (Figure 5). Figure 4 shows the results for participants whose response allocations during baseline and DRA carried over into the treatment integrity phases. For these 5 participants, the contingencies in place during the preceding phase (baseline or full-integrity DRA) influenced responding during the subsequent 50% integrity phases. For example, all of these participants engaged in higher rates of problem behavior during the 50% integrity conditions that followed baseline and usually displayed

higher levels of appropriate behavior during the 50% integrity condition that followed the full-integrity DRA.

The remaining 8 participants in Subset 4 did not show consistent carryover during the 50% integrity failure conditions. Figure 5 shows the results for these participants. Carryover occurred in some phases and seemed particularly likely to occur in the first exposure to the integrity failure condition following DRA (in which carryover occurred for 7 of 8 participants). However, as the experiment progressed, these participants responded in an unexpected manner during the integrity failures: Response allocation completely switched from the previous phase. In other words, they primarily



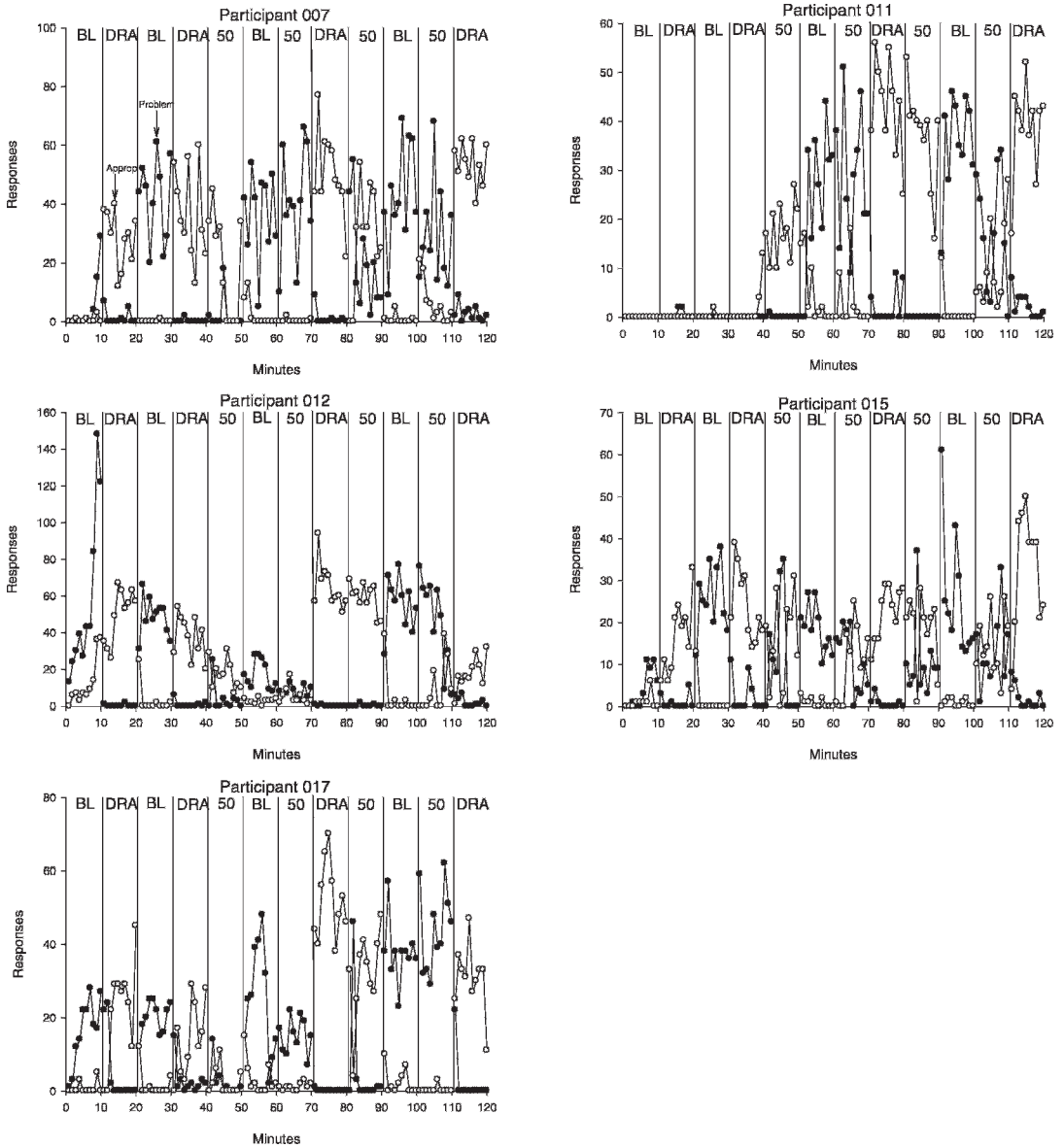


Figure 4. Partial results for participants in Subset 4, who were exposed to 50% integrity following baseline and full treatment. Each panel shows results for a participant. All of these participants showed consistent carryover between baseline or treatment and error phases. Filled circles denote rates of clicking on the black object (problem behavior), and open circles denote rates of clicking on the red object (appropriate behavior). Condition labels show baseline (BL), DRA, and treatment integrity failure phases. During integrity failures, condition labels show treatment integrity as omission integrity or commission integrity.

engaged in appropriate behavior during 50% integrity phases that followed baseline and in problem behavior during 50% integrity phases that followed DRA. This change in allocation

occurred for the second, third, and fourth exposures to 50% integrity for Participants 001 and 007, the first and fourth exposures for 019, the second and third exposures for 005, the

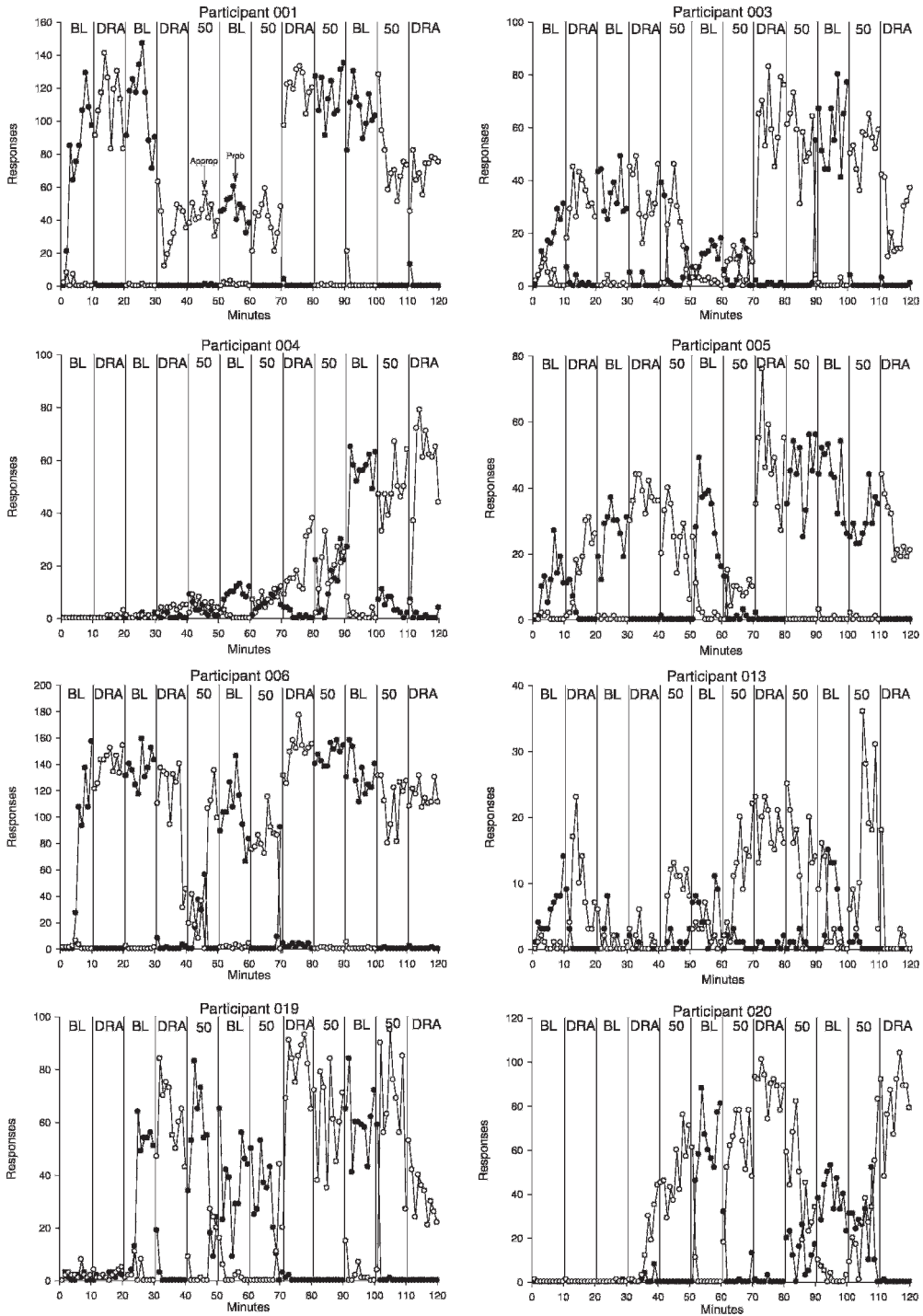


Figure 5. Results for some of the participants in Subset 4, who were exposed to 50% integrity following baseline and full treatment. Each panel shows results for a participant. All of these participants showed some evidence of switching between baseline or treatment and error phases. Filled circles denote rates of clicking on the black object (problem behavior), and open circles denote rates of clicking on the red object (appropriate behavior). Condition labels show baseline (BL), DRA, and treatment integrity failure phases. During integrity failures, condition labels show the treatment integrity percentage.

second and fourth exposures for 013 and 020, and the fourth exposure for 003 and 004. In almost all of these conditions, the participant engaged in high rates of one response and low rates of the other response throughout the phase.

The changes in response allocation may have been due to the reinforcement schedules that we used during the baseline and DRA conditions. As mentioned previously, we chose to use FR 1 and extinction schedules in these conditions because those schedules were used most commonly in previous DRA research. However, it is possible that the continuous reinforcement schedule exerted sufficient control over responding, such that the introduction of the 50% integrity condition signaled participants to alter response allocations. For example, it could be that once the condition changed from DRA to 50% integrity, the first unreinforced appropriate response resulted in switching to problem behavior. Alternatively, it is possible that the first reinforcer delivery for one response resulted in a change in response allocation to that response (i.e., reinforcer delivery altered response allocation).

The switching of response allocation observed for those participants in Subset 4 also may be due to the experimental arrangement. Because the ideal response pattern was alternation between problem behavior and appropriate behavior every 10 min for the first four phases, it is possible that participants generated a rule about responding, such as “switch every 10 min.” Although they did not have any exteroceptive means of timing the conditions, it is possible that the passage of time exerted control over responding.

In sum, the results from Experiment 1 demonstrated that the efficacy of DRA treatments decreased based on different kinds of treatment integrity failures. Errors of commission had a greater impact on responding than did errors of omission, but only at relatively low levels of treatment integrity (20% and 40%). Thus, the DRA arrangements presented in this

experiment appeared to be a relatively robust “treatment” overall, in that generally lower levels of integrity continued to support treatment efficacy (decreased levels of problem behavior and maintenance of appropriate behavior). This outcome may be because participants who experienced a DRA treatment with errors of commission only can maximize reinforcement by continuing to respond on the richer reinforcement schedule (i.e., the FR 1 schedule for appropriate behavior).

Combined errors seem to have the same effects on behavior as commission errors only. This suggests that, when both types of failures are combined into a single measure, commission errors may be more responsible for detrimental effects than omission errors. To illustrate, at 50% combined treatment integrity failures, both the reinforcement schedules in effect and the participant’s recent reinforcement history influenced problem behavior. This is evidenced by the differences in responding during the 50% integrity phases that followed DRA implemented with full integrity and those that followed baseline (Figure 4).

Combined errors may have detrimental effects because participants attempt to maximize reinforcement. As mentioned previously, the participants exposed to combined errors (Subset 3) allocated the majority of their responses to the richer reinforcement schedule. This result may have important implications for application because caregivers may be likely to make combined errors (commission and omission) when attempting to implement DRA treatments. For example, caregivers with a long history of reinforcing problem behavior may be prone to reinforcing problem behavior on a relatively rich schedule and appropriate behavior on a relatively lean schedule, similar to the lower levels of treatment integrity experienced by Subset 3. If this is the case, those caregivers may see little to no improvement in the behavior, despite their attempts to implement the treatment as it was designed.

Results from Subset 4 highlight the possibility that treatment integrity failures, particularly those that do not clearly favor one response over another, may be particularly detrimental if implemented following baseline probabilities of reinforcement. For example, if a caregiver implements the DRA with less than optimal (e.g., 50%) integrity following training, there may be no improvement in the problem behavior, at least initially. Alternatively, if the caregiver can create a reinforcement history for alternative behavior initially by implementing DRA with high integrity, later treatment integrity failures may not be as detrimental. Recent reinforcement history influences the impact of integrity failures, thereby emphasizing the importance of initially implementing the DRA with high integrity.

Although the results of Experiment 1 have implications for application, the generality of these results was unclear. In particular, we obtained the results with a nonclinical population responding on an analogue task, which may not have generality to the treatment of problem behavior displayed by individuals with disabilities. Thus, Experiments 2 and 3 extended the results of Experiment 1 to clinical application.

## EXPERIMENT 2: APPLIED EVALUATION OF COMBINED ERRORS

The purpose of Experiment 2 was to evaluate the effects of combined errors on the occurrence of problem and appropriate behavior during DRA, in an attempt to replicate the results of Subset 3 from Experiment 1. Thus, we designed this experiment as one means of validating the results from the human operant laboratory. We chose to replicate the results of Subset 3 because it seemed probable that naturally occurring integrity failures during DRA included both commission and omission errors and because of the detrimental effects of these errors on DRA efficacy observed during the combined integrity errors in Experiment 1.

### *Method*

*Participant and setting.* Helena was a fourth-grade student who had been diagnosed with autism. She spoke in complete sentences, independently requested items and activities, and participated in a regular education classroom for the majority of her school day. Helena's teacher had referred her to a school-based program for the assessment and treatment of off-task behavior (defined below).

All sessions were conducted in an empty therapy room (4 m by 4 m) in Helena's school that was equipped with a table, chairs, and leisure items appropriate for a variety of different ages and skill levels. Sessions were conducted 2 or 3 days per week, with two to four sessions per day. All sessions were 5 min in duration.

*Data collection and interobserver agreement.* Data for all sessions were collected using handheld computers. Observers were undergraduate or graduate students in behavior analysis who had been trained to mastery criterion previously for data collection by attaining interobserver agreement scores of 90% or above for three consecutive sessions with a previously trained observer. Observers typically sat in the corner of the therapy room, about 1 m away from Helena and the therapist.

Trained observers collected data on student and therapist behavior. Student responses included off-task behavior, on-task behavior, and task completion. We defined *off-task behavior* as engaging in an alternative activity (not related to the task). Examples of off-task behavior included putting her head on her desk, playing with pencils or a pencil box, or drawing. We defined *on-task behavior* as having her pencil in her hand and her eyes oriented toward a worksheet. Both off-task and on-task behavior were scored as duration measures. Observers also collected frequency data on the number of tasks that Helena completed as a secondary measure of on-task behavior. In addition to these responses, observers collected data on

attention delivery from the therapist (the therapist looked at Helena and made a vocal statement) that was also scored as a duration measure.

A second observer simultaneously and independently collected data during 43% of reinforcer assessment sessions and 49% of treatment evaluation sessions. Interobserver agreement was calculated by dividing each session into 10-s bins and comparing the number of responses scored within each bin across observers by dividing the smaller number of responses in that bin by the larger number of responses and converting the ratio to a percentage (Shirley et al., 1997). The percentages were then averaged across all bins in the session to yield an overall percentage. During the reinforcer assessment, mean interobserver agreement was 100% for on-task behavior, 100% for off-task behavior, and 82% (range, 74% to 87%) for therapist attention. During the treatment evaluation, mean agreement was 92% (range, 73% to 100%) for on-task behavior, 94% (range, 80% to 100%) for off-task behavior, and 83% (range, 63% to 100%) for therapist attention.

Observers also collected data on treatment integrity, which we defined as the degree to which the therapist implemented the contingencies as designed. During the phases in which treatment integrity was manipulated deliberately, integrity scores were based on the therapist's adherence to a randomized computer output, as described below. Treatment integrity for all sessions was 100%.

*Reinforcer assessment.* The therapist conducted a reinforcer assessment to determine possible reinforcers for Helena's behavior. During the reinforcer assessment, the therapist provided Helena with two identical worksheets that differed only in that one had the word BREAK printed across the top, and the other had the word TALK printed across the top. Both worksheets were available concurrently throughout all reinforcer assessment sessions. When

Helena oriented toward the break worksheet, the therapist prompted her every 5 s to complete a task. Completion of a task on the break worksheet resulted in the removal of both worksheets for 15 s. When Helena oriented toward the talk worksheet, the therapist avoided eye contact with her and did not speak until she completed a task. Completion of a task on the talk worksheet resulted in eye contact and praise, lasting approximately 15 s. If Helena's behavior was influenced by escape, we expected her to either work on the break worksheet (to obtain programmed breaks) or to orient toward the talk worksheet but not complete any tasks (to maintain therapist silence and lack of prompts); in other words, either of these responses would result in the termination of either demands or therapist attention. If Helena's behavior was influenced by attention, we expected her to either complete tasks on the talk worksheet (to maintain access to attention in the form of praise) or to orient toward the break worksheet but not complete any tasks (to maintain access to attention in the form of prompts). Thus, both worksheets were associated with potential positive and negative reinforcement contingencies.

The therapist conducted forced exposure trials before the start of the reinforcer assessment. During the forced exposure, only one worksheet was available. The therapist vocally prompted Helena to complete a task on that worksheet and provided the associated consequence. The therapist provided one exposure to each consequence before the start of the reinforcer assessment.

Before the start of each session, the therapist instructed Helena that she could work on whichever worksheet she wanted and could switch between sheets at any time. We initially selected a double-digit addition task as the worksheet content. After four sessions using the double-digit addition task, we changed the task to writing vocabulary words to determine the generality of the reinforcer assessment findings

across tasks. Vocabulary tasks were used only for the reinforcer assessment; double-digit addition was used for the treatment integrity evaluation.

*Treatment integrity evaluation.* To make Helena's duration-based measures more similar to the rate measures in Experiment 1, we considered each consecutive 15 s that Helena spent on task or off task as a response. During baseline sessions, the therapist attended to Helena in each 15-s interval that she was off task and ignored her on-task behavior. Attention during baseline took the form of coaxing her to work, including saying statements like, "It's math time, Helena," "You have a math worksheet and a pencil," "Don't you think you should work on your math?," and "You won't get your math done if you keep playing." These statements were similar in form to the typical reaction of Helena's teachers when she engaged in off-task behavior in the classroom.

During DRA phases, the therapist attended to Helena when she was on task for 15 s and ignored off-task behavior. The form of attention during DRA was changed to praise statements and comments about Helena's work, like "great working," "You got that one right!," and "You're so smart." Again, these statements were similar to the praise typically delivered in Helena's classroom.

We evaluated four levels of combined omission and commission errors: 80%, 60%, 40%, and 20% integrity. These levels and the programming of the integrity failures were similar to Subset 3 of Experiment 1 (using RR schedules). For example, 80% of on-task behavior and 20% of off-task behavior resulted in attention during 80% integrity phases. During all integrity failure phases, the therapist held a clipboard with a computer-generated sequence of which responses should result in reinforcement. The form of attention was similar to that provided in baseline and DRA conditions (i.e., coaxing and praise following off- and on-task behavior, respectively). We counterbalanced the order of the integrity

failure phases across replications in a reversal design.

### *Results and Discussion*

Helena selected the talk worksheet during 96% of the reinforcer assessment (range, 70% to 100%; results available from the first author), suggesting that praise served as a reinforcer for on-task behavior. Anecdotally, Helena began scratching out BREAK at the top of the break worksheet and writing TALK on that worksheet before the third session. This did not affect the consequences that the therapist provided contingent on working on the two sheets, but provided additional evidence that attention was a reinforcer for Helena's behavior. She chose to work on the talk worksheet regardless of the specific task (double-digit addition or copying vocabulary words).

Figure 6 shows the results of Helena's treatment integrity analysis. The top panel shows the percentage of the session time that Helena spent on or off task. The bottom panel shows the rate at which she completed double-digit addition problems. She initially engaged in on-task behavior. Over the course of the baseline, however, on-task responding decreased and off-task responding increased; by the last two sessions in baseline, Helena was engaging exclusively in off-task behavior. During the first exposure to DRA implemented with full integrity, off-task behavior decreased and on-task behavior increased.

Helena's responding seemed to be affected by the degree of the treatment integrity failure during the integrity failure phases. She engaged in more on-task than off-task behavior during the 80% and 60% failure phases. However, she allocated more time to off-task behavior than on-task behavior during the 40% and 20% integrity conditions. These results replicate those obtained in Experiment 1, which also showed that DRA lost its efficacy when implemented at less than 50% integrity with combined omission and commission errors. The condition sequence did not influence Helena's behavior

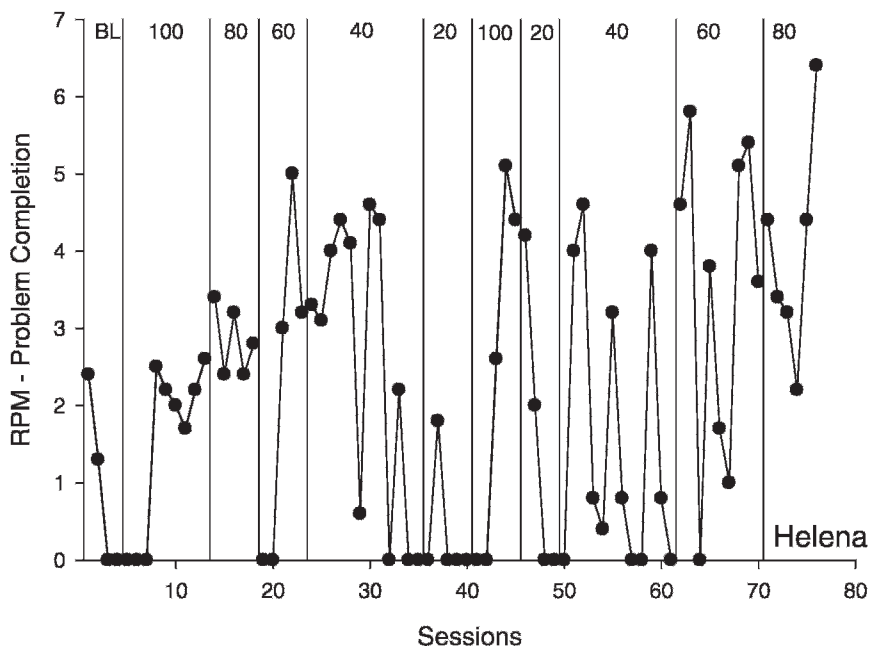
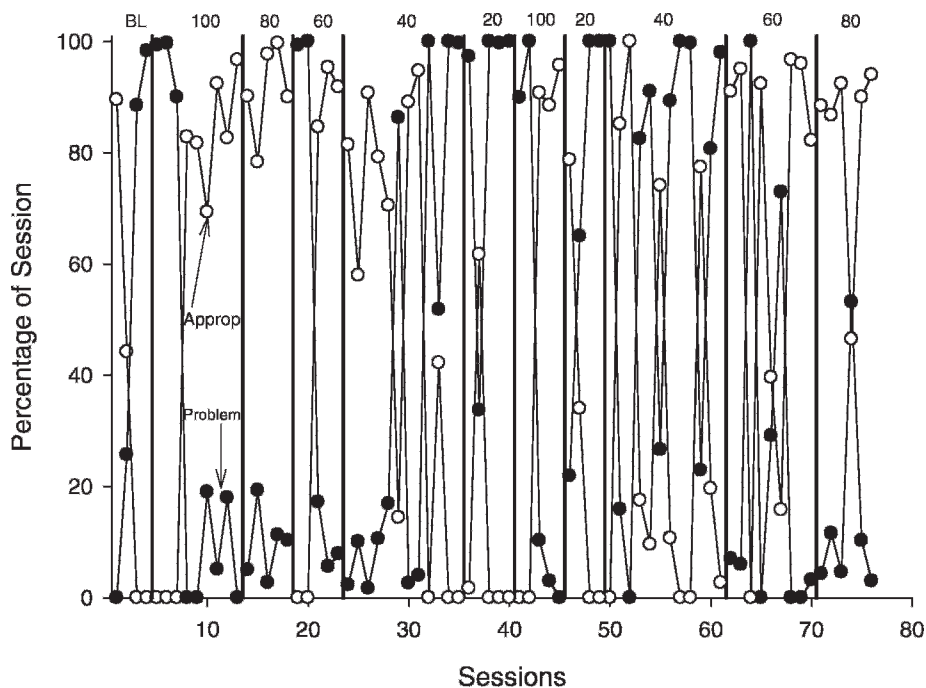


Figure 6. Treatment integrity assessment results for Helena. The top panel shows percentage of the session spent on or off task by the open and filled circles, respectively. Condition labels show baseline (BL), DRA, and treatment integrity failure phases. During integrity failures, condition labels show the treatment integrity percentage. The bottom panel shows the rate at which Helena completed math problems during the assessment.

strongly during the integrity failure phases, insofar as her behavior during the replications matched the results obtained from the initial exposures. Taken together, these results suggest that when the reinforcement schedules select for a particular response, the sequence of conditions is less important (as was the case with Helena); however, when the schedules do not select for a particular response (as in the case of 50% integrity from Experiment 1), sequence may become a key factor in response allocation.

The secondary measure of on-task behavior (the number of double-digit addition problems completed; Figure 6, bottom) supports the analysis based on time allocation. Helena completed a mean of 2.8 and 3.0 tasks per minute during the 80% and 60% integrity failure phases, respectively. When treatment integrity decreased to 40%, she completed a mean of 2.1 tasks per minute. When treatment integrity was 20%, she completed a mean of only 0.9 tasks per minute, equal to the mean rate of responding during baseline.

The results of Experiment 2 suggest that treatment integrity failures may be detrimental to DRA interventions if the integrity failures occur often. However, DRA seems to be resistant to detrimental effects of integrity failures as a whole, in that treatment integrity decreased to 40% before detrimental effects occurred. Yet the results from Subset 4 of Experiment 1 suggested that condition sequence may play an important role when DRA was implemented with 50% integrity. Thus, the purpose of Experiment 3 was to examine the possible influence of sequence effects on integrity failures by replicating the results of Subset 4 from Experiment 1 in a school setting with a child with developmental disabilities.

### EXPERIMENT 3: APPLIED EVALUATION OF SEQUENCE EFFECTS

#### *Method*

*Participant and setting.* Jake was an adolescent who had been classified as trainable mentally

handicapped; he was enrolled in a center school for children with disabilities. Jake's school administrator referred him to a school-based treatment program for the assessment and treatment of aggression (described below). Jake communicated using gestures or single-word vocal utterances and spent his school day in a special education classroom for students with moderate mental impairment. He was able to follow one-step directions and complete a variety of self-care tasks independently or with minimal prompting.

We conducted all sessions in an empty classroom (approximately 6 m by 10 m) in Jake's school. The classroom was equipped with tables, chairs, and leisure items appropriate for a variety of different ages and skill levels. We conducted sessions 2 to 4 days per week, with three to five sessions per day. All sessions were 5 min in duration.

*Data collection and interobserver agreement.* The data-collection system and interobserver agreement arrangements were identical to Experiment 2. Student responses included aggression (physical contact between Jake's open hand and the therapist's body) and greetings (i.e., saying, "hi"). Observers scored aggression and greetings as frequency measures. Therapist responses included attention delivery (the therapist looked at Jake and made a vocal statement) and escape delivery (the therapist said "take a break," removed instructional materials, and turned away from Jake).

A second observer simultaneously and independently collected data during 32% of functional analysis sessions and 44% of treatment evaluation sessions. Interobserver agreement was calculated as described for Experiment 2. During the functional analysis, mean interobserver agreement was 98% (range, 75% to 100%) for aggression, 100% for greetings, and 97% (range, 81% to 100%) for therapist behavior. During the treatment evaluation, mean agreement was 93% (range, 60% to 100%) for aggression, 92% (range, 64% to



100%) for greetings, and 84% (range, 72% to 95%) for therapist behavior. Treatment integrity, as defined for Experiment 2, was 100% for all sessions except Session 13, in which the therapist inadvertently reinforced one instance of problem behavior.

*Functional analysis.* We conducted a functional analysis to determine possible reinforcers for Jake's aggression. Initial functional analysis sessions were similar to those described by Iwata, Dorsey, Slifer, Bauman, and Richman (1982/1994) and consisted of play, attention, and escape conditions. During the play sessions, Jake had access to leisure items and continuous therapist attention. During attention sessions, he had access to leisure items, but the therapist ignored him until he engaged in aggression. Contingent on aggression, the therapist provided a brief reprimand, such as "Don't hit me, Jake." During escape sessions, the therapist asked Jake to engage in an academic task, such as letter sorting, using a three-prompt instructional sequence (verbal prompt, model, physical guidance, with 5 s between prompts). Contingent on aggression, the therapist allowed Jake to take a 30-s break from the task, signaled by the therapist saying "take a break" and removing the instructional materials.

When low or decreasing rates of aggression occurred in the initial sessions, additional conditions were included in the functional analysis. For example, Jake's teacher reported that he was most likely to engage in aggression when she was talking with another person or during transitions, when people were frequently in close physical proximity to Jake. Based on these reports and classroom observations, three additional functional analysis conditions were included: neutral attention, proximity, and diverted attention. The neutral attention condition was identical to the attention condition described above, except that the form of attention was changed from a reprimand to a neutral statement, such as "What's going on, Jake?" During the proximity condition, the

therapist sat within 0.5 m of Jake and moved away from him contingent on aggression. During the diverted attention condition, two therapists talked with each other. Contingent on aggression, one therapist ceased conversation with the other therapist, turned to Jake, and made a brief neutral comment, similar to those used in the neutral attention condition.

*Treatment integrity evaluation.* The baseline conditions were similar to the diverted attention functional analysis condition. Two therapists sat across the table from each other, with one therapist seated next to Jake (within 1 m). The therapists talked with each other and ignored Jake until he engaged in aggression. Contingent on each aggression response, the therapists stopped talking, and one therapist turned to Jake and made a brief neutral comment. The therapists did not reinforce greetings.

We conducted an evaluation of DRA after the conclusion of baseline. Before the first DRA session, the therapist prompted Jake to greet her by saying, "Jake, if you want to talk to me, say 'hi.'" When Jake said "hi," the therapist turned to him and made a brief neutral comment. The therapist repeated this prompting procedure 10 times before the start of the initial DRA phase. No additional prompts were used after these initial 10 trials. During DRA sessions, two therapists sat talking with each other and ignored aggression (extinction). When Jake said "hi," one therapist turned to him and made a brief neutral comment, such as "What are you looking at, Jake?"

We used a 50% integrity condition to examine the effects of sequence effects on responding during treatment integrity failures, which was similar to the treatment integrity failures experienced by Subset 4 of Experiment 1. During this condition, 50% of aggression responses and 50% of greetings resulted in brief attention, according to RR 2 schedules. Two therapists talked with each other and ignored Jake. The therapist seated across the table from Jake held a clipboard with a computer-generat-

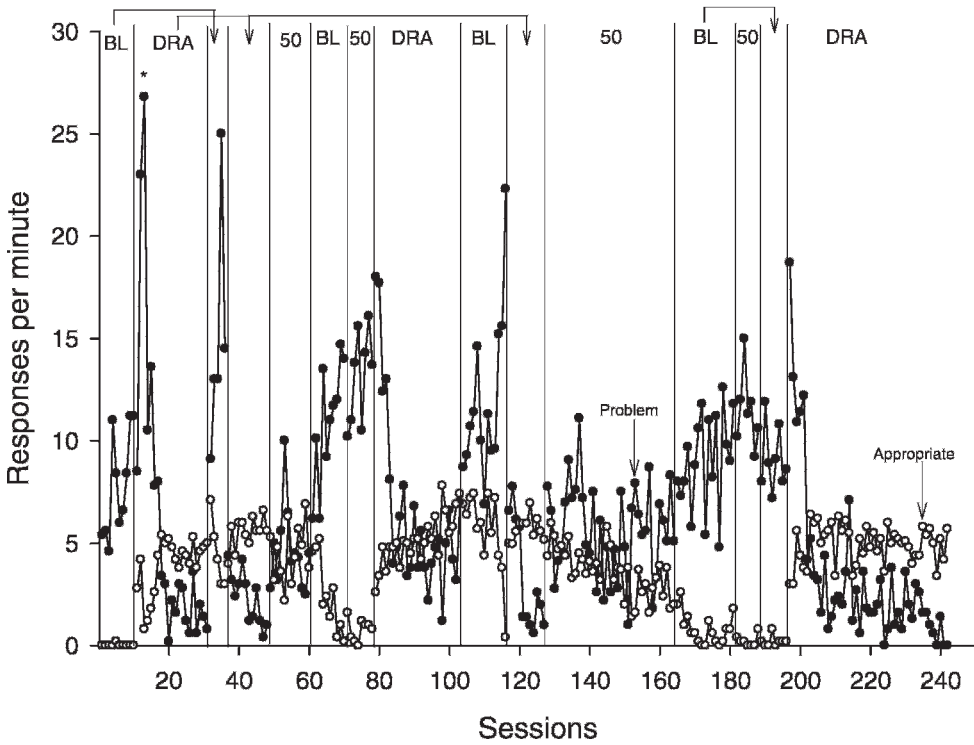
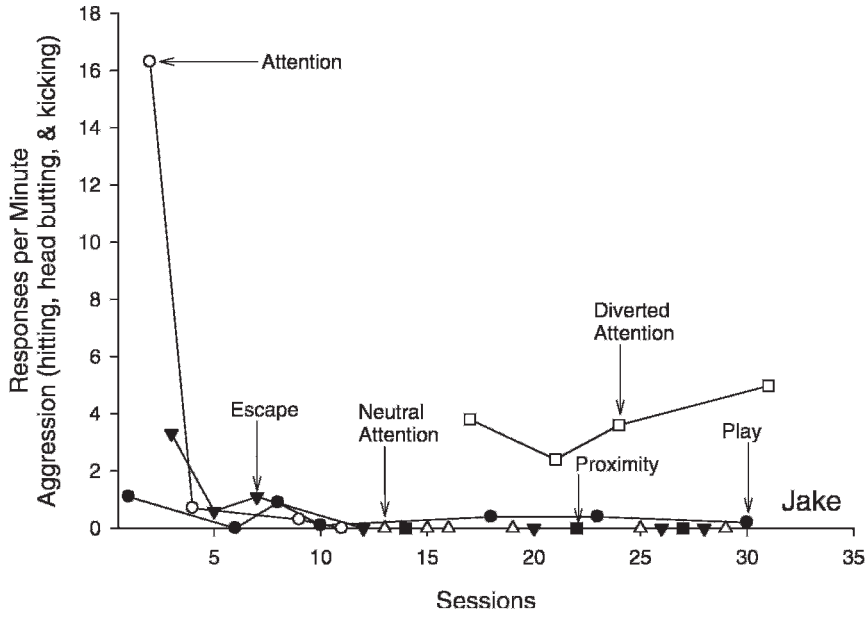


Figure 7. Results of the functional analysis (top) and treatment integrity evaluation (bottom) conducted with Jake. Responses per minute of Jake's behavior are shown along the y axis, with sessions shown on the x axis for both panels. In the bottom panel, condition labels show baseline (BL), DRA, and treatment integrity failure phases. During integrity failures, condition labels show the treatment integrity as a percentage. Filled circles denote rates of aggression, and open

ed sequence of which responses should result in reinforcement and cued the therapist sitting closest to Jake as to which response the therapist should reinforce. Reinforcers consisted of brief, neutral statements, similar to those provided in the baseline and DRA conditions. The 50% condition followed baseline twice and DRA twice to assess possible sequence effects using a reversal design.

### *Results and Discussion*

Figure 7 shows the results of Jake's functional analysis and treatment integrity evaluation. Although rates of aggression were elevated initially in the play, attention, and escape conditions, rates eventually decreased to near-zero levels across all functional analysis conditions (top). Following the introduction of the three additional test conditions, Jake engaged in elevated rates of aggression exclusively in the diverted attention condition, suggesting that adult attention served as a reinforcer for aggression, particularly when two adults were talking to each other but were ignoring Jake.

Figure 7 (bottom) shows the results of Jake's integrity analysis. During all baseline phases, rates of aggression occurred at moderate to high rates ( $M = 10.2$  responses per minute), and greetings occurred at low or decreasing rates ( $M = 2.2$  responses per minute). During the first exposure to DRA implemented with full integrity, rates of aggression increased initially before decreasing to low levels, and rates of greetings increased to moderate, stable levels. During subsequent DRA phases that followed baseline, aggression decreased to low rates ( $M = 4.4$  responses per minute across all DRA conditions), and appropriate behavior increased to moderate rates ( $M = 4.9$  responses per minute across all DRA conditions).

During the 50% integrity phases that followed DRA, a mixture of aggression and greetings occurred, with some bias toward aggression. Jake engaged in a mean rate of 5.4 aggression responses per minute and 3.9 greetings per minute during these two phases. This result contrasts with previous studies (e.g., Vollmer et al., 1999) that showed that DRA had relatively robust effects when implemented at reduced levels of integrity following DRA with full integrity. The bias toward problem behavior observed in the current experiment could be due to Jake's extraexperimental history (i.e., caregivers attending to aggression before experimental sessions). Unfortunately, we were unable to collect data to evaluate this possibility; future studies could investigate the extent to which extraexperimental histories influence responding during integrity failures.

The results in the 50% integrity conditions following DRA differ from the results of the 50% integrity conditions following baseline. During the integrity failures following baseline, rates of greetings remained low or near zero ( $M = 0.5$  responses per minute), and rates of aggression remained high and stable ( $M = 12.4$  responses per minute). For Jake, integrity failures (at least the 50% integrity condition used in the current experiment) were more detrimental to the treatment when they followed baseline than when they followed treatment with perfect integrity. These results suggest that the initial implementation of the treatment may affect responding during later integrity failures. Specifically, if caregivers initially implement DRA with a high level of integrity, later integrity failures may not be as detrimental. However, if a caregiver who previously provided a rich reinforcement schedule for problem behavior initially implements a

---

←

circles denote rates of greetings. The asterisk above Session 13 (bottom) denotes an accidental treatment integrity failure, in which the therapist reinforced one instance of aggression.

treatment with moderate or poor integrity, problem behavior is unlikely to improve, as in the 50% integrity conditions following baseline in the current experiment. These findings replicate those of Subset 4 in Experiment 1.

Exposures to treatment integrity failures also seemed to exert influence on a later implementation of DRA at 100% integrity. In particular, treatment effects during 100% integrity were more difficult to regain following a 50% integrity phase than following baseline. This effect can be seen in Sessions 79 to 102 (Figure 7, bottom). These results suggest that initial implementation of a treatment at reduced levels of integrity may impede treatment effects during later implementation of DRA with high levels of integrity. Evaluation of these effects was not a focus of this study, so we did not manipulate condition sequence in a way that permitted conclusions about the consistency of this effect. However, the effect of integrity failures on subsequent conditions should be evaluated in future research.

## GENERAL DISCUSSION

The present collection of experiments examined the effects of two types of treatment integrity failures (failure to reinforce appropriate behavior and reinforcing problem behavior, here termed omission and commission errors, respectively) on responding during DRA. We first investigated the effects of treatment integrity failures in a human operant procedure, in which participants responded on an analogue task (Experiment 1). Commission errors in isolation and omission and commission errors in combination resulted in increases in arbitrarily defined problem behavior and decreases in appropriate behavior. In addition, condition sequence played a role in the effects of integrity failure phases (as evidenced by results obtained from Subset 4). Next, we replicated the results of the combined errors phase with a student diagnosed with autism and with a student labeled as trainable mentally handicapped. In

both cases, the treatment-based replications reproduced results obtained in the human operant laboratory. Specifically, errors of commission and omission decreased treatment efficacy, and problem behavior was more likely to recur when treatment integrity failures occurred following baseline conditions than when otherwise identical integrity failures occurred following full treatment conditions.

Overall, DRA procedures seem to have relatively robust effects despite the occurrence of treatment integrity failures. Errors of omission in isolation did not affect treatment outcome detrimentally, regardless of the level of treatment integrity (Subset 1, Figure 1). However, participants engaged in somewhat lower rates of appropriate behavior during the first exposure to 40% and 20% integrity. This decrease in response rates as the density of the reinforcement schedule was reduced was somewhat unusual and should be examined further.

Detrimental effects of commission errors (Subset 2, Figure 2) and combined error types (Subset 3, Figure 3) occurred only when integrity levels decreased to 40% or below. When 50% integrity was evaluated (Subset 4, Figure 4), the detrimental effects were attenuated when the error phase followed a perfectly implemented DRA. The relatively robust effects suggest that DRA may yield favorable treatment outcomes even when implemented with decreased integrity.

The present experiments extend prior research by examining the effects of both omission and commission errors on differential reinforcement procedures and by providing an example of translational research spanning the basic-to-applied continuum. Prior research on treatment integrity failures in DRA treatments showed that the treatment generally was robust, but that sequence effects could be a problem (Vollmer *et al.*, 1999). In the current research, treatment integrity failures seemed to have fewer detrimental effects when the level of integrity descended (the sequence of integrity

failures from 80% to 20%) than when levels gradually ascended (the sequence of integrity failures from 20% to 80%). These effects can be seen with Participants 205, 222, 307, and 319. For each of these participants, rates of problem behavior were greater in the 40% and 20% integrity conditions during the ascending sequence than during the descending sequence. Unfortunately, the order of sequence presentation is confounded because we exposed all participants to the ascending sequence before the descending sequence. Results for participants in Subset 4 provide additional support for the idea that sequence plays a role in treatment integrity failure phases. In general, treatment integrity failures appeared to have less of an impact on treatment efficacy when failure phases followed DRA implemented with full integrity than when they followed baseline.

In addition to the results relating directly to treatment integrity failures, the current studies also have implications for the utility of human operant methods in examining problems of applied significance. In Experiment 1, we completed each participant's data set in just over 2 hr, resulting in rapid generation of complete data sets. Although limited by brief exposures to the contingencies, the results from the present experiments suggested which variables influenced responding and which do not. Using a similar approach, the most influential variables can be evaluated further using more extended exposures either in the human operant laboratory or in application (as in Experiments 2 and 3). The use of human operant methods in the current study informed Experiment 2 by suggesting that further information on the effects of combined errors, but not omission errors alone, would be likely to help inform treatment development. Thus, when replicating the results of Experiment 1, we focused on combined error types instead of omission errors. In addition, the reproducibility of the effects obtained in Experiment 1 permitted replication using only 1 participant during each of Experiments 2 and 3.

Finally, human operant procedures permit researchers to identify likely functional relations without the research constraints associated with applied contexts (e.g., time restrictions or difficulties associated with persistence or deliberate exacerbation of problem behavior). This freedom may allow researchers to identify rapidly which variables may affect applied outcomes and further explore the mechanisms responsible for behavior change during treatment procedures. The current experiments provide a model for this type of translational research method. In these experiments, we obtained similar results across both human operant and applied procedures. To illustrate, the results from Experiment 1 were used to inform subsequent applied replications, which focused only on variables that influenced behavior in the human operant laboratory. Similarly, Experiment 2 examined only combined treatment integrity failures (as opposed to the three different types of failures examined in Experiment 1), in that combined errors were detrimental to treatment effects during the human operant evaluation. In addition, the results of Experiment 1 suggested that sequence effects might be most pronounced during 50% integrity conditions. This result shaped the procedures for Subset 4 and for Experiment 3, both of which demonstrated effects of sequence during 50% integrity. In general, the human operant results allowed a more informed, targeted evaluation in clinical contexts than would have been otherwise possible.

There are limitations associated with the use of an analogue procedure in the human operant laboratory. First, Experiment 1 involved college students as participants, which may limit the generality of the results. Despite this limitation, the results of Experiment 1 seemed to generalize to participants with disabilities (based on the results obtained in Experiments 2 and 3). Second, the response chosen, a mouse click, was a simple response that existed in all participants' repertoires. Although we did not

specifically instruct participants to click the mouse buttons, we told them to “use only the mouse to earn as many points as possible.” The reinforcement contingencies in place, in conjunction with this instruction, frequently resulted in very high response rates, which often exceeded 200 responses per minute. These rates are well above those typically observed when treating problem behavior and may have influenced the outcome of the human operant experiments. This limitation could be addressed in future research by changing the form of the response to a topography that more closely approximates problem behavior (e.g., increasing the effort or duration of the target response).

Third, the duration and sequence of exposure to the contingencies in Experiment 1 may have influenced the outcome. We exposed participants to each phase for only 10 min, kept the sequence of conditions constant, and changed phases at set points in time, regardless of the pattern or rate of responding. Therefore, phase changes often occurred when behavior was on an upward or downward trend. Researchers should attempt to replicate these results using more extended exposures and conducting phases until responding stabilizes. Extending the exposures to the contingencies also may help clarify the relative lack of within-subject replication obtained during the human operant experiments; often, response rates or patterns obtained during the first exposure to a particular level of integrity failure were not replicated in a subsequent exposure. This lack of within-subject replication necessarily tempers the conclusions that can be drawn from the data. Increasing the duration of each phase and conducting phases until behavior stabilizes may increase the degree to which levels and allocations of behavior are reproduced across replications.

Other limitations are based on the specific procedures used in these experiments, independent of the human operant procedure. For example, the baseline and differential reinforcement schedules were restricted to only one set of

parameters: FR 1 and extinction schedules. Different parameters of the baseline and treatment schedules may have resulted in different patterns of behavior during the integrity failure phases. For example, intermittent baseline and DRA schedules may have altered the effects of treatment integrity failures. We chose to examine FR 1 schedules because of their frequent use in prior DRA research (e.g., Shirley *et al.*, 1997; Vollmer *et al.*, 1992, 1999; Walsh, 1991).

We examined only two types of possible errors in the current set of studies: errors of omission and commission that occurred according to probabilistic (RR) schedules. This frequently resulted in very high reinforcement rates during commission phases, which probably contributed to the detrimental effects of those errors. However, probabilistic schedules are only one way that errors of omission and commission might occur. These errors could also occur based on a variety of different other types of schedules, including interval-based schedules. The use of interval-based treatment integrity failures would limit the degree to which participants could maximize reinforcers by responding at high rates during errors of commission. In addition, errors of omission and commission are only two types of possible errors that may occur. Other errors might include differential delays to reinforcement, differential reinforcer magnitudes, or alternating periods of full treatment and no treatment. Future research should examine parametric variables associated with errors of omission and commission carefully, as well as different types of treatment integrity failures.

Despite the limitations, the results of the current experiments have important implications for application. First, the effects of errors on differential reinforcement procedures could be used to inform caregiver training. Currently, many caregiver-training procedures stress the importance of delivering reinforcers following appropriate behavior or following some speci-

fied period of time in which no instances of problem behavior occurred. However, the results of the current experiments suggest that accidentally failing to deliver an earned reinforcer (an error of omission) probably is not highly detrimental to the overall treatment effects. Caregiver training procedures may instead focus on the importance of the extinction component to differential reinforcement procedures. If caregivers do not implement this component with a high level of integrity (in other words, if they make commission errors), the treatment effects could be degraded. Unfortunately, extinction is not always possible, such as with dangerous behavior reinforced by social consequences. Thus, evaluations of differential reinforcement procedures that minimize the reinforcement of problem behavior, rather than eliminating it, are warranted.

Second, the results imply that initial monitoring of a caregiver's implementation of procedures may improve the chance that treatments later have robust effects. Monitoring and feedback during initial implementation could have two overall effects on the treatment. One, it could increase the overall levels of integrity with which the caregiver implements the procedure over time by providing a solid training foundation. Therefore, treatment integrity failures may never become an issue because the caregiver consistently implements the treatment with a high level of integrity. Two, monitoring may help ensure a high degree of treatment integrity during the initial implementation of the treatment. The results of Experiment 3 suggest that high levels of initial integrity may reduce the detrimental effects of subsequent treatment integrity failures. Unfortunately, this implies that caregiver-training procedures may be time intensive initially to ensure the best long-term outcome. Future research should examine the types and levels of treatment integrity failures made by caregivers following a variety of training procedures.

In conclusion, the methods used in the current studies seem to be useful for studying complicated questions like those related to treatment integrity. Across experiments, reduced levels of treatment integrity had detrimental effects on responding. The current studies may help inform future treatment integrity research, which could focus more on errors of commission and combined errors than errors of omission alone. In addition, more research is needed on the effects of treatment integrity failures on other types of behavioral treatments, such as noncontingent reinforcement procedures. Finally, the types and levels of integrity failures that occur in application should be examined through descriptive studies. The values obtained through these descriptive analyses then could be replicated and manipulated in the laboratory.

## REFERENCES

- Carr, E. G., & Durand, V. M. (1985). Reducing behavior problems through functional communication training. *Journal of Applied Behavior Analysis, 18*, 111–126.
- Codding, R. S., Feinberg, A. B., Dunn, E. K., & Pace, G. M. (2005). Effects of immediate performance feedback on implementation of behavior support plans. *Journal of Applied Behavior Analysis, 38*, 205–219.
- Hanley, G. P., Iwata, B. A., & Thompson, R. H. (2001). Reinforcement schedule thinning following treatment with functional communication training. *Journal of Applied Behavior Analysis, 34*, 17–38.
- Iwata, B. A., Dorsey, M. F., Slifer, K. J., Bauman, K. E., & Richman, G. S. (1994). Toward a functional analysis of self-injury. *Journal of Applied Behavior Analysis, 27*, 197–209. (Reprinted from *Analysis and Intervention in Developmental Disabilities, 2*, 3–20, 1982)
- Kelley, M. E., Lerman, D. C., & Van Camp, C. M. (2002). The effects of competing reinforcement schedules on the acquisition of functional communication. *Journal of Applied Behavior Analysis, 35*, 59–63.
- Lerman, D. C. (2003). From the laboratory to community application: Translational research in behavior analysis. *Journal of Applied Behavior Analysis, 36*, 415–419.
- Mazaleski, J. L., Iwata, B. A., Vollmer, T. R., Zarcone, J. R., & Smith, R. G. (1993). Analysis of the reinforcement and extinction components in DRO contingencies with self-injury. *Journal of Applied Behavior Analysis, 26*, 143–156.

- Mueller, M. M., Piazza, C. C., Moore, J. W., Kelley, M. E., Bethke, S. A., Pruett, A. E., et al. (2003). Training parents to implement pediatric feeding protocols. *Journal of Applied Behavior Analysis, 36*, 545–562.
- Northup, J., Fisher, W., Kahng, S. W., Harrell, R., & Kurtz, P. (1997). An assessment of the necessary strength of behavioral treatments for severe behavior problems. *Journal of Developmental and Physical Disabilities, 9*, 1–16.
- Shirley, M. J., Iwata, B. A., Kahng, S. W., Mazaleski, J. L., & Lerman, D. C. (1997). Does functional communication training compete with ongoing contingencies of reinforcement? An analysis during response acquisition and maintenance. *Journal of Applied Behavior Analysis, 30*, 93–104.
- Tiger, J. H., Hanley, G. P., & Bruzek, J. (2008). Functional communication training: A review and practical guide. *Behavior Analysis in Practice, 1*, 16–23.
- Volkert, V. M., Lerman, D. C., Call, N. A., & Trosclair-Lasserre, N. (2009). An evaluation of resurgence during treatment with functional communication training. *Journal of Applied Behavior Analysis, 42*, 145–160.
- Vollmer, T. R., & Iwata, B. A. (1992). Differential reinforcement as treatment for behavior disorders—procedural and functional variations. *Research in Developmental Disabilities, 13*, 393–417.
- Vollmer, T. R., Iwata, B. A., Smith, R. G., & Rodgers, T. A. (1992). Reduction of multiple aberrant behaviors and concurrent development of self-care skills with differential reinforcement. *Research in Developmental Disabilities, 13*, 287–299.
- Vollmer, T. R., Roane, H. S., Ringdahl, J. E., & Marcus, B. A. (1999). Evaluating treatment challenges with differential reinforcement of alternative behavior. *Journal of Applied Behavior Analysis, 32*, 9–23.
- Walsh, P. (1991). The use of differential reinforcement in affecting multiple behavioural change in a woman with severe mental handicap. *Irish Journal of Psychology, 12*, 382–392.
- Worsdell, A. S., Iwata, B. A., Hanley, G. P., Thompson, R. H., & Kahng, S. W. (2000). Effects of continuous and intermittent reinforcement for problem behavior during functional communication training. *Journal of Applied Behavior Analysis, 33*, 167–179.

*Received March 17, 2008*

*Final acceptance June 25, 2009*

*Action Editor, Henry S. Roane*