An Evaluation of Probabilistic Delay and Denial Training: Durability Against Treatment Relapse

Kendall Mae Kastner
Marquette University

Follow this and additional works at: https://epublications.marquette.edu/dissertations_mu

Recommended Citation
Kastner, Kendall Mae, "An Evaluation of Probabilistic Delay and Denial Training: Durability Against Treatment Relapse" (2024). Dissertations (1934 -). 3227.
https://epublications.marquette.edu/dissertations_mu/3227
AN EVALUATION OF PROBABLISTIC DELAY AND DENIAL TRAINING: DURABILITY AGAINST TREATMENT RELAPSE

by

Kendall M. Kastner, M.S., BCBA

A Dissertation submitted to the Faculty of the Graduate School, Marquette University, in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

Milwaukee, Wisconsin

May 2024
Delay and Denial Teaching is a prevalent prescribed intervention following successful implementation of Functional Communication Training (FCT). Although there is strong empirical evidence in support of delay and denial teaching as an intervention to reach socially acceptable outcomes, there has yet to be systematic assessments of this intervention’s susceptibility to treatment relapse.

The goal of this evaluation is to arrange treatment challenges which mimic the conditions that threaten treatment outcomes in the natural environment, specifically transferring treatment to a new environment or treatment integrity omission errors. Researchers replicated recent iterations of delay and denial teaching procedures (Hanley et al., 2014; Rose & Beaulieu, 2019), and included treatment challenges based on procedures by Saini et al., (2018), and Fisher et al., (2019).
Introduction

In the treatment of socially mediated problem behavior, Functional Communication Training (FCT; Carr & Durand, 1985) has been widely demonstrated as an efficacious, function-based treatment when combined with extinction (Fisher et al., 2000; Hagopian et al., 1998; Tiger et al., 2008, Wu et al., 2022). However, common problems arise when attempting to integrate FCT procedures into the individual’s daily life. For example, after successfully teaching a functional communication response (FCR) as a replacement for problem behavior, levels of engagement in the FCR may occur too often for it to be sensibly reinforced (i.e., manding during nonreinforcement periods, or when it is impractical to do so). There is also a possibility that a resumption in problem behavior occurs when the FCR is no longer continuously reinforced (Briggs et al., 2018; Muething et al., 2021). Therefore, while providing immediate reinforcement for the FCR is suitable for initial treatment, the schedule of reinforcement should then be thinned to resemble the natural environment more closely (Durand & Moskowitz, 2015, Tiger et al., 2008).

The process of formal programming for generalization and maintenance by exposing individuals to worsening conditions, specifically a less dense reinforcement schedule within their natural environment is termed schedule thinning. In general, these recommendations involve arranging procedures to systematically lean a reinforcement schedule until a socially valid and practical terminal schedule is reached (Kranak & Brown, 2023). In sum, the purpose of schedule thinning is to maintain the strength of the FCR while also decreasing the overall rate of manding, and to prevent a relapse in problem behavior during extinction periods. Several arrangements have been utilized in the literature to facilitate successful reinforcer schedule thinning; including multiple schedules (Betz et al., 2013.; Fisher et al., 2015; Hanley et al., 2001; Saini et al., 2016),
chained schedules (e.g., Hagopian et al., 2011; Lalli et al., 1995), response restriction (Campos et al., 2017), and delay schedules or probabilistic delay procedures (Fisher et al., 1993; Fisher et al., 2000; Ghaemmaghami et al., 2016; Hanley et al., 2014; Rose & Beaulieu, 2019).

Even with the best practices in assessment and treatment of problem behavior, one or more form of treatment relapse can often occur during the schedule thinning process (Briggs et al., 2018). Treatment relapse describes the phenomena of a resumption of behavior after having previously extinguished or minimized its occurrence (Wathen & Podlesnik, 2018). Specifically, the relapse of problem behavior can be dangerous and lead to various undesirable consequences and impact the long-term maintenance of treatment effects. The overall durability of treatment packages against treatment relapse is an ever-growing line of research within the context of severe problem behavior. To successfully transition clients out of behavioral services while maintaining low levels of problem behavior, it is essential to know the extent to which the obtained treatment effects will persist through various forms of treatment challenges in the natural environment. Treatment challenges can be described as less than idyllic conditions that an individual encounters in their environment which increases the risk for behavioral relapse. To account for these environmental threats to treatment the goal is to increase the efficacy of interventions and move toward behavioral inoculation, or to decrease the prevalence of treatment relapse through the systematic and proactive programming of stimulus and environmental changes (Kimball et al., 2023).

There are several types of treatment relapse, renewal describes a return of problem behavior, or the reemergence of a previously extinguished behavior following a context change. Context changes in the literature have been used to describe a shift in environments such as a setting change (e.g., clinics, schools, homes). With a consecutive case-series analysis, Muething
et al. (2020) demonstrated instances of renewal in 42% of the 182 context changes reviewed. Behavioral relapse often occurs as clinicians transfer treatments which have been successful in a clinical context to the natural environment such as at home, school, or to the community. Saini et al. (2018) utilized applied a three-phase arrangement to assess renewal following successful treatment effects. That is, researchers initially assessed levels of problem behavior in the home with the caregiver (A), implemented FCT treatment in a clinic with trained clinicians (B), and subsequently assessed renewal during FCT treatment in the home with trained caregivers (A). In sum, Saini et al. found operant renewal to occur in a majority of participants (75%) when only the stimulus context changed (i.e., change from FCT in clinic with clinicians to FCT in home with caregivers). These findings suggest that although common and empirically sound prescribed treatments can be robust in their effects in the trained environment, changes to only the stimulus context can still pose challenges resulting in a resumption of problem behavior.

Additionally, treatment relapse in the form of resurgence can occur when a behavior is extinguished in the presence of alternative reinforcement and remerges during suspension of the alternative reinforcement (Epstein, 1985), or under a worsening of reinforcement conditions (e.g., schedule thinning; Kimball et al., 2023). Typically, tests for resurgence of behavior have also been studied and completed in three phases, where Phase 1 describes conditions where reinforcement is provided for the target behavior (e.g., reinforcement of problem behavior during baseline). In Phase 2 a target behavior is placed on extinction, and a new alternative behavior contacts reinforcement (e.g., in an FCT arrangement placing problem behavior on extinction and reinforcing an FCR). Last, during Phase 3 the target and alternative behaviors either both contact extinction, or reinforcement for the alternative response is worsened. Resurgence is then said to be demonstrated by a resumption of target behavior during Phase 3 although both responses
contact extinction or a decrease in reinforcement. The conditions where resurgence may occur includes common treatment integrity omission errors (i.e., failure to deliver reinforcement for an alternative response). In the natural environment the individual may also encounter unplanned and extended extinction-like conditions if an untrained caregiver fails to respond to the newly trained response. Resurgence may also occur during schedule thinning when the schedule for reinforcement becomes leaner (a worsening of reinforcement conditions). Each time the schedule requirement changes, learners experience brief periods of extinction-like conditions. (Briggs et al., 2018).

There are several promising procedures that have been used to promote mitigation of behavioral resurgence during schedule thinning. Some of which includes manipulations to the schedules of reinforcement during each phase of the treatment comparison (e.g., reinforcement magnitudes during baseline or treatment; Fisher et al., 2019; Nevin et al., 2016), operant-expansions (Neely et al., 2020), and discrimination training (Betz et al., 2013; Fisher et al., 2020; Fuhrman et al., 2016). Most notably, researchers are evaluating the extent to which resurgence occurs during schedule thinning with a multiple schedules arrangement (Fisher et al., 2020, Fuhrman et al., 2016). A multiple schedule is an evidence-based procedure which involves the presentation of two or more schedule components with discriminative stimuli, and typically the reinforcement component signals that the alternative response produces immediate reinforcement for every response (see Saini et al., 2016, for a review). The use of multiple schedules following FCT for schedule thinning is done by manipulating the durations of the two components and serves as a form of discrimination training. For example, Hanley et al. (2001) successfully utilized a multiple schedule for destructive behavior and thinned the overall schedule to be practical for the individual’s natural environment. Specifically, after initially teaching an FCR,
they gradually increased each component to reach a terminal schedule of a 1-min reinforcement period which alternated with a 4-min extinction period while maintaining low levels of problem behavior. In another study, Betz et al. (2013) arranged a component analysis of signaled and un-sighed schedules during schedule thinning. It was found that programming schedule-correlated stimuli to signal reinforcement and extinction components was more effective at maintaining behavior reduction than non-sighed components. These advancements in the understanding of the utility of multiple schedules have informed procedural arrangements for schedule thinning and established the importance of schedule correlated stimuli.

Another increasingly more popular evidence-based methodology for schedule thinning is both termed Probabilistic Delay Thinning, or Delay and Denial Training. Similar to multiple schedules, individuals are first taught FCRs with continuous reinforcement. Subsequently, clinicians systematically introduce delays to reinforcement following a target communication response. Clinicians will prescribe the use of probabilistic delay thinning for several reasons, one being that contingencies closely emulate natural environments (i.e., caregivers denying access to an item or interaction and telling the child to wait) at the onset of intervention, another rationale is the programmed exposure to unpredictability, and also the ability to program teaching alternative behaviors during the delays (Ghaemmaghami et al., 2016). For example, Hanley et al. (2014) initially taught a simple FCR and subsequently shaped the FCR into a more complex form (i.e., a FCR with an autoclitic frame). Once the complex FCR occurred at high stable levels with minimal occurrence of problem behavior, researchers implemented probabilistic delay thinning where 60% of the FCRs contacted a delay cue 40% contacted immediate reinforcement. The delays were systematically increased until a terminal delay. This procedure resulted in high levels of FCRs and low levels of problem behavior across participants thus strengthening delay
tolerance and contextually appropriate behavior. Within schedule thinning a tolerance for delays to reinforcement, or delay tolerance, can be conceptualized as an absence of manding at high levels, and near-zero levels of problem behavior during delay periods (Ghaemmaghami et al., 2016).

There is evidence of the efficacy of probabilistic delay and denial training, or hereby referred to as probabilistic delay teaching, to decrease the overall amount of reinforcement, maintain lower levels of mands and low levels of problem behavior (Ghaemmaghami et al., 2016; Hanley, et al., 2014; Rose & Beaulieu, 2019). New prevalence data captured by Mitteer et al. (2024), indicated that credentialed Board-Certified Behavior Analysts (BCBAs) and Board-Certified Behavior Analysts-Doctoral (BCBA-D) are most commonly selecting delay and denial teaching as their schedule thinning intervention. However, less is known about the durability of treatment effects during common treatment challenges. Said differently, it is unknown the prevalence of treatment relapse following probabilistic delay training. Some research exists on the topic area including a study by Rose and Beaulieu (2019) where they taught a simple FCR, shaped the FCR to a complex FCR (i.e., “Excuse me” “May I have the _____, please” ), and subsequently introduced probabilistic delay teaching. Researchers increased the delays to reinforcement systematically as individuals demonstrated mastery at each step of schedule thinning while requiring participants to engage with an alternative activity.

Rose and Beaulieu (2019) conducted additional challenges to treatment at the conclusion of the delay teaching for one of their participants. First researchers exposed one of their participants to two 20-min “denial probe” sessions. Although researchers did not describe these as tests for behavioral resurgence, the denial probe sessions share characteristics of a test for behavioral resurgence (e.g., programmed extinction for alternative and target behaviors). During
the denial probe researchers provided a less preferred alternative activity and did not provide access to the requested item at any point after. It has been established within the literature that providing alternative functional reinforcers (Austin & Tiger, 2015; Driftke et. al, 2020) and alternative items or activities (Ghaemaghami et. al, 2016; Muharib et. al, 2022) during delays can mitigate a return of problem behavior but, the presence and absence of less preferred items has not been directly evaluated in the context of delay and denial procedures. Rose and Beaulieu also conducted “treatment extensions,” where procedures remained the same however, sessions were conducted with researchers in a different room in the individual’s house (novel with no history of teaching). Further, sessions were conducted with the caregiver implementing treatment in this novel room. Although researchers did not call these procedures renewal tests, the context changes loosely resemble conditions which might occasion behavioral renewal (i.e., exposing individuals to experimental Phase 3, or novel context). Nevertheless, responding characteristic of resurgence or renewal was not observed during these evaluations.

Although, a methodology for evaluating treatment relapse has been established in the context of multiple schedules, no applied studies have systematically studied renewal and resurgence as a part of delay and denial teaching treatment packages. Thus, the purpose of this study is to replicate the probabilistic delay and denial-teaching procedures described by Hanley et al. (2014) and Rose and Beaulieu (2019) and extend these procedures by programming post intervention treatment challenges. That is, we replicated the procedures described in previous literature for assessing renewal and resurgence of socially maintained problem behavior to determine the extent to which relapse was present following probabilistic delay teaching (Saini et al., 2018; Fisher et al., 2019). Additionally, researchers were interested in the mitigation effects of programming access to alternative items during denial conditions.
Method

Participants, Setting, and Materials

Researchers recruited participants from a university-based clinic serving individuals referred for treatment of problem behavior. All participants were referred for various concerns which included increasing social skills, communication skills, and problem behavior reduction. Participants received clinical services five, three, two, and three days per week for Noah, David, Alabama, and Mario, respectively. During a two-to-four-hour appointment experimental sessions were conducted for a minimum of half the appointment and maximum the entire daily appointment. Researchers included the first four participants admitted into the clinic who met inclusion criteria. Inclusion criteria were that the participant (1) was between the ages of 4-17, (2) identified functions responsible for maintenance of problem behavior were socially mediated, (3) caregivers identified delay tolerance as a clinical goal, (4) had an echoic repertoire, and (5) were not visually or hearing impaired to the point where that may interfere with abilities to participate.

Research sessions were conducted in one of two clinical 3 x 3 m therapy rooms with a one-way mirror, intercom system, and furniture (e.g., tables and chairs). For the purpose of this study two types of therapy rooms, hereby referred to as contexts 1 & 2, were utilized. That is, Context A describes a therapy room located on a separate floor of the clinic, which had all walls painted orange and orange stimuli (e.g., paper, therapist shirt). The second therapy room, or Context B was located within the typical clinic setting on a floor different from Context A. Context A and Context B were otherwise identical aside from location in the building, color of the walls, and some of the stimuli. Sessions were all conducted by graduate students and overseen by a BCBA-D.
Noah was a 9-year-old, Latinx male diagnosed with autism spectrum disorder (ASD) who was referred for the assessment and treatment of aggressive and destructive behaviors. His problem behaviors included aggression in the form of hitting and kicking, and property destruction. He communicated vocally with single word mands, tacts, by drawing pictures, and had a limited echoic repertoire. David was a 6-year-old Latinx male diagnosed with ASD who was referred for the assessment and treatment of aggressive and destructive behavior. David’s problem behavior included aggression in the form of hitting and pinching, and property destruction. He communicated with mands, tacts, and gestures and had an echoic repertoire. For both Noah and David Spanish was spoken primarily at home, and English was spoken primarily at school and in the clinic. Alabama was a 7-year-old, Caucasian male who was diagnosed with ASD; and was referred for the assessment of aggressive behavior. Alabama engaged in aggressive problem behaviors in the forms of hitting, and kicking and he had an extensive vocal verbal repertoire where he spoke in full sentences. Mario was a 9-year-old Caucasian male diagnosed with ASD referred for the assessment and treatment of aggressive and destructive behaviors including hitting, kicking, biting and property destruction as well as the ingestion of non-food items (pica). Mario had an extensive vocal verbal repertoire and spoke in full sentences.

**Response Measurement, Interobserver Agreement**

Trained observers independently collected data on laptop computers behind an observation window. Observers scored each occurrence of problem behavior including aggression, and property destruction. Aggression included hitting, kicking, pushing, pinching, scratching, or throwing objects at the therapist. Property destruction included hitting or kicking furniture or the walls, throwing objects, tearing one’s own clothing, swiping materials, and
turning over furniture. Researchers combined each topography into a single measure of problematic behavior and converted the occurrence into a response rate by dividing the total number instances of behavior by the session duration (e.g., 20 instances of aggression in a 10-min session, 20/10=2.0).

Observers also scored the occurrence of complex FCRs (FCRcomplex), simple FCRs (FCRsimple), tolerance responses, and the duration of toy engagement. Functional communication responses were varied based on each participants identified functional reinforcer (See Table 1). Tolerance responses were defined as a child independently saying, “Okay” within 5 s of a therapist delivered delay cue. Toy engagement was defined as independent manipulation with an alternative activity. The percentage of session engagement was equal to the total duration in second of engagement divided by the session duration and multiplied by 100.

A second independent observer collected data simultaneously or asynchronously from video recordings of sessions 32%, 38%, 30%, and 40% of sessions for Noah, David, Alabama, and Mario’s sessions, respectively. Observers’ scoring records from each session were divided into 10-s intervals and were compared on an interval-by-interval basis using the proportional-agreement method. Intervals in exact agreement received a score of 1. Intervals not in exact agreement received a score by dividing the smaller number of scored responses by the larger number of responses. The sum of all interval scores were then divided by the total number of intervals, and the resulting quotient was converted into a percentage. This resulted in mean agreement scores of 94% for problem behavior (range, 82% to 100%), 99% complex FCRs (range, 90% to 100%), 93% for tolerance responses (range, 85% to 100%), and 93% for engagement (range, 82% to 100%) for Noah; 98% for problem behavior (range, 90% to 100%), 99% agreement for complex FCRS (range, 93% to 100%), 92% for tolerance responses (range,
87% to 100%), and 93% for engagement (range, 79% to 100%) for David; 99% for problem behavior (range, 92% to 100%), 98% for complex FCRs (range, 85% to 100%) and 98% for tolerance responses (range, 89% to 100%), 99% (range, 91% to 100) for engagement for Alabama; and 95% for problem behavior (range, 90% to 100%), 99% complex FCRs (range, 97% to 100%), 98% for tolerance responses (range, 95% to 100%), and 97% for engagement (range, 92% to 100%) for Mario.

**Procedural Integrity**

To ensure the independent variables were implemented by the session therapist as programmed we collected data synchronously or via video recordings across phases and for at least 30% of sessions for each participant from FCT, probabilistic delay and denial training, and the treatment challenges. Data were collected on implementation of extinction for problem behavior, FCR prompts, and delivery of the delay cue, programmed delay interval duration and reinforcement after programmed engagement requirement. From those measures we calculated a mean integrity value for each participant. Mean procedural integrity data were 99% (range 89%-100%) for all of Noah’s sessions, 98% (range 85%-100%) for David, 98% (range 87%-100%) for Alabama, and 99% (range 90%-100%) for Mario.

**Preference Assessment**

Researchers conducted a Multiple Stimulus Without Replacement (MSWO; DeLeon & Iwata, 1996) with caregiver-nominated items to identify the highest preferred items (e.g., iPad, blocks etc.). That is, researchers interviewed caregivers to select toys to include in an array for participants to make selections. The results of this procedure identified high, moderate and low-preferred items to deliver during FCT and delay tolerance training. The highest preferred items were the items most often selected during the preference assessments, and lowest were those
selected least often. Researchers sorted six least selected items and identified those as low-preferred items for inclusion in the toys available during delays (Figure 1).

**Functional Analysis**

Researchers conducted a functional analysis of problem behavior with each participant, based on semi-structured interviews with caregivers and the procedures of Iwata, Dorsey, et al., (1994a) with each participant. All test conditions were based initially on the most common variables responsible for the maintenance of problem behavior (Iwata et al., 1994b), meaning each functional analysis included tests for attention, escape from demands, and access to tangible items.

**Initial Functional Communication Training Evaluation**

**Experimental Design.** Researchers utilized an ABAB reversal design. All sessions were 10 mins in length. First, researchers established a baseline of problem behavior. Researchers began sessions by disrupting reinforcement access and reinforced problem behavior on a FR1 schedule. Following stable baseline responding, researchers then implemented FCT (B phase). During the B phase researchers taught a simple FCR as a replacement to target problem behavior (problem behavior was placed on extinction) and continued until stable responding was observed. Researchers then conducted a reversal to baseline to demonstrate the treatment effects of FCT (A phase). Once researchers observed stable responding, a return to FCT was completed to replicate the treatment effect (B).

At any point during participant engagement in any experimental sessions, if problem behavior occurred at dangerous or destructive levels and participant-specific criteria for session termination were met (e.g., 15 instances of head-directed self-injury within a 10-min session), persistent emotional behavior (e.g., crying for a period lasting over 2 mins, and any direct mands
to stop (e.g., “I don’t want to do this anymore”) sessions would end, and researchers would wait until problem behavior was absent for at least 3 min before resuming participation. This procedure was not implemented as these criteria were not met by any of the participants.

**Baseline.** The baseline conditions of the initial evaluation are identical to participant specific evocative condition identified from the functional analysis (A phase). Sessions during this phase were 10 min in length and any instance of target problem behavior was reinforced on a FR-1 schedule. Researchers delivered reinforcement for 30 s during tangible or escape test sessions, and for attention sessions a brief reprimand was delivered contingent on problem behavior. Researchers conducted baseline session in the *orange* context, or Context A. Context A for Noah (SR-) was arranged to include a novel therapist wearing a blue shirt with blue materials.

**Pre-Teaching.** Researchers replicated the procedures in Rose and Beaulieu (2019) and Hanley et al. (2014) when teaching increasingly more complex FCRs. Researchers taught any new components of the target FCR trials prior to returning to the session format. Researchers started with simple FCRs (FCR$_{\text{simple}}$) and progressed to the complex FCRs (FCR$_{\text{complex}}$), which included tolerance responses. Said differently, after mastery at a previous step researchers introduced each new teaching step in a trial-based teaching format utilizing behavior skills training (BST). Researchers utilized BST by describing the target FCR, modeling the expected target behavior, and role-playing with the participant until the participant engaged in an independent response in a trial-based format. These procedures were used for each step of the FCR shaping procedures. Noah, David, Alabama, and Mario required approximately 2-6 trials before an independent response were observed for each step. The researcher conducted all sessions in Context B (i.e., the context associated with a standard therapy room). For Noah (SR-)
a therapist wearing pink colored shirt and pink materials was associated with all teaching sessions.

**FCT.** Based on the procedures of Rose and Beaulieu (2019) researchers began each session by disrupting reinforcement access (e.g., removal of preferred toy or attention). If the participant did not emit the FCR independently, researchers delivered an echoic prompt after 5 s; problem behavior resulted in no programmed consequences (i.e., extinction). Following the FCR, the researcher delivered the reinforcer for 30 s. After 30 s elapsed, the researcher would disrupt reinforcement access (e.g., say, “My turn,” and removed access to the preferred item). Researchers replicated the FCR teaching steps of those used by Ghaemmaghami et al., (2016, Table 1). Researchers begun with the initial step, the FCRsimple (e.g., “iPad, please”), and systematically shaped the FCR in a stepwise manner across five steps to the terminal FCRcomplex (e.g., emitting the response: “Excuse me,” tolerating a wait period where they were to wait for up to 3-s for the adult to respond and say “yes” and then saying, “May I have the iPad, please”; Step 4).

During FCT for the social negative evaluation for Noah (SR-) we systematically shaped the FCRsimple (e.g., “Break, please”) to the FCRcomplex (e.g., emitting the response: “Excuse me,”, tolerating a wait period where he waited for up to 3-s for the adult to respond and say “yes” and then saying “Can I have a break, please”). We introduced the teaching of each new step once participants engaged in the target FCR independently in 90% of opportunities and engaged in a 90% reduction in problem behavior across three sessions for the previous step. Mastery for FCT was defined as engagement in the independent FCRcomplex at 90% independence, and a 90% reduction in problem behavior from baseline levels.

**Probabilistic Delay Tolerance Teaching**
**Pre-Teaching.** Similar to the procedures used for FCT, we used BST to expose each participant to the participant-specific tolerance response (e.g., “When I say, ‘Not right now,’ you can let me know you heard me and say, “Okay’”). Before experiencing any delay to reinforcement researchers modeled the expected target behavior and role played in a trial-based format until independent participant engagement in the tolerance response occurred. Noah, David, Alabama, and Mario required approximately 1-6 trials before independent engagement in tolerance responses occurred.

**Delay Tolerance Teaching.** During delay tolerance session-based teaching, the participant was required to emit the FCR complex and a tolerance response (i.e., “Okay”) following a delay or denial cue (e.g., “Not right now”, “Later”, or “Wait”). Identical to Rose and Beaulieu (2019), and Hanley et al. (2014), a session consisted of five practice opportunities with the evocative context. Two of every five FCR complex produced immediate reinforcement, and three of every five FCR complex produced a delay to reinforcement.

Prior to the onset of delay tolerance teaching sessions, researchers randomized the presentation order of immediate or delayed reinforcement trials (e.g., delay, immediate, immediate, delay, delay), the programmed delay duration (e.g., 0 s, 30 s, or 45 s), the delay cue (e.g., “Wait,” “Not right now,” and “Later”), and the alternative toys or activities available (e.g., Group 1: coloring materials, cars, Group 2: sand, Jenga, Group 3: blocks, music toys). For Noah (SR-) evaluation related delay cues (e.g., “Time to work”, “Work time”, and “Work first”) and previously mastered work tasks (conditional discriminations of clothing and food, tracing letters and shapes, and receptive identification and “search and find”) were randomized in the same fashion. For example, an FCR might result in a 45 s delay to reinforcement following the researcher-delivered delay cue, “Later.” These variables were rotated in a quasi-random and
counterbalanced fashion, and once a variable was selected it was not selected again until the next session. Multiple delay cues were programmed to offer a variety and so that stimulus control was not associated with only one cue. To facilitate this, prior to the start of a session, the first author randomized each variable and identified and denoted each component for each trial with an in-room datasheet (see supplemental materials). Consistent with Rose and Beaulieu (2019) the delay was measured by a timer, and the delay time paused during any instances of problem behavior, repeated FCRs, and extended if problem behavior occurred in the last 5 s of the delay (i.e., momentary Differential Reinforcement of Other Behavior; mDRO 5s). If the participant engaged in repeated FCRs researchers re-presented the delay cue once and provided no consequences for any additional FCRs.

Termination of the delay period was based on activity engagement and the absence of problem behavior (Drifke et al., 2020; Ghaemmaghami et al., 2016; Hanley et al., 2014; Rose & Beaulieu, 2019). The delay progression is depicted by Table 2, with five delay progression steps centered around a mean delay. For example, for Delay Step 3, the programmed delays that were randomized and presented in those sessions include 30 s, 60 s, and 90 s. Consistent to Rose and Beaulieu (2019), and due to the increased delay durations in Steps 4 and 5, the reinforcement time was increased and randomized. Meaning the programmed reinforcement time ranged from 30 s up to 50% of the average duration of the delay (e.g., if the mean delay was 2 mins, maximum reinforcement time could be 1 min). Changes in reinforcement were to slow changes in unit price, or the ratio of work output and duration of reinforcement (Roane et al., 2007). We increased the delay period once participants met our progression criteria, that is, when there were three consecutive sessions with independent FCRs during 90% of opportunities and a reduction in problem behavior to 90% of baseline levels. Delay tolerance program mastery was defined as
three consecutive sessions with independent FCRs during 90% of opportunities and a reduction in problem behavior to 90% of baseline levels at the terminal delay step.

Assessment of Durability of Treatment Effects

Renewal test. The renewal test was arranged similarly to Saini et al. (2018). That is, following BL in Context A and delay and denial teaching in Context B, we returned to Context A while continuing to arrange the contingencies described in the delay tolerance teaching sessions (i.e., ABA renewal). Aside from the context change, there were no other changes from delay tolerance teaching, which included the continued presence of same sets of alternative toys. Elevation in problem behavior relative to the last three delay and denial sessions would be indicative of renewal in this context. The renewal test phase was stopped after three consecutive sessions with problem behavior at or below an 85% reduction or after 10 total sessions. If problem behavior persisted to 10 sessions, we returned to Context B and conducted delay and denial teaching at the terminal step until three consecutive sessions with independent FCR engagement occurred during 90% of opportunities and a reduction in problem behavior to 90% of baseline levels.

Resurgence Test. Researchers arranged a resurgence treatment challenge. Sessions were conducted consecutively at the end of a therapeutic day where it was feasible to deny delivery of functional reinforcers. Meaning that after the first FCR_{complex}, reinforcement was completely withheld (i.e., extinction) for both problem behavior and appropriate FCRs. Procedures were arranged based on those by Fisher et al., (2019). That is, researchers arranged for at least 3 consecutive sessions to occur in a row. If more than three resurgence test sessions were conducted in an appointment, a different researcher took the participant for a walk minimizing attention and restricting access to tangible items before returning to Context A. However, in line
with the procedures by Rose and Beaulieu (2019), researchers provided less preferred activities or toys during the resurgence test. This was done to replicate the treatment extensions conducted at the end of their treatment evaluation, and to be able to investigate the role of these toys in the following comparison. This arrangement is unlike traditional resurgence test evaluations (e.g., Fisher et al., 2019) where access to all items is withheld. The resurgence test phase was conducted until there were three consecutive sessions with problem behavior at or below an 85% reduction or after 10 total sessions.

**Alternative Items Comparison.** Researchers arranged for conditions to better understand the role of mitigation with alternative items by conducting a comparison with three of the four participants. After arranging a replication of Rose and Beaulieu (2019) where researchers arranged a resurgence treatment challenge with the presence of alternative items, researchers arranged an additional resurgence treatment challenge by removing the alternative items from the therapy room. Sessions were otherwise conducted or ceased in the same manner as the resurgence test. For two of the participants, researchers conducted a second set of resurgence treatment challenges with and without access to alternative items in reverse order to evaluate sequence effects.

**Results**

Functional analyses for Noah, David, Alabama, and Mario are depicted on Figure 1. Social-positive reinforcement to tangible items was found to function as a reinforcer responsible for maintenance of Noah (top left panel), David (top right panel), and Mario’s (bottom right panel) problem behavior, where access to attention was responsible for maintaining Alabama’s (bottom left panel) target problem behavior. Noah’s problem behavior was found to be multiply maintained as social negative reinforcement in the form of escape from instruction was also
identified. Thus, we first targeted the social positive function and subsequently targeted the social negative function.

**FCT**

The results of FCT for Noah, David, Alabama, and Mario are depicted on Figure 3 for Noah, and Figure 4 for David, Alabama, and Mario. For all participants, Noah SR+ ($M = 1.3$), David ($M = 4.8$), Alabama ($M = 2.2$), Mario ($M = 1.6$) and Noah SR- ($M = 1.1$) we observed elevated rates of problem behavior in an initial baseline. After implementing FCT problem behavior decreased to near zero levels for all participants. Meaning for Noah (SR+: $M = <0.1$), David ($M = 0.1$), Alabama ($M = 0.2$), Mario ($M = <0.1$) and Noah (SR-: $M = 0$). FCT was effective in reducing problem behavior. In a contingency reversal back to baseline participants once again all engaged in elevated rates of problem behavior (Noah SR+: $M = 1.7$, David $M = 4.1$, Alabama $M = 3.4$, Mario $M = 2.0$, and Noah SR-: $M = 1.6$). During the final three sessions of FCT all participants engaged in a $M = 0.0$ instances of problem behavior and independent engagement in the FCR complex (Noah SR+: $M = 100\%$, David $M = 100\%$, Alabama $M = 98\%$, Mario $M = 98\%$ and Noah SR-: $M = 85\%$). Sessions to mastery of the FCR complex varied across participants. FCT shaping was mastered in 88, 29, 21, 16 and 12 sessions for Noah (SR+), David, Alabama, Mario, and Noah (SR-) respectively.

**Probabilistic Delay Tolerance Teaching**

The results of probabilistic delay teaching for Noah, David, Alabama, and Mario are depicted respectively in the first, second, third, and fourth quadrants across figures. Combined problem behavior is graphed first on each panel on Figure 5 for Noah (Sr+), David, Alabama, and Mario and Figure 6 for Noah’s escape treatment. First, in a return to baseline all participants, engaged in elevated rates of problem behavior, Noah (Sr +: $M = 3.4$), David ($M = 4.4$), Alabama
As a note, Noah took an extended 1-month family vacation, as denoted by the break in the x axis in Figure 5. Next, researchers arranged delay and denial teaching for all participants. For David and Alabama, the schedule increase evoked no elevated rates of problem behavior, differing from Noah and Mario, who observed an increase in problem behavior after each step increase. However, at the end of this phase all participants engaged in in zero levels of problem behavior during the last three sessions at the terminal delay. The second panel for each participant’s data depicts functional communication responses graphed as a rate across baseline, and delay teaching conditions. Last, the third panel for each participant’s data displays tolerance responses as a rate on the left y-axis, and alternative toy engagement as a percentage on the right y-axis. In sum, mastery was met after 45, 17, 25, 34, and 30 probabilistic delay sessions for Noah (Sr+), David, Alabama, Mario, and Noah (Sr-).

Assessment of the Durability of Treatment Effects

Renewal. Graphed on Figures 5 and 6 are the results of the renewal treatment challenge. During the programmed renewal treatment challenge Noah (Sr+) engaged in a rate of 0.1 instances of problem behavior during the first session of the renewal test, meeting criteria for termination of the treatment challenge after 3 sessions (top left panels). David engaged in a mean rate of 0.2 instances of problem behavior during the renewal test, meeting criteria for termination of the treatment challenge after three sessions (top right panels). Alabama engaged in mean rate of 0.1 instances of problem behavior during the renewal test, meeting criteria for termination of the treatment challenge after 3 sessions (bottom left panels). Finally, Mario engaged in mean of 0.1 instance of problem behavior during the renewal test, meeting criteria for termination after 6 sessions (bottom right panels). However, we conservatively conducted an extra three session beyond the reduction criteria due to elevated rates and the historical variability of problem
behavior. During the renewal test for Noah (Sr-), he engaged in a rate of 0 instances of problem behavior during the first session of the renewal test, meeting criteria for termination of the treatment challenge after 3 sessions.

**Resurgence and Alternative Toy Comparison.** Results of the resurgence treatment challenge vary (Figures 5 and 6). Noah engaged in 0 levels of problem behavior during the resurgence treatment challenge with, and without alternative toys (top left panel). David engaged in 0 levels of problem behavior when toys were present, and upon removal of the alternative toys David engaged in a mean rate of 2.7 instances of problem behavior, meeting termination criteria to end only due to predetermined cutoff criteria (i.e., 10 session cap). Researchers then returned to delay and denial teaching (mean of 0.0 instances of problem behavior in the last three sessions) to reintroduce treatment before conducting the resurgence test in the reverse order. During the second set of treatment challenges David engaged in a mean of target problem behavior of 2.2 instances per min without the presence of the alternative toys. Once toys were returned to the room David engaged in 0.0 levels of problem behavior (top right panels).

Alabama engaged in zero levels of problem behavior during the resurgence test with toys. When the toys were removed from the therapy room Alabama engaged in elevated levels of problem behavior at mean rate of 1.0 per min in the first session before returning to zero levels (bottom left panels). During the resurgence treatment challenge with alternative toys present Mario engaged in variable rates of problem behavior ($M=0.5$ instances per min). Mario was not included in the resurgence without toys evaluation as there were safety concerns related to reports of automatically maintained pica (bottom right panels). Lastly, Noah (SR-) engaged in 0.6 levels of problem behavior during the resurgence treatment challenge (figure 6) but met criteria for termination after 6 sessions.
Discussion

Importantly, we taught four participants to engage in a FCRComplex that replaced problem behavior, and to tolerate delays to reinforcement by engaging in an alternative task. We were able to successfully utilize probabilistic delays to increase durations for delay and denial tolerance to reinforcement.

Furthermore, we systematically assessed the durability of probabilistic delay and denial-training. We did so by first programming treatment challenges, first with a test for treatment relapse in the form of renewal. To do so we utilized two unique contexts for intervention, context A (orange room, or novel blue therapist) which was associated with a history of reinforcement for problem behavior and no teaching, and Context B (standard therapy room, or pink therapist) with history of teaching and reinforcement for appropriate behavior and extinction for problem behavior. During our programmed renewal treatment challenges, we conducted probabilistic delay teaching in Context A where we observed a return in problem behavior for one out of five datasets. We coded both renewal and resurgence consistent with previous studies (Muething et al., 2020, Saini et al., 2018). To depict renewal, we calculated a mean rate of problem behavior for the first three renewal sessions for each participant and divided by the mean rate of problem behavior from their baseline sessions (Figure 7).

In the resurgence treatment challenge, we programmed extinction for both problem behavior and the newly trained alternative behaviors. We observed an increase in problem behavior consistent with resurgence for two of the five datasets (Mario and Noah, SR-). During the social positive treatment challenge and consistent with procedures by Rose & Beaulieu (2019) researchers provided access to the less preferred toys which were available during delay and denial teaching procedures. As a supplemental analysis, researchers were interested in how the presence of competing items may play a role in resurgence responding. After the initial
resurgence treatment challenge was conducted an additional resurgence treatment challenge was completed for three of the four social positive participants. This challenge was the same as the first resurgence test but was conducted without the availability of alternative toys during delays to the functional reinforcer- in line with previous resurgence test procedures (Fisher et al., 2019). For Noah and Alabama, the presence of toys had no impact on problem behavior, however for David the presence of the toys appeared to moderate levels of problem behavior. To represent levels of resurgence we also calculated a mean rate of problem behavior for all resurgence test, and the alternative toys resurgence tests. For each participant we divided those values by the mean rate of problem behavior from their baseline sessions. This calculation (Figures 8, and 9) and represents resurgence as a proportion of baseline.

There are several potential reasons as to why we did not observe a return of problem behavior during the renewal test for four out of the five datasets. Anecdotally, all four participants engaged in similar statements (i.e., “The orange room isn’t as fun”, “Why do we have to go to the orange room”), indicating a known contingency difference between Context B (typical therapy room) and Context A (orange room); Researchers required participants to tact the orange colored construction paper posted on the door to the orange room prior to entry, and to tact the color of the shirt therapists wore once inside the orange room, but it is worth noting that unlike previous studies researchers did not provide any instruction or rules related to the contingencies. Future research could extend these data and examine the influence of rule governed behavior and its impact on problem behavior renewal following context changes.

There are additional theoretical explanations for the present findings with regards to renewal and resurgence mitigation. Kimball et al. (2023) presents a summary of several mitigation strategies found in the basic, translational and applied literature. Probabilistic delay
and denial teaching may include several of the strategies described, for example, the manipulation of treatment duration. This version of FCT included a shaping process for the FCR complex, where mastery was met for Noah, David, Alabama, and Mario after 97, 29, 21, and 26 10-minute sessions, respectively. Conversely at minimum, if not otherwise programmed clinicians may be only implementing three sessions of FCT to see a trend before beginning schedule thinning, which is much less therapeutic time than participants in this study experienced. Basic and applied research has attempted to isolate the impact of increased experiential durations of treatment on relapse (Leitenberg et al., 1975; Nall et al., 2018; Shahan et al., 2020; Sweeney & Shahan, 2013; Trask et al., 2018; Winterbauer et al., 2013; Fuhrman et al., 2018; Greer et al., 2022), however, applied research has been mixed.

Other strategies that are included in delay and denial teaching are extinction cues (Nieto et al., 2020), in this case delay cues which indicate the unavailability of reinforcement. At the onset of the delay researchers would indicate the delay period by saying one of the randomized delay cues, which may have served as a signal associated with extinction. The inclusion of delay cues may have some protectiveness of treatment relapse, this may be an area for future research. Additionally, we hypothesize that the presence of the alternative toys provided stimulus control over responding. Programming discriminative stimuli is another treatment relapse mitigation strategy (Betz et al., 2013; Browning and Shahan, 2021; Fuhrman et al. 2016; Fisher et al. 2020; Greer et al. 2019). The present study provided alternative toys during delay teaching and as a result, the toys may have served as reinforcement correlated stimuli. This idea is supported by the high levels of engagement with the toys despite contacting no reinforcement during these sessions. Researchers Craig, Browning and Shahan (2017) have demonstrated that the presence of food correlated stimuli, in their basic preparation, an illuminated food aperture, decreased
relapse in resurgence tests. Therefore, perhaps the presence of the alternative toys could be serving as intervention-correlated stimuli to decrease the likelihood of treatment relapse. Future studies might examine this phenomenon in more depth. This hypothesis may account for David and Alabama’s engagement in problem behavior once the toys were removed from the room. Another explanation for the return of problem behavior without toys present is that they served as competing items for problem behavior. It is well understood that the presence of competing items during delays (Austin & Tiger, 2015; Driftke et. al, 2020; Ghaemmaghami et. al, 2016; Muharib et. al, 2022) can produce durable intervention outcomes. Nevertheless, this leads us to an important avenue for future research.

Another strength of delay and denial teaching is the FCT shaping process. For Noah (SR+) a resumption of problem behavior occurred at each shaping step increase. As each new step required another variation of the FCR, the previous iteration was no longer reinforced. Similarly, Noah and Mario engaged in a resumption of problem behavior each probabilistic delay step increase (i.e., 0 s delay to 30 s mean delay, 30 s to 60 s and so on). The schedule requirement created experiential periods of extinction-like conditions, a worsening for participants, which may have impacted resurgence responding during the treatment challenge. Future applied research is both needed to systematically identify if the FCT shaping process or if longer exposure to schedule thinning impacts mitigation.

There are a few limitations in the present study. First, unlike Rose and Beaulieu (2019) and others who have recently reported the use of delay schedules (Ghaemmaghami et al., 2016; Hanley et al., 2014), researchers of this study first conducted isolated-contingency functional analyses. This choice differs from the procedures we replicated and may have an effect on the duration of time spent in baseline (time spent reinforcing problem behavior has been reported as
a mitigation strategy in basic and translational research; Winterbauer et al. 2013; Todd et al. 2012 Bruzek et al., 2009; Lambert et al., 2020; Smith & Greer, 2022). This choice may have had an effect on the overall intervention durability against treatment relapse. Second, engagement in the tolerance response was variable throughout probabilistic delay teaching across participants. Researchers provided a praise statement (i.e., “Nice job saying, ‘Okay’”) following participant engagement in the tolerance response, however the delay began in lieu of participant responding. In prior delay and denial teaching application these responses persisted as it stands researchers are unsure why the present findings differ from previous literature.

A strength of probabilistic delay and denial teaching is that it reproduces social situations. Providing an alternative activity or item mimics the natural environment (rarely are we asked to wait for things without an alternative activity, (e.g., waiting at the doctor’s office with magazine or cell phone). Second, these procedures emulate the unpredictability that an individual may encounter in the natural environment. Researchers provided three randomized and counterbalanced delay time requirements, delay cues, and toy groups. Specifically, future research is needed to determine the impact of predictable versus unpredictable delay requirements. Another consideration is the use of naturalistic cues in delay and denial teaching in place of contrived discriminative stimuli. This differs from the stimuli which are used to maintain stimulus control in multiple schedule arrangements. That said, the purpose of this evaluation was not to directly compare schedule thinning procedures, and rather to assess this schedule thinning tool and its durability against treatment relapse. There are several empirically validated procedural arrangements to achieve schedule thinning (i.e., multiple schedules, chained schedules, probabilistic delay arrangements). Individuals who would be good clinical candidates for delay and denial procedures differ greatly from those who would benefit from multiple or
chained schedule arrangements (i.e., the prerequisite skills needed for an individual to engage in a FCR complex, ability to attend to delay cues, and or engage independently with alternative items).

Generalization of our services is extremely important for the continued success of the populations we serve. Successful generalization includes a firm understanding of how treatment will persist under less ideal conditions. Researchers want to place emphasis on the importance of embedding treatment relapse mitigation strategies into interventions to facilitate sound outcomes. To better understand the maintenance of treatment gains with delay and denial teaching we arranged challenges to treatment so to identify the endurance of newly taught skills and the impact on the resumption of problem behavior.

The results of the present study provide both clinical and theoretical implications. From a clinical standpoint, the present findings replicate and extend the results obtained by Rose and Beaulieu (2019) which show probabilistic delays as an effective intervention to 1.) teach and shape communication responses, 2.) train an alternative appropriate waiting response, and 3.) reduce problem behavior. Furthermore, the data we’ve collected support probabilistic delays as a fairly robust intervention in respect to treatment relapse. These findings highlight the durability of delay and denial teaching and provide evidence towards its use as a prescribed intervention for the treatment of severe or otherwise challenging behavior. It is the hope of researchers to be able to add to the body of research advancing our understanding of treatment relapse and aligning intervention towards behavioral inoculation.
Bibliography


behavior. *Journal of the Experimental Analysis of Behavior, 111*(1), 75–93.

https://doi.org/10.1002/jeab.488


https://doi.org/10.1037/bdb0000085


https://doi.org/10.1007/s40614-022-00333-2


http://dx.doi.org/10.1037/bdb0000108


Trask, S., & Bouton, M. E. (2016). Discriminative properties of the reinforcer can be used to attenuate the renewal of extinguished operant behavior. *Learning & Behavior, 44*(2), 141-161.  
https://doi.org/10.3758/s13420-015-0195-9

https://doi.org/10.1901/jeab.2011.96-261

http://dx.doi.org/10.1037/bar0000119

https://doi.org/10.1016/j.lmot.2012.03.003

https://doi.org/10.1177/10534512211032628
Table 1.

_example of shaping steps functional communication response_

<table>
<thead>
<tr>
<th>FCT Step</th>
<th>Target FCR</th>
<th>Reinforcement time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>“iPad, please”</td>
<td>30</td>
</tr>
<tr>
<td>2</td>
<td>“iPad, please”</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>+ tolerate removal and hands down</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>“Can I have the iPad please?”</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>+ tolerate removal and hands down.</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>“Excuse me”</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>“Can I have the iPad please?”</td>
<td></td>
</tr>
<tr>
<td></td>
<td>+ tolerate removal and hands down</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>“Excuse me”</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>+ 3s wait.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>“Can I have the iPad please?”</td>
<td></td>
</tr>
<tr>
<td></td>
<td>+ tolerate removal and hands down</td>
<td></td>
</tr>
</tbody>
</table>

Note. Shaping steps for a social positive tangible functional communication intervention.
Table 2.

*Probabilistic Delay Progression*

<table>
<thead>
<tr>
<th>Delay Step</th>
<th>Programmed Mean Delay (s)</th>
<th>Programmed Mean Work Requirement (Noah SR-, responses)</th>
<th>Reinforcement Time (s)</th>
<th>Max Session Cap (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>30</td>
<td>600</td>
</tr>
<tr>
<td>2</td>
<td>30 (0-45)</td>
<td>2 (0-4)</td>
<td>30</td>
<td>600</td>
</tr>
<tr>
<td>3</td>
<td>60 (30-90)</td>
<td>5 (2-10)</td>
<td>30</td>
<td>600</td>
</tr>
<tr>
<td>4</td>
<td>120 (60-180)</td>
<td>10 (5-15)</td>
<td>30-60</td>
<td>900</td>
</tr>
<tr>
<td>5</td>
<td>240 (120-360)</td>
<td>20 (15-25)</td>
<td>30-120</td>
<td>1500</td>
</tr>
</tbody>
</table>

*Note.* Numbers in parentheses are a range.
Figure 1.

Results of Preference Assessments

Note. The y-axis depicts the percentage of selections across trials, and the x-axis the stimulus selected. Top left are Noah’s preference assessment data, David on the top right, Alabama on the bottom left, and Mario on the bottom right.
Figure 2.

Results of Functional Analyses

Note. The y-axis depicts the rate of challenging behavior across the sessions on the x-axis.
**Figure 3.**

*Results of Functional Communication Teaching for Noah*

*Note.* The top panel for both social positive and social negative evaluations depict the rate of challenging behavior on the y-axis across sessions on the x-axis. The bottom panel depicts FCRs as a rate, across sessions.
Figure 4.

*Results of Function Communication Teaching for David, Alabama, and Mario*

![Graph showing results](image)

*Note.* The top panels for David, Alabama, and Mario depict the rate of challenging behavior on the y-axis across sessions on the x-axis. The bottom panel depicts FCRs as a rate, across sessions.
Figure 5.

*Results of Delay and Denial Teaching (SR+)*

*Note.* The left y-axes for the top panel for Noah (Sr+), David, Alabama, and Mario depict the rate of challenging behavior, the middle panel the FCRs as a rate, and the bottom tolerance responses as a rate. On the bottom panel for each participant the right y-axis graphs engagement as a percentage with the alternative items. The x-axis for all graphs depicts sessions.
Figure 6.

Results of Delay and Denial Teaching for Noah (SR-)

Note. The left y-axes for the top panel for Noah (Sr-) depict the rate of challenging behavior, the middle panel the FCRs as a rate, and the bottom tolerance responses as a rate. On the bottom panel the right y-axis depicts engagement as a percentage with the alternative items. The x-axis for all graphs depicts sessions.
Figure 7.

Renewal as a Proportion of Baseline

*Note:* Proportion of baseline responding during renewal test sessions for each participant dataset.
Figure 8.

Resurgence as a Proportion of Baseline

Note: Proportion of baseline responding during resurgence test sessions with toys for each participant dataset. David experienced two programmed resurgence tests, and both were calculated separately.
Figure 9.

*Resurgence during Alternative Toys as a Proportion of Baseline*

Note: Proportion of baseline responding during resurgence test sessions without alternative toys for each participant dataset. David experienced two programmed resurgence tests without alternative toys, and both were calculated separately.
### Supplemental Information

#### Appendix A

<table>
<thead>
<tr>
<th>Session #</th>
<th>Trial: <em>Immediate or delay</em></th>
<th>Delay (s)</th>
<th>Delay Cue</th>
<th>Toy Group</th>
<th>SR+ Duration</th>
<th>Session Therapist Initials</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Imm</td>
<td>Delay</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Imm</td>
<td>Delay</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Imm</td>
<td>Delay</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Imm</td>
<td>Delay</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Imm</td>
<td>Delay</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Imm</td>
<td>Delay</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Imm</td>
<td>Delay</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Imm</td>
<td>Delay</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Imm</td>
<td>Delay</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Imm</td>
<td>Delay</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Imm</td>
<td>Delay</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Imm</td>
<td>Delay</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Imm</td>
<td>Delay</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Imm</td>
<td>Delay</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Imm</td>
<td>Delay</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Imm</td>
<td>Delay</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Imm</td>
<td>Delay</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Imm</td>
<td>Delay</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Imm</td>
<td>Delay</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Imm</td>
<td>Delay</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>