An Evaluation of Procedures that Affect Response Variability

By

Joseph D. Dracobly

Submitted to the graduate degree program in the Department of Applied Behavioral Science and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

____

Chairperson Claudia Dozier, Ph.D.

____

Pamela Neidert, Ph.D.

____

Derek Reed, Ph.D.

____

David Jarmolowicz, Ph.D.

____

Barbara Thompson, Ph.D.

Date Defended: 07/15/2014

The Dissertation Committee for Joseph D. Dracobly

certifies that this is the approved version of the following dissertation:

An Evaluation of Procedures that Affect Response Variability

Chairperson Claudia L. Dozier, Ph.D.

Date approved: 07/15/14

Abstract

Response variability has traditionally been studied as both a by-product of schedules of reinforcement and as a dimension of operant behavior. More recently, researchers have focused on inducement (via extinction), direct reinforcement (via percentile and lag schedules), and stimulus control of response variability. The purposes of the current study were to (a) determine general levels of response variability across a large number of children, (b) replicate and extend previous research on effects of various procedures on the production and maintenance of both variable and novel responses, and (c) determine if stimuli correlated with response variability and response repetition contingencies could immediately affect response variability. In Study 1, there was a nearly bi-modal distribution of participants who emitted low and high variability. In Study 2, for the majority of children, variability increased when exposed to extinction but both fixed-lag 4 and variable-lag 4 schedules produced the highest levels of variability and novelty. Finally, in Study 3, stimuli correlated with each contingency were effectively used to evoke relatively rapid alternation between repetition and variation. Implications and considerations for future researchers are discussed.

Acknowledgements

I would like to acknowledge my mentor, Dr. Claudia L. Dozier, for her all of her contributions to the development of this project and writing of this proposal. I would also like to acknowledge my lab mates for their above-and-beyond help in the development and conduct of this project. Finally, I would like to acknowledge my wife, Libba, for all her support (including excellent proofreading).

Table of Contents

An Evaluation of Procedures that Affect Response Variability

Although early researchers observed systematic changes in response form (e.g., Antonitis, 1951; Ferraro & Branch, 1968), they were primarily interested in studying the functional properties of the operant class. Over time, researchers became more interested in investigating the effects of various environmental manipulations on response variability (i.e., systematic changes in response topography from instance to instance). Basic researchers have conducted the majority of these studies and have focused on fine-grained analysis of the effects of extinction, indirect reinforcement (e.g., intermittent schedules of reinforcement, schedule-induced behavior, etc.), and direct reinforcement on both moment-to-moment and longer-term response variability. More recently, there has been an increased focus in applied research on the effects of procedures to increase response variability of socially relevant responses. For example, researchers have provided evidence that promoting response variability may result in the production of creative behavior (e.g., Goetz & Baer, 1973), the teaching of new responses (e.g., Lalli, Zanolli, & Wohn, 1994; Pryor, Haag, & O'Reilly, 1968), and the reduction of repetitive behavior (e.g., Esch, Esch, & Love, 2009: Napolitano et al., 2010).

Research on response variability has both conceptual and applied implications (for extensive discussion see reviews by Lee, Sturmey, & Fields, 2007 and Neuringer & Jensen, 2013). First, numerous researchers (e.g., de Souza Barbara, 2012a; Goetz & Baer, 1976; Machado, 1989; Page & Neuringer, 1985; Pryor et al., 1969; Schwartz, 1982) have suggested that it is important to study response variability because it occurs in various species, across operant classes, and in most situations and contexts. Therefore, continued study of this phenomenon allows us to learn more about how complex repertories come about (e.g., Harding, Wacker, Berg, Rick, & Lee, 2004; Holman, Goetz, & Baer, 1977). A common interpretation is that response variability is a dimension of operant behavior that is affected by antecedents and consequences in a similar way to other operant dimensions (e.g., magnitude, location, etc.; Denney & Neuringer, 1988; Joyce & Chase, 1990; Machado, 1989; Page & Neuringer, 1985). In fact, in support of this interpretation, researchers have reliably shown that (a) contingencies can modify response variability in the same way they modify other operant dimensions (Antonitis, 1951; Kinloch, Foster, & McEwan, 2009; Morgan & Lee, 1990; Morris, 1987), (b) response variability can be altered by direct reinforcement (e.g., Goetz & Baer, 1973; Lee, McComas, & Jawor, 2002; Machado, 1989; Page & Neuringer, 1985), and (c) response variability can come under stimulus control (e.g., Denney & Neuringer, 1998; Page & Neuringer, 1985; Ward, Kynaston, Bailey, & Odum, 2008). In addition, it is likely that response variability offers many of the same survival advantages as other operant dimensions (Page and Neuringer, 1985).

In addition to the importance of the conceptual understanding of response variability, it is important to study the application and refinement of procedures for producing response variability in various populations across numerous contexts. This allows us to determine under what conditions more complex human behavior (e.g., creativity and problem solving) come about. Furthermore, increasing response variability is important in clinical populations. For example, individuals with autism spectrum disorder often engage in repetitive or stereotypic behavior (e.g., Turner, 1999), which is inversely related to the amount of response variability displayed by these individuals. Determining procedures for increasing response variability (and decreasing repetition) in language, motor behavior, and social interaction allows for an increase in different topographies of responses within various response classes, which in turn allows for greater levels of reinforcement. To these ends, the majority of research on response variability has focused on three areas: induction of variability, direct reinforcement of variability, and stimulus control of variability.

Induction of Response Variability

Previous researchers have suggested that various contingencies can produce response variability. These contingencies are typically discussed as either inducing response variability (e.g., extinction and intermittent schedules or reinforcement) or directly producing (programming for) response variability (e.g., direct reinforcement of novel responding, percentile schedules, and lag schedules of reinforcement). Although previous researchers have found that intermittent schedules of reinforcement induce more response variability than continuous reinforcement schedules (e.g., Machado, 1989; Neuringer, 1991; Odum et al., 2006; Tatham et al., 1993), the majority of research on the induction of response variability has focused on the role of extinction. Extinction involves the removal of the response-consequence relationship by no longer delivering a reinforcer following a response (Skinner, 1933). Typically, extinction results in the reduction and elimination of responding. However, before this occurs, there is often an increase in responding, which often involves an increase in the variability of response topography (e.g, Antonitis, 1951; Morgan & Lee, 1996).

There has been limited basic research on the direct effects of extinction on response variability. However, researchers (e.g., Kinloch, Foster, & McEwan, 2009; Morgan & Lee, 1996; Souza, Abreu-Rodrgues, & Baumann, 2010) have identified several general effects of extinction on response variability. First, basic researchers have found that extinction will produce increases in variability in a variety of dimensions, including inter-response time (IRT). Second, extinction will inhibit subsequent development of response variability by reinforcement.

Third, variability induced by extinction may be higher after a history of reinforcement for variation and lower after a history of reinforcement for repetition.

To date, several applied studies have been conducted on the effects of extinction on the induction of variability of socially relevant responses (e.g., Grow, Kelley, Roane, & Shillingsburg, 2008; Lalli, Zanolli, & Wohn, 1995; Valentino, Shillingsburg, Call, Burton, & Bowen, 2011). Overall, researchers have found that extinction is an effective procedure for increasing variability of socially relevant responses but that this induction decreases over time. For example, Grow et al. (2008) evaluated the effects of extinction on inducing an appropriate mand (e.g., signs, gestures, vocal requests) in three individuals with severe problem behavior. Following a functional analysis, Grow et al. evaluated the effects of extinction plus reinforcement for inducing and maintaining an appropriate mand to access the functional reinforcer. Grow et al. found this procedure to be effective in quickly inducing an appropriate mand. This study provided initial evidence that extinction could effectively, and quickly, induce an appropriate, socially relevant response.

Valentino et al. (2011) conducted a replication of the Grow et al. (2008) study using two forms of appropriate behavior (signs [a higher rate response] and vocalizations [a lower rate response]). Valentino et al. found that extinction plus reinforcement was effective in increasing vocalizations across five out of six topographies of vocalizations, and because reinforcement occurred during baseline, it was clear that extinction was the primary variable that resulted in the increase in vocalizations.

Although extinction may be used to produce response variability, there are several limitations associated with its use. First, the effects of extinction on the induction of response variability may occur for a short period of time because behavior is no longer contacting reinforcement (e.g., Miller & Neuringer, 2000). In addition, all behavior is reduced during extinction (i.e., low-probability and high-probability responses are affected equally). Therefore, when reinforcement is again available, behavior may not actually contact reinforcement. Furthermore, if behavior does contact reinforcement, it is likely to be high-probability responses. This could be particularly problematic if one's repertoire contains a limited number of responses (e.g., a child with autism learning to emit mands) – even under extinction, the rapid alternation of limited responses will be repetitious. Finally, researchers have found that extinction can produce a variety of undesirable side effects, such as aggression and emotional responding (for a review, see Lerman, Iwata, & Wallace 1999).

Direct Production of Response Variability

Although researchers have found that response variability can be induced, greater information can be gained by understanding how reinforcement directly affects response variability. As mentioned above, this is particularly important in applied settings where the direct production of response variability can be useful for teaching complex behavior (e.g., Harding et al., 2004) and teaching variation of responding to children with disabilities that are likely to emit repetitious responding (e.g., Esch et al., 2009; Miller & Neuringer, 2000). The majority of recent research has focused on two reinforcement schedules for directly reinforcing variability: percentile schedules (e.g., Duker & van Lent, 1991; Machado, 1989; Machado, 1992; Machado, 1997; Miller & Neuringer, 2000) and lag schedules (e.g., Cammilleri & Hanley, 2005; Esch et al., 2009; Glover & Gary, 1976; Harding et al., 2004; Lee et al., 2002; Lee & Sturmey, 2006; Napolitano et al., 2010; Page & Neuringer, 1985; Susa & Schlinger, 2012).

Much of the initial research on the direct production of response variability was focused on reinforcement of novel behavior. Pryor et al. (1969) conducted the first evaluation of the effects of reinforcement on the production and maintenance of novel behavior. Pryor et al. identified various "trick" responses that could be made by porpoises. To shape these responses, Pryor et al. reinforced novel approximations of these responses. They found this to be an effective procedure for teaching novel behavior, to the point that the study was terminated due to the complexity of responses.

Goetz and Baer (1973) also conducted an early evaluation of reinforcement on the production of novel behavior. Goetz and Baer first identified various components of block designs that could be made by young children and then instructed the children to make a block design. During baseline, no programmed consequences occurred following the completion of a structure. During the novelty phase, each time the child made a structure that had a form that had not previously been made (i.e., contained a different combination of components), the teacher provided enthusiastic, descriptive praise (e.g., "Good job, that's different"). During the repetition phase, each time the child made a structure that was the same as a form that was previously made in the session, the teacher provided enthusiastic, descriptive praise (e.g., "Good job, that's the same"). Goetz and Baer found that form diversity (i.e., the number of novel forms per session) increased when reinforcement was available for novel responses, whereas form diversity decreased to baseline levels when reinforcement was available for response repetition.

Although the reinforcement of novel behavior is a useful method for increasing novel behavior, there are significant limitations to this approach. First, it is not clear how large the response class must be to continually produce novel behavior as opposed to simply producing low-frequency responses (e.g., Goetz & Baer, 1974). Second, if novel behavior involves multiple components, behavior may quickly become so complex as to preclude reliable observation and measurement (e.g., Pryor et al., 1969). Finally, in many situations, it is not necessary to produce novel behavior but rather simply produce variation in responding, such as teaching language to children with disabilities (e.g., Esch et al., 2009; Lee et al, 2002).

Percentile Schedules

One type of reinforcement schedule that has been used to produce variability in responding is the percentile schedule. In the percentile schedule, responses that are occurring below a certain probability produce reinforcement. Basic researchers (e.g., Machado, 1989; Machado, 1992; Machado, 1997) have found that the development of moment-to-moment repetition and higher order stereotypy (e.g., the repetition of a pattern of responses on a left and right key, such as LRRLL, LRRLL, LRRLL, LRRLL or a reliable pattern of mands, such as candy, book, ball, iPad, tickle, candy, book, ball, iPad, tickle) that are associated with other procedures are almost completely eliminated when using a percentile schedule. For example, Machado (1989) taught pigeons to peck one of two keys. Machado then provided reinforcement for pecks that occurred below a specific relative-frequency value (e.g., 30%). Machado found that response variability was positively correlated with percentile schedule requirements. That is, leaner percentile schedules produced higher levels of response variability. In subsequent replications, Machado (1992 & 1998) also found that variability, of both pecks and switching, could be increased by implementing a percentile schedule, with greater variability produced by more stringent percentile schedules.

Only two applied studies (Duker & van Lent, 1991; Miller & Nueringer, 2000) have evaluated the effects of percentile schedules on variability of socially relevant behavior. Although researchers found that percentile schedules were effective for increasing response variability and decreasing the likelihood of higher order stereotypy, application of this procedure is unlikely in socially relevant environments because of the difficulty of implementation. As mentioned above, the percentile schedule requires continuous frequency and probability calculations, which makes using this schedule with non-automated responses very difficult. Because most socially relevant behavior is currently unable to be transduced via automated devices, this makes the percentile schedule of little use in applied work.

Lag Schedules of Reinforcement

Lag schedules of reinforcement are the most commonly used form of reinforcement for directly producing response variability. A lag schedule involves delivery of reinforcers contingent on a response topography that differs from a specified number of previous responses. For example, under a lag 2 schedule, the topography of the response must differ from the previous two response topographies for reinforcer delivery; under a lag 5 schedule, the topography of the response must differ from the previous five response topographies for reinforcer delivery. Both basic (e.g., Machado, 1998; Morris, 1987; Morris, 1989; Page & Neuringer, 1985) and applied (Cammilleri & Hanley, 2005; Esch et al., 2009; Glover & Gary, 1976; Harding et al., 2004; Heldt & Schlinger, 2012; Lee et al., 2002; Napolitano et al., 2010; Susa & Schlinger, 2012) researchers have demonstrated that lag schedules are highly effective at directly producing and maintaining response variability.

Basic research. Although previous researchers (e.g., Pryor et al., 1969) had demonstrated that specific arrangements of reinforcement contingencies could produce varied and novel behavior, their contingencies were never explicitly arranged for response variation. Page and Neuringer (1985) conducted one of the first extensive studies on response variability. In a series of studies, Page and Neuringer addressed the limitations of an earlier study on the development of stereotyped behavior (Schwartz, 1982) and directly evaluated the effects of lag schedules of reinforcement on increases in response variability. In Experiment 1 and 2, Page and

Neuringer replicated the Schwartz experiment and demonstrated that Schwartz's "no more than four" requirement (i.e., in every response sequence, no more than four responses could occur on either the L or R key, which significantly limited the total number of possible response sequences) that precluded response variability. Given that Page and Neuringer were able to demonstrate in Experiments 1 and 2 that response variability was sensitive to reinforcement, the remainder of their experiments focused on determining the effective features and functions of lag schedules of reinforcement. In Experiment 3, Page and Neuringer examined the effects of various lag values (5, 10, 15, 25, and 50) on the production of response variability of eight-response sequences (i.e., sequences comprised of eight responses on either an L or R key). In Experiment 4 Page and Neuringer manipulated the number of responses required for a sequence - 4, 6, and 8 under a lag 5 schedule. In Experiment 5, Page and Neuringer compared the effects of a lag 50 schedule and a yoked variable ratio (VR) schedule. With all the experiments taken together, Page and Neuringer noted that variability appeared to be a fundamental dimension of operant behavior. Specifically, sequences of responses emitted by pigeons could be made highly variable by making reinforcement contingent on response variability. Page and Neuringer provided the strongest, and most direct, demonstration that lag schedules of reinforcement are effective at directly increasing response variability.

Because of the robustness of the effects obtained by Page and Neuringer (1985), there have been a limited number of basic research studies examining the effects of lag schedules on response variability. Morris (1987), in a comparison of developing response variability under free-operant and discrete-trial arrangements, found that a lag 2 schedule was sufficient to increase response variability and that response variability was higher under the discrete-trial procedure than under the free-operant procedure. In a subsequent replication, Morris (1989) again found all lag values (2, 4, and 6) reliably increased response variability, with the highest levels of response variability occurring under the discrete-trial arrangement. Additionally, as found by Page and Neuringer, Morris found that larger lag values produced higher levels of variability than smaller lag values.

Neuringer (1991) compared the effects of a lag 4 schedule with a yoked schedule on the variability of four-response sequences across two response keys (left and right) and found that the lag 4 schedule produced significantly higher levels of response variability than the yoked reinforcement schedule. Machado (1997), in Experiment 3, compared the effects of a lag 25 schedule with the effects of yoked control phase, similar to Experiment 3 in Page and Neuringer (1985). Machado found that this procedure was sufficient to produce high levels of variability (variable responding on an average of 64% trials in the first vary condition and 74% of trials in the second vary condition).

However, the promising results of the effects of lag schedules on increasing response variability are not without some limitations. First, lag schedules are simplest to implement with simple responses, particularly when using larger lag values. This may make the implementation of large lag schedules, which produce the highest sustained response variability, difficult to implement in applied settings, particularly with complex responses (i.e., a human observer may have difficulty observing and recording all differences from instance to instance of a complex responses). Second, lag schedules, as compared to percentile schedules, are much more likely to produce higher order stereotypy (e.g., Machado, 1997). This can be limited by using larger lag values; however, because these larger lag values may be impractical in certain applied settings and with complex responses, researchers need to investigate methods for manipulating low-value lag schedules to reduce higher order stereotypy. Interestingly, because most basic researchers have investigated the effects of various independent variables (e.g., drugs; McElroy & Neuringer, 1990) on response variability, attempts to address this issue have primarily been conducted by applied researchers, for whom the development of higher order stereotypy is particularly problematic (e.g., in individuals with autism, who are already prone to stereotyped responding, the programmatic development of additional stereotyped responding could be very problematic).

Applied research. Because of the potential utility, and relative simplicity, of implementing lag schedules of reinforcement, applied researchers have begun evaluating the effects of lag schedules of reinforcement on various socially relevant responses (e.g., Cammilleri & Hanley, 2005; Esch et al., 2009; Glover & Gary, 1976; Harding et al., 2004; Heldt & Schlinger, 2012; Lee et al., 2002; Napolitano et al., 2010; Susa & Schlinger, 2012). Lag schedules of reinforcement allow for direct reinforcement of true response variability. That is, a response can be repeated provided it is repeated at a frequency above the lag value. For applied researchers, clinicians, teachers, and therapists, lag schedules are also far simpler to implement than other schedules of reinforcement that have been shown to produce response variability (e.g., percentile schedules; Machado, 1990, 1992) because lag schedules do not require real-time probability and frequency calculations.

Lee et al. (2002) conducted one of the first applied, direct evaluations of the effects of lag schedules of reinforcement on response variability. Lee et al. evaluated the effects of a lag 1 schedule of reinforcement on varied appropriate responses to a social question by three adults with autism. During baseline, any appropriate response to the question resulted in reinforcement (either a token or praise and physical interaction). During the differential reinforcement of alternative behavior (DRA) plus lag 1 phase, any appropriate response that was different from the immediately preceding response produced reinforcement. Lee et al. found that the DRA plus lag 1 procedure was effective for increasing the variability of appropriate responses to the question for two out of three participants. Lee et al. also analyzed across-session patterns of responding and found that although the DRA plus lag 1 procedure was effective for producing increases in variable responding within sessions, the overall number of different responses remained relatively low across sessions.

Subsequent studies have continued to provide evidence that lag schedules of reinforcement can effectively increase variability of socially relevant behavior. For example, Lee and Sturmey (2006) compared the effects of a lag 0 (any response, regardless of repetition, produces reinforcement) and a lag 1 schedule on the variability of responses to a question (e.g., “What do you like to do?’) by three teenagers with autism. Lee and Sturmey found that the lag 1 schedule was highly effective in increasing the percentage of responses that were varied and appropriate for two out of three participants. However, as Lee et al. (2002) found, the overall number of novel responses remained low – higher order stereotypy developed under the lag 1 schedule.

The applied studies on lag 1 schedules provide evidence that this schedule can increase variability of socially relevant behavior; however, one of the primary limitations of each of the studies discussed above is the lack of reduction of response variability when the lag 1 schedules were removed. This may have occurred because a lag 1 schedule of reinforcement may be more likely to produce higher order stereotypy (repetition of sequences of responses). Because only a small number of different responses are required to meet the lag 1 requirement, the most efficient form of responding is to alternate between two responses. Therefore, although a lag 1 schedule can increase response variability, it may not be the best schedule to use if a high rate of response variability is desired.

To date, five applied studies (Cammilleri & Hanley, 2005; Glover & Gary, 1976; Harding et al., 2004; Heldt & Schlinger, 2012; Susa & Schlinger, 2012) have included a lag schedule that was not lag 1. Cammilleri and Hanley (2005) were the first to systematically evaluate the effects of a lag "infinite" schedule on response variability on the activity selections of two typically developing children. Cammilleri and Hanley found that the lag "infinite" schedule was effective in increasing the variability of activity selections. In addition, during the return to baseline phase, there was an immediate and sustained decrease in the variability of activity selections. A similar study was conducted by Harding et al. (2004), who evaluated the effects of this lag schedule on increasing variable punching and kicking techniques displayed by martial arts students. Harding et al. found that the lag infinite schedule was effective for increasing the number of novel counterstrikes (punches and kicks) in a drill (feedback) and sparring (no feedback) condition.

Heldt and Schlinger (2012) evaluated the effects of a lag 3 schedule on the tacting of two children with intellectual and developmental disabilities and found that the lag 3 schedule produced almost immediate and sustained increases in the variability of tacts across both participants. In addition, Susa and Schlinger (2012) evaluated the effects of a progressive lag schedule on the verbal responding of a child with autism. Following baseline, in which the participant was asked, "How are you?," the participant was exposed to a progressively increasing lag schedule (1, 2, 3). Across all lag schedules, responses variability increased, with the level of variability increasing in direct proportion to the lag schedule in effect.

Although these studies provided evidence that a leaner lag schedule could produce immediate, sustained, and reversible increases in response variability, the lag "infinite" schedules more closely resembled an extinction procedure and therefore the increased variability may have, in part, been induced. In addition, it would be very difficult to implement a lag "infinite" schedule on a long-term basis because this always requires a novel response to produce reinforcement. This would require extensive data on all responses, which is likely impractical in most settings. Therefore, it would be useful to evaluate the effects of various lag schedules with values between 1 and "infinity." This would serve two important purposes. First, researchers would be able to determine which lag values produce acceptable levels of response variability without over-burdening data collectors. Second, these evaluations would allow researchers to compare levels of response variability across a variety of lag schedules (as in Page & Neuringer, 1985) with respect to both short-term (i.e., within-session) and long-term (across-session) response variability. This could allow clinicians, teachers, and the like, to select the lag schedule that meets their specific needs (i.e., they could select the lag schedule depending on how much response variability they needed and how much higher order stereotypy they could tolerate).

Despite these limitations, there were some interesting findings on the across-session effects of lag schedules that also warranted further investigation. For example, as Page and Neuringer (1985) noted, subjects' behavior under lag schedules may be "quasi-random." That is, the behavior may have higher order patterns, often termed higher order stereotypy. In fact, this quasi-random behavior has been found by some applied researchers under direct reinforcement of variability using lag schedules (e.g., Goetz & Baer, 1973; Lee et al., 2002). This higher order stereotypy involves a change in responses from instance-to-instance but a fixed pattern in the sequence of responses, such as, "LRLRLRLRLR" across two, L and R, response keys (as discussed above). Higher order stereotypy is particularly problematic with small response classes, in which one might easily notice the repetition, and with certain populations, such as individuals with intellectual and developmental disabilities, who may be prone to emit stereotyped responding. To date, basic researchers have not directly addressed this problem. However, because of the problem higher order stereotypy may present in certain applied settings (e.g., in working with children with autism, who are already prone to stereotyped responding), applied researchers have begun to address this issue. For example, Lee et al. (2007) provided some suggestions for reducing higher order stereotypy, such as implementing variability schedules with multiple response classes, alternating tasks and responses across teaching trials, and engaging in alternate tasks, such as vocal interaction, between teaching trials. For example, if one is interested in increasing the variability of already acquired responses, alternating responses across trials could interrupt the development of repetitive patterns of responding, particularly if the response sequence occurs quickly. By precluding the rapid emission of topographies from one response, fixed progression through the response class may be stopped, thus increasing the likelihood that other responses would be emitted (i.e., this may produce increases in response variability through the same, or similar, mechanism as percentile schedules). In addition, if one is establishing a new response class and wants to ensure a variety of responses are included, programming incompatible responses between teaching trials may preclude the initial development of response patterns (or hierarchies). In turn, this may allow for greater ease in increasing, and maintaining, variability once the response class is established (i.e., because no higher order patterns of responses will be established when initially forming the response class).

Stimulus Control of Response Variability

Researchers have reliablu demonstrated that lag schedules are effective at increasing response variability. Although this finding is promising, if response variability is a fundamental dimension of operant behavior, it is important to demonstrate that this dimension, like others, can come under stimulus control. Stimulus control of response variability could be particularly useful in certain applied settings, where it may be important to alternate between variation and repetition.

Previous researchers (Denney & Neuringer, 1998; Page & Neuringer, 1985; Souza & Abreu-Rodrigues, 2010; Ward et al., 2008) have evaluated the effects of discriminative stimuli on response variability for the purpose of demonstrating that response variability can come under stimulus control. Discriminative stimuli are stimuli correlated with the availability of reinforcement (e.g., Lattal, 1975), and stimulus control refers to differential control by two or more stimuli, typically a discriminative stimulus that is correlated with availability of reinforcement and an s-delta, a stimulus correlated with the absence of reinforcement (e.g., Holz & Azrin, 1961). Stimulus control of response variability is important for both the demonstration that variability is a dimension of operant behavior and for the solving of various socially relevant problems. Specifically, stimulus control of response variability may promote the generalization of response variability when lag schedules are discontinued (by continuing to present the stimulus correlated with the lag schedule) and programming stimulus control may allow for rapid alternation between variation and repetition, which is necessary in some social settings (e.g., repeating your name at a party when you meet new people but talking about various topics with people you already know).

Page and Neuringer (1985) conducted the first evaluation of stimulus control of response variability. In Experiment 6 of their study, Page and Neuringer evaluated the effects of a stimulus control procedure on response variability by bringing variable and repetitious responding under stimulus control, presenting these stimuli in a multiple schedule, and reversing the correlation of the stimuli. Page and Neuringer found that pigeons would reliably switch from variable responding to stereotyped responding under control of changes in stimuli, with their responding becoming more accurate across the phase.

Denney and Neuringer (1998) conducted the second evaluation of stimulus control of response variability using rats. In the presence of the houselight and no tone (first 10 rats) or houselight off and tone on (second 10 rats), reinforcement was available for variable responding. However, in the presence of no houselight and a tone (first 10 rats) or a houselight and no tone (second 10 rats), noncontingent reinforcement (NCR) in which reinforcer delivery was yoked to equate reinforcement in the variable responding component was in effect. Denney and Neuringer found that (a) response variability was higher in the presence of the light previously correlated with the variable responding component and (b) response variability increased nearly immediately after presentation of the light correlated with the variable responding component. In addition, when the lights were removed, response variability in the variable responding component reduced, whereas response variability in the NCR component increased to nearly identical levels as those in the variable responding component. Ward et al. (2008) extended these results by including contingency reversals, which is a much stronger method of demonstrating stimulus control because one is able to demonstrate that the stimulus directly produces changes in responding (i.e., when the reversal occurs, if there is strong stimulus control, behavior should change to be "in line" with the new stimulus). Ward et al. found that response variability was more sensitive to the contingency reversals than response repetition.

Although research has provided evidence that response variability can come under stimulus control, there are still areas that require further clarification, particularly in the application of promoting stimulus control of variable responding with socially relevant responses. First, researchers should evaluate the effects of variability-correlated stimuli on the production of variability during novel tasks or in novel situations. That is, researchers should determine if variability-correlated stimuli can signal "when" to behave variably, regardless of the response form. This could be particularly useful in the treatment of individuals with intellectual and developmental disabilities, who may have difficulty learning when variable responding is necessary. Second, researchers should determine if variability-correlated stimuli can facilitate the long-term production of variable responding. For example, researchers should determine if alternating between variability-correlated and repetition-correlated stimuli is more effective than constant exposure to direct reinforcement of variability in the production of long-term variability and reduction of higher order stereotypy.

Overall, both basic and applied researchers have found that a variety of operant processes can affect response variability. Response variability can be induced by extinction; however, the most efficient and effective method for producing response variability is direct reinforcement, either in the form of percentile schedules or lag schedules of reinforcement. Because of the difficulty in using percentile schedules, most researchers have focused on evaluating the effects of lag schedules of reinforcement on response variability. Both basic and applied researchers have found that even low lag values (e.g., lag 1) can produce immediate and sustained response variability. In addition, various applied researchers have found that reinforcement of novel behavior (a lag "infinite" schedule) is even more effective than low-value lag schedules at producing response variability. However, even when reinforcing of novel responses only, researchers have found that when examining across-session responding, a higher order pattern of responding, often termed higher order stereotypy, emerges. To date, only researchers using percentile schedules have found a solution to this problem. Despite this problem, researchers have also found that response variability can come under stimulus control and can be maintained for extended periods of time. Taken together, researchers have found robust effects of direct reinforcement of response variability; however, many questions remain unanswered, including how to produce both within-session and across-session response variability, how to improve the effectiveness of lag schedules at producing across-session response variability, and if response variability, once brought under stimulus control, can be rapidly started and stopped.

Purpose

There are multiple purposes of the current study. First, I replicated and extended previous research by determining the levels of variability on a simple task with a large number of children (Study 1). These data provided local norms with respect to variability and allowed me to identify children to include in Study 2. Second, I evaluated the utility of various procedures to increase response variability both within and across sessions (Study 2). Specifically, I evaluated the effects of extinction, fixed-lag schedules, and variable-lag schedules on the production of response variability. Third, I evaluated whether stimuli correlated with reinforcement of variability produced variability on a novel task and whether variability could be turned "on and off" by rapidly alternating between stimuli correlated with variability and repetition (Study 3).

Study 1 Method: Response Variability Assessment

Purpose

The purpose of Study 1 was to determine general levels of response variability with a large number of preschool-aged children. This was done by determining levels of variability under a fixed-ratio (FR) 1 schedule of reinforcement, which is a common variability-independent schedule. This allowed me to determine the degree of response variability under conditions that are typical of the natural environment (i.e., involve the delivery of reinforcers for all responses regardless of degree of variability).

Participants and Setting

Twenty-four preschool-age children (ranging in age from 2 ½ to 5 years of age) who attended the Edna A. Hill Child Development Center (CDC) at the University of Kansas participated in this study. These children included both typically developing children (i.e., children with no known diagnosis) and children with intellectual and developmental disabilities (IDD; particularly, autism spectrum disorders) that had been diagnosed by a neurologist or psychologist as evidenced by their intake paperwork. Detailed demographic information for each child is displayed in Table 1. Children were selected from two preschool classrooms, which are attended by a combination of approximately 35 typically developing children and children with IDD; and two early-intervention classrooms, which are attended by 10 children diagnosed with autism spectrum disorders. To participate in this study, each child had to be able to follow simple directions (i.e., one-step and two-step instructions, such as, “Pick a block” and “Choose four blocks and place them on the card in front of you”) and had a well-developed listener repertoire (i.e., receptively and expressively identified shapes, colors, locations) as evidenced by curriculum assessments conducted throughout the year in their classrooms.

All sessions were conducted in an individual session room located near each child's classroom. Sessions were conducted 2 to 5 times per day, 3 to 5 days a week. The room contained a table, chairs on each side of the table (or a bench on the child's side), chairs for data collectors, and materials relevant to each condition. The room also had a ceiling to mid-wall one-way mirror.

Materials

During all sessions, a white laminated card depicting four squares with black borders that are printed horizontally (see Figure 1) was available. In addition, four piles of four different colored blocks, all grouped by color (i.e., pile of four red blocks, pile of four blue blocks, pile of four green blocks, pile of four yellow blocks) were available. Additionally, a container with preferred edible items was available. A picture of the materials and arrangement used in each session is depicted in Figure 2.

Response Measurement and Interobserver Agreement

Trained observers collected data on child and therapist behavior using a pencil and paper. Data were collected on the response form that was based on the order of block placement. That is, data collectors recorded the color of the blocks the child placed, from the therapist's left to right, on each trial after all four blocks were placed on the card. For example, if on a particular trial the child placed, from left to right, a red block, a green block, a blue block, and a yellow block, the data collector recorded "R G B Y" on the data sheet for that trial. If no response was made within 1 min of a trial starting (i.e., after the consumption of the edible item), data collectors recorded "NR." The primary dependent variable was the number of different response forms in a session (within-session variability). A different response form was defined as any difference of the order of colored blocks on a particular trial as compared to all other trials. For example, if during trial 1, the response form was R G B Y, and during trial 2, the response form was R G Y B, these were counted as two different response forms. This was calculated after each session by counting the number of different response forms observed during a particular session. The secondary dependent variable in all sessions was the cumulative number of novel response forms across sessions (across-session novelty). A novel response form was defined as a response form that has not been observed in any previous session. This was calculated after each session by counting the number of novel response forms displayed across all sessions. Because the response requirement was four blocks per trial and there were four different colored blocks from which to choose, the total number of possible responses was 256 (i.e., to calculate possible combinations with repetition allowed, the formula is *n* to the *r* power, where *n* represents the number of options from which to choose and *r* represents the number that must be selected on each trial, so because there are four options and four blocks must be chosen on each trial, our calculation is 4 to the 4^{th} power, for a product of 256).

Data were also collected on therapist delivery and non-delivery of edible items and praise. Therapist delivery of edible items and praise was defined as placing the edible item on a plate near the participant and making a praise statement within 5 s of the participant placing a fourth block on the response card. Therapist non-delivery of edible items and praise was defined as not placing the edible item on a plate near the participant and not starting a praise statement within 5 s of the participant placing a fourth block on the response card. Data collectors recorded, on the same data sheet they record block placement, a “Y” (for “Yes”) if the therapist placed the edible item on the plate and made a praise statement and an “N” (for “No”) if the therapist did not place an edible item on the plate and did not make a praise statement on each trial.

To determine interobserver agreement (IOA), a second observer collected data for at least 30% of sessions across children. To calculate IOA, for each separate measure, trials with exact agreement were scored a 1 and trials with a disagreement were scored 0. The trials with agreement were then summed, divided by the number of trials with agreement plus disagreement, and multiplied by 100%. IOA was 100% for Lorena (33.33% of sessions), Mona (50% of sessions), Bruto (42.9% of sessions), Matteo (47.1% of sessions), Nico (50% of sessions), Armondo (50% of sessions), Olga (42.9% of sessions), Gastone (33.3% of sessions), Horatio (40% of sessions), Luigi (42.9% of sessions), Nadia (37.5% of sessions), Furio (50% of sessions), Kajetan (33.3% of sessions), Imilia (37.5% of sessions), Mordecai (42.9% of sessions), Cassio (33.3% of sessions), Bea (57.1% of sessions), and Matilda (63.67% of sessions). IOA was calculated for 42.9% of sessions for Arlo and was 96. 7% (range: 90% - 100%). IOA was calculated for 33.3% of sessions for Biago and was 96. 7% (range: 90% - 100%). IOA was calculated for 66. 7% of sessions for Maria and was 97.5% (range: 90% - 100%). IOA was calculated for 45. 5% of sessions for Manlio and was 98% (range: 90% - 100%). IOA was calculated for 40% of sessions for Evangelina and was 98.8% (range: 95% - 100%). IOA was calculated for 50% of sessions for Carlo and was 99% (range: 95% - 100%).

Procedure

Prior to the start of the Study 1, a five-item paired-stimulus preference assessment similar to that used by Fisher et al. (1992) was conducted. The purpose of this preference assessment was to identify highly preferred edible items to use in all three studies. The preference assessment consisted of 10 trials in which each item was paired with every other item once. On a trial, two items were presented in front of the child, and the child was instructed to pick one item. Once the child had chosen an item, he or she was presented the chosen item to consume.

The non-chosen item was removed. Once the child had consumed the item, the next trial began. Once all 10 trials were completed, the number of times each item was chosen was summed and a rank, from most chosen to least chosen, was assigned. In the case of a tie for rank one, the data from the trial in which the two items were presented together was used to determine the rank (i.e., the item chosen on that trial was ranked one and the item not chosen was ranked two). At any point during the three studies if a child stated he or she no longer liked the food item, an additional preference assessment was conducted. Additionally, prior to each session, the top two items were presented and the child was asked to choose one item to use during that session.

All experimental sessions included 20 trials and a minimum of six sessions were conducted for all participants. Additional sessions beyond the first six were conducted until data were stable, based on visual inspection (i.e., either a stable increasing trend, stable responding, or a decreasing trend to low levels). During each session, the white laminated card depicting four squares and the four piles of different colored blocks (with four blocks in each pile) was present. Prior to the start of each trial, the blocks were always placed in the same position, from left to right (from the perspective of the therapist). For example, if prior to session 1 the piles of blocks were, from left to right, R Y G B, for all subsequent sessions in all studies, the piles of blocks were always placed, from left to right, in this same order. See Figure 2 for an example of how materials were set up for each trial for a particular participant. Prior to the start of each session, the child was given instructions. The therapist presented the materials and said, “I want you to place a block in each one of the squares on the card. If you put four blocks on the card, you will get a treat. You can start when I say, ‘Go.’” On each trial the therapist said, “Go” or a very similar instruction. The participant had 1 min to start the response sequence (the therapist counted this time silently). If the child did not start the sequence within 1 min, or if the child did not complete the sequence within 1 min after starting the sequence, the data collectors scored "NR." During this 1-min period, the therapist occasionally (e.g., every 5 – 10 s) reminded the child to put blocks on the card. This prompting occurred with only a few participants and very infrequently, typically at the beginning of the study. Once a child placed a block on the card and removed his or her fingers from the block, the block was required to remain in that position. If a child attempted to move a block that he or she had placed, the therapist told the child, "Once a block is down, it must stay there." This happened with only a few participants and very infrequently, most typically at the very beginning of the study. Once the child placed four blocks in the four squares, the therapist gave the child one piece of edible while providing praise (e.g., saying, "Good job putting the blocks in the squares"). Next, the therapist removed the blocks, placed the blocks back in the relevant pile, and then implemented the next trial until all 20 trials had been conducted. If the child did not start the response sequence within 1 min from the end of the last trial, or did not complete the response sequence within 1 min of starting the sequence, the therapist picked up the card for 5 s and then started the next trial. Responding (regardless of degree of variability) was reinforced on an FR 1 schedule of reinforcement by delivering one highly preferred edible item and praise.

Results and Discussion

The results of Study 1 are depicted in Figures 3 through 9. In addition, the mean number of different responses (mean variability) for each child is depicted in the third column of Table 1 and the mean number of novel responses (mean novelty) for each child is depicted in the further column of Table 1. In each figure, the number of different responses (variability; within session) is depicted by white circles and the cumulative number of novel responses (novelty; across sessions) are depicted by black bars. Data are grouped by level of variability and novelty. Low variability and novelty was defined as emitting an insufficient number of different responses in each session to meet a lag 4 schedule (i.e., the largest, on average, schedule in Study 2) and fewer than 5% (10) of the total possible responses (256). Moderate novelty was defined as alternating between emitting an insufficient and sufficient number of different responses in each session to meet a lag 4 schedule and between 5% (10) and 10% (25) of the total possible responses (256). High variability and novelty was defined as emitting a sufficient number of different responses in each session to meet a lag 4 schedule and greater than 10% (25) of the total possible responses (256). These levels of variability and novelty were determined at the conclusion of Study 1 based on the responding of the majority of children (21/24). That is, the majority of children either emitted variability that was insufficient to meet a lag 4 schedule and correspondingly less than 5% of the total possible responses (12/24) or emitted variability that was sufficient to meet a lag 4 schedule and correspondingly greater than 10% of the total possible responses (10/24). These two groups were classified as low and high variability and novelty. Therefore, the two children who emitted variability that was sometimes sufficient to meet a lag 4 schedule and correspondingly between 5% and 10% of the total possible responses were classified as in the moderate variability and novelty group. Data were grouped based on both variability (within session) and novelty (across sessions) because each level of novelty required a corresponding similar level of within-session variability. That is, to generate a high level of novelty (i.e., greater than 25 different responses), a participant would have to emit at least five different responses per session (which would be equal to the number of responses required to maximize reinforcement under a lag 4 schedule of reinforcement).

Figures 3-6 depict data for participants who displayed low levels of variability. In Figure 3, the data for Nico are depicted in the top panel, the data for Miles are depicted in the middle panel, and the data for Bruto are depicted in the bottom panel. Nico emitted only 1 response across all sessions in Study 1. Miles emitted a mean of 2.6 (range: 1-4) different responses within session and 4 novel responses across sessions, which stopped after the first session (i.e., regardless of how many responses he emitted each session, these responses were always from the same group of four responses). Finally, Bruto emitted a mean of 3.8 (range: 1-4) different responses within session and a total of 6 novel responses across sessions, with novel responses, after the first session, only in sessions 3, 4, and 6.

In Figure 4, the data for Imilia are depicted in the top panel, the data for Luigi are depicted in the middle panel, and the data for Arlo are depicted in the bottom panel. Imilia emitted a mean of 3.9 (range: 3-4) different responses within session and after the first session, she did not emit any novel responses. Arlo and Luigi both emitted a mean of 4 (no range) different responses within session, and after the first session, neither child emitted additional novel responses.

In Figure 5, the data for Olga are depicted in the top panel, the data for Cassio are depicted in the middle panel, and the data for Mona are depicted in the bottom panel. Both Olga and Cassio emitted a mean of 4 (no range) different responses within session and after the first session, neither child emitted additional novel responses (more than six sessions were conducted with Cassio despite the stable level of responding because he was the first participant in Study 1 and the six-session criteria had not yet been established). Finally, Mona emitted a mean of 4.3 (range: 4-5) different responses within session and 6 novel responses across sessions and stopped emitting novel responses after session 4.

In Figure 6, the data for Gastone are depicted in the top panel, the data for Matilda are depicted in the middle panel, and the data for Carlo are depicted in the bottom panel. Gastone emitted a mean of 4.7 (range: 4-5) different responses within session and 10 novel responses across sessions, with, after the first session, one novel response in sessions 2, 3, 5, 7, and 9. Matilda emitted a mean of 3.2 (range: 1-6) different responses within session and 14 novel responses across sessions, which she stopped emitting after session 7. Interestingly, although Matilda emitted a level of variability (M=3.2) that alternated between being sufficient and insufficient for meeting a lag 4, she did emit greater than 12 novel responses across sessions (in Table 1, the "M" next to her mean level of novelty denotes that level met the moderate variability criterion). However, because she did not emit a sufficient level of variability within sessions, she was classified in the low variability and novelty group. Finally, Carlo emitted a mean of 5.1 (range: 4-8) different responses within session and 10 novel responses across sessions, which stopped after session 6. Interestingly, Carlo's level of variability (M=5.1) was sufficient, in some sessions, to meet a lag 4 schedule (in Table 1, the "M" next to his mean level of variability denotes that level met the moderate variability criterion). However, because he did not also emit more than 12 novel responses across sessions, he was classified in the low variability and novelty group.

In Figure 7, the data for Nadia are depicted in the top panel and the data for Armondo, are depicted in the middle panel. Nadia and Armondao emitted moderate levels of variability and novelty. Nadia emitted a mean of 6.1 (range: 3-11) different responses within session and 25 novel responses across sessions, with at least 3 novel responses in each session except session 7. Armondo emitted a mean of 7.5 (range: 6-8) different responses within session and 18 novel responses across session, with no novel responses in session 6. In Figure 7, the data for Furio, who emitted high levels of variability and novelty, are depicted in the bottom panel. Furio emitted a mean of 6.7 (range: 2-11) different responses within session and 29 novel responses across sessions, which stopped after session 5.

Figures 8-10 depict data for additional participants who emitted high levels of variability and novelty. In Figure 8, the data for Manlio are depicted in the top panel, the data for Horatio are depicted in the middle panel, and the data for Biagio are depicted in the bottom panel. Manlio emitted a mean of 6.7 (range: 4-14) different responses within session and 33 novel responses across sessions, with at least 1 novel response in all sessions after session 5. Horatio emitted a mean of 6.9 (range: 4-11) different responses within session and 32 novel responses across sessions, with at least one novel response in each session. Finally, Biagio emitted a mean of 8 (range: 3-13) different responses within session and 33 novel responses across sessions, with at least one novel response in all but the last session.

In Figure 9, the data for Bea are depicted in the top panel, the data for Evangelina are depicted in the middle panel, and the data for Mordecai are depicted in the bottom panel. Bea emitted a mean of 8.1 (range: 1-14) different responses within session and 28 novel responses across sessions, with at least one novel response in all but two sessions (sessions 3 and 5). Evangelina emitted a mean of 9.4 (range: 5-14) different responses within session and 28 novel responses across sessions, with one novel response in all but two sessions (sessions 6 and 7). Finally, Mordecai emitted a mean of 10.4 (range: 7-15) different responses within session and 49 novel responses across sessions, with at least three novel responses in each session.

Finally, in Figure 10, the data for Lorena are depicted in the top panel, the data for Kajetan are depicted in the middle panel, and the data for Maria are depicted in the bottom panel. Lorena emitted a mean of 11.5 (range: 7-17) different responses within session and 45 novel responses across sessions, with at least one novel response in each session. Kajetan emitted a mean of 12 (range: 10-14) different responses within session and 43 novel responses across sessions, with at least two novel responses in each session. Finally, Maria emitted a mean of 13.7 (range: 10-19) different responses within session and 61 novel responses across sessions, with at least four novel responses in each session.

Taken together, the majority of children in Study 1 emitted either low or high variability. Specifically, 12 children emitted low variability and novelty. That is, they emitted an insufficient number of different responses in each session to, on average, meet a lag 4 schedule and less than 12 novel responses (i.e., less than 5% of total possible responses out of 256). Two children emitted moderate variability and novelty. These children alternated between a sufficient and insufficient number of different responses in each session to meet a lag 4 schedule and 13 to 25 responses novel responses (i.e., 5% to 10% of total possible responses out of 256). Finally, 10 children emitted high variability and novelty. That is, they engaged in a sufficient number of different responses in each session to, on average, meet or exceed a lag 4 schedule and greater than 25 responses novel responses (i.e., more than 10% of total possible responses out of 256; See Table 1). Interestingly, this would suggest that in many cases, the effects of variability-independent consequences are mostly binary – these consequences will either produce low variability and low novelty or high variability and high novelty. In addition, there was significant correspondence between within-session response variability and across-session novelty, although it is unclear how or why a variability-independent schedule produced such level of novelty. One possibility is the effects of adventitious reinforcement of variability. For example, the majority of the children emitted at least two different responses within session. Therefore, it is possible that although reinforcement was delivered for responding per se, rather than form, this reinforcement after changing form selected changing responses across trials (which could be seen most clearly with the children who emitted high variability). Regardless, for some individuals, a variability-independent schedule may be sufficient for producing higher levels of variability and novelty (i.e., levels of variability and novelty that are expected under leaner lag schedules, such as a lag 4). This is a particularly promising finding because these variability-independent schedules are much simpler to implement (i.e., one must only attend to responding per se rather than the form of responding) and if faded to an intermittent schedule (thus allowing extinction to also induce variation), could be very useful in maintaining high levels of variability. It is also possible that these children would have emitted this pattern of responding in the absence of reinforcement for responding (i.e., extinction could induce this level of variability); however, it is unlikely that this responding would maintain over a long period of time.

Given that the majority of children emitted either low variability and novelty or high variability and novelty, identifying shared characteristics of children in each group could be useful in quickly assessing the necessity of variability-dependent schedules in producing high levels of variability and novelty. First, there were no systematic differences between the two groups based on exposure to a differing number of sessions. In the low variability and novelty group, only 6 sessions were conducted with 2 of 12 children and in the high variability and novelty group, only 6 sessions were conducted with 3 of 10 children. Across all children in the low variability and novelty group, the mean number of sessions was 8.5 and across all children in the high variability and novelty group, the mean number of sessions was 8.2. In addition, 8 out of 10 of the children in the high variability and novelty group met the high novelty criterion (more than 25 different responses across sessions) on or before session 6. Finally, there were significant differences (see Table 1) between the mean number of novel responses in the low variability and novelty group (range: 0.25 – 1.3) and the mean number of novel responses in the high variability and novelty group (range: 2.8 – 10.2).

There were also no systematic differences between groups with respect to age or gender (see Table 2). The mean age of the low variability and low novelty group, without two of the three children with autism (Nico and Carlo, who were of a significantly older age due to their clinical need for intensive behavioral intervention and thus were able to remain enrolled at an age well above that of their typically developing peers) was 3.9 (range: 2.7 – 4.8). Including these two children with autism, the mean age of the low variability and novelty group was 4.2. The mean age of the high variability and novelty group was 3.6 (range: 2.9 – 4.6) and nearly half of the total participants of each gender were in the low variability and novelty and high variability and novelty groups (there were more males than females in Study 1).

There were also no systematic differences between children with IDD and typically developing children (see Table 2), although the sample in this study was quite small (four children). Overall, three children with IDD emitted low variability and novelty, and one child with IDD emitted high variability and novelty. However, Carlo, who was in the low variability and novelty group, emitted the highest level of variability of any child in the low variability group and met one of the two criteria to be included in the moderate variability group (mean variability of 5.1 different responses per session, with half of his sessions sufficient and half of his sessions insufficient to meet a lag 4). This was a particularly interesting finding for two reasons. First, given a core feature of autism is repetitive patterns of behavior, it was interesting that some children with IDD who emitted high levels of repetition with other responses (e.g., motor movements, vocal sounds, mands, etc.) did not emit repetition on this simple motor task. That is, if repetition produces some form of automatic reinforcement for these children, repeating a pattern in Study 1 could have increased the amount of reinforcement available for the repetition – both automatic reinforcement and the social reinforcement of praise and food delivery. For example, Carlo and Nico have been observed to emit motor, vocal, or both types of stereotypy, on average, during greater than 80% of 5-min intervals across the six hours they attend their early intensive behavioral intervention program (based on data collected by respective their one-to-one teachers). Second, the majority of applied research on lag schedules of reinforcement has occurred with individuals with IDD (e.g., Esch et al., 2009; Napolitano et al., 2010; Susa & Schlinger, 2012), which would suggest that using lag schedules is an important part of intervention for children with IDD. However, the results of Study 1 provide evidence that it may not be the disability per se (i.e., repetition in some response classes does not necessarily predict repetition in all response classes) that necessitates the use of lag schedules but rather a specific history of reinforcement that engenders low levels of variability. For example, for many of the children, I observed higher levels of variability in the first few sessions. Over a very short period of time, however, this variability decreased and I observed a co-occurring decrease in novelty. This would provide evidence that low levels of variability are directly produced by the environment; however, it would be important to specifically evaluate this, perhaps by alternating between variability-independent (i.e., ratio-based reinforcement schedules) and variability-dependent schedules (i.e., lag or percentile schedules).

If it is the case that the environment directly produces reductions in variability and novelty (i.e., reductions are observed regardless of an individual's probability of emitting stereotyped responding with other topographies), it would be important to determine what features of environments promote low levels of variability in many children with IDD. This could allow teachers and clinicians to design environments that prevent the development of low levels of variability, which could have great benefits to these children as they age (e.g., prevent the development of socially stigmatizing stereotypy, reduce the need for future behavior-reduction interventions, thus allowing more exposure to skill acquisition training, etc.). This could also be useful in many preschool settings, as teachers could use this information to design procedures that also help promote high variability (and possibly creativity) for all children, regardless of disability.

Although the majority of children in Study 1 emitted either low variability and novelty or high variability and novelty, it is not clear which of these patterns of responding is the most appropriate. For example, these general patterns provide some information on local "norms" of variability, which could be useful in treatment programs in the center in which the children are enrolled (i.e., teachers and clinicians in the center have some information on what levels of variability should be trained and maintained such that children are responding with similar patterns to their peers). However, these data provide only limited information about what levels of variability and novelty would be useful in environments outside of the center. For example, in many social settings, such as at a party, talking about twenty different topics may be inappropriate - listeners may have difficulty rapidly switching between topics or may wish to talk about a single topic for an extended period of time (e.g., a sports team). Likewise, in other settings, such as an art studio, one may be asked to emit a large number of different and novel responses. In those settings, individuals who emit only low levels of variability and novelty may be viewed as missing some inherent "creativity" or social skills when, in fact, the lack of variable and novel responses is the result of a specific history of reinforcement.

Therefore, it would be important for subsequent research to investigate three areas with respect to levels of variability and novelty. First, researchers should determine what levels of variability and novelty are considered “normal” or “appropriate” across a wide variety of environments. Second, researchers should determine if changes in reinforcement contingencies can, over shorter periods of time, produce increases and decreases in variability and novelty. Finally, researchers should determine if these schedules can then be correlated with antecedent stimuli to produce real-time changes in levels of variability and novelty (i.e., to mirror what must occur in many social settings, such as when one is at a party and often must repeat a simple greeting response but then must be able to talk about varied, and sometimes novel, topics, depending on the other person(s) in the conversation).

Study 2: Response Variability Evaluation

Purpose

The purpose of Study 2 was to evaluate the effects of different procedures for producing response variability and novel behavior. First, I replicated and extended previous research (e.g., Goetz & Baer, 1973; Grow et al., 2008; Harding et al., 2004; Lalli et al., 1994; Lee et al., 2002; Lee & Sturmey, 2006; Machado, 1992; Miller & Neuringer, 2000; Page & Neuringer, 1985) on the use of extinction and lag schedules for increasing response variability and novel behavior in young children. That is, I first compared the effects of extinction (EXT) and a typical lag schedule (i.e., fixed-lag 1 [FL 1]) to determine the effects on response variability and novel responding. Next, if FL 1 schedules, compared to R 1 schedules, did not produce high levels of variability, or if they produced higher order stereotypy (i.e., low levels of novel responding across sessions, based on the same criteria as in Study 1, regardless of the levels of variability within sessions) with a particular child, I compared the effects of a larger fixed-lag schedule (i.e., FL 4) and a matched variable-lag (VL) schedule (i.e., VL 4) for producing response variability within sessions and novel responding across sessions. In addition, a yoked schedule (i.e., reinforcers delivered on the same trials as those in the immediately preceding session with a lag schedule) was initially used as a control procedure. However, if the yoke schedule did not produce reductions in response variability, reinforcement of repetition (R 1) was used as a control. The inclusion of a VL schedule was to obtain (potentially) the benefits of percentile schedules (i.e., a lag schedule that is sensitive to local changes in probabilities of specific response topographies) without requiring continuous frequency and probability calculations. I also wanted to compare the effects of direct reinforcement of variability with the yoked schedule (the initial control procedure) to determine the "robustness" of direct reinforcement of variability on the production of both within-session and across-session response variability.

Participants and Setting

Participants in Study 2 (see names with asterisks in Table 1) were seven children (Arlo, Imilia, Matilda, Matteo, Furio, Carlo, and Nico) from Study 1 that emitted low levels of variability and novelty, as evidenced by emitting trial-to-trial variability that was typically insufficient to meet a lag 4 schedule (our largest fixed-lag schedule and the mean value of our variable-lag schedule). That is, to be included in Study 2, children had to frequently emit no more than the same four or five responses in a row (e.g., RRRRR, YYYY, GGGG, BBBB, RRRR), which could also be considered higher order stereotypy, such that there were few trials that would meet a lag 4 (or greater) schedule requirement. The setting was identical to that in Study 1.

Materials

Materials during Study 2 were identical to Study 1. That is, during Study 2, the same response card used in Study 1 (see Figure 1), four piles of four different colored blocks, grouped by color, and preferred food items were present in all sessions (except extinction sessions; see Figure 2).

Response Measurement, Interobserver Agreement, and Data Analysis

As in Study 1, data were collected on the number of different response forms and novel responses across sessions. In addition, data were collected on therapist delivery and non-delivery of edible items and praise as in Study 1. Finally, for some participants, data were analyzed with respect to the number of switches and the distribution of responses across response forms. To determine the number of switches made each session, data collectors counted the number of trials in which the response form on a given trial differed from the response form on the preceding trial. For example, if the response form on trial 10 was R R R R, a switch was recorded if the response form on trial 9 was any form except R R R R. Because a determination of “switch” could not be made on trial 1, there were 19 possible switches in each session. To determine the distribution of responses across response forms, the number of trials with each response form was summed. For example, if the response form R R R R occurred on trials 2, 3, 5, 7, 12, 15, 16, 17, and 18; and the response form B B B B occurred on trials 1, 6, 9, 13, the distribution of R R R R would be nine and the distribution of B B B B would be four. The purpose of calculating switches was to graphically display differences in trial-by-trial responding of participants who generated similar totals of variable and novel responses across sessions. That is, for some participants (see below), the number of different responses and novel responses in each session, across R 1 and lag phases, was similar or identical; however, their responding in each session, on a trial-by-trial basis, was often very different across R 1 and lag phases. For example, in R 1 phases, a child might emit four different responses across trials 1 – 4 then emit only one of those four responses on trials 5 – 20, whereas in lag phases, a child might alternate between four responses in a reliable pattern for all 20 trials.

For the purpose of procedural integrity, I calculated (based on data collected on therapist delivery of reinforcers) whether the therapist correctly delivered edible items and praise in a particular session. Procedural integrity was evaluated for at least 25% of sessions across all conditions and children. Following sessions in which procedural integrity was calculated, data collectors counted the number of trials in which the therapist correctly delivered or did not deliver an edible and praise. In addition, data collectors counted the number of trials in which the therapist incorrectly delivered or did not delivery an edible and praise. To calculate procedural integrity, the number of trials in which the therapist correctly responded were divided by the total number of trials, and multiplied by 100%. Across all participants, procedural integrity was 100%

Interobserver agreement (IOA) for response form and procedural integrity was determined as was done in Study 1 using the exact-agreement method. IOA was calculated for at least 30% for all participants and was 100% Carlo (39.4% of sessions), Nico (41.5% of sessions), and Miles (30.19% of sessions). IOA was calculated for 30% of sessions for Luigi and was 99.6% (range: 95% to 100%). IOA was calculated for of 37.5% of sessions for Matilda and was 99.7% (range: 95% - 100%). IOA was calculated for 61.6% of sessions for Furio and was 99.9% (range: 95% - 100%). IOA for the procedural integrity measure was also calculated and was 100% for all participants.

As in Study 1, data were analyzed to determine response variability within session and response novelty across sessions. This allowed for an evaluation of the effectiveness of each procedure at producing immediate and sustained response variability. Within-session response variability was calculated by determining the number of different responses emitted during that session. For within-session data analysis purposes, the "different response" criterion was session-dependent (i.e., the novelty of a response was based solely on the other responses emitted in the session). For one participant (Nico), sessions were initially videotaped to determine if other dimensions of responding (e.g., serial order of selection and placement of blocks, etc.) would become more variable. However, technical difficulties prevented some sessions from being videotaped or stored for later viewing. Therefore, data collectors were instructed to record other dimensions of responding for all participants (i.e., the typical pattern of responding was established and data collectors recorded deviations) including total session duration. Across all participants, there were no systematic differences in non-targeted dimensions of responding or duration of sessions (i.e., session durations varied but not based on the specific schedules in effect); however, there were some idiosyncratic effects observed in individual participants, which will be discussed below.

Across-session response variability was calculated by determining the cumulative number of novel responses as defined in Study 1 (i.e., a response was considered novel only the first time it occurred). For across-session data analysis purposes, the novelty criterion was both phase-dependent (i.e., the novelty of a responses was based on all the other responses emitted during the phase) and study-dependent (i.e., the novelty of responses was based on all other responses emitted during Study 2). That is, for the phase-dependent requirement, within each phase, once a response was emitted, it was no longer considered novel for the duration of that phase; for the study-dependent requirement, once the response was emitted, it was no longer considered novel for the duration of the study (Study 2), regardless of phase. For example, a response that occurred multiple times per session could still allow for the schedule requirement to be met (i.e., if it was emitted at a frequency that met the lag-schedule criterion). However, the response would only count as novel on a within-phase basis the first time it was emitted in that phase and would only count as novel on a within-study basis the first time it was emitted in Study 2. By including both the phase-dependent and study-dependent criteria, I was able to more precisely determine the parameters of higher order stereotypy (i.e., whether the child was alternating among a larger subset of the response class or alternating among the entire response class). For example, on a phase-dependent basis, a child could appear to be emitting 20 novel responses in each lag phase. However, it could be the case that the child was simply emitting the same 20 responses in each lag phase, which could only be detected by analyzing novel responses on a study-dependent basis.

Procedures

Procedures were similar to those conducted in Study 1. That is, sessions included the same materials and included 20 trials. However, the container of edible items was not present during extinction sessions. In addition, the same rules were provided at the beginning of each session, with the exception of extinction sessions, in which the therapist did not include a statement about the delivery of edible items. In all sessions, if a child asked about edible items (i.e., after not receiving an edible item on a trial), the therapist responded with a statement such as, “Sometimes you get treats and sometimes you don’t.” However, edible items and praise were delivered based on different patterns of responding across different conditions (see condition descriptions below). That is, during a particular phase, if the response met the schedule requirement programmed for that phase, the therapist delivered one edible item and praise with 5 s. If I did not observe increases in variable responding, relative to the preceding R

1 phase, for participants by the end of the first EXT phase, the participant was not included in the remainder of Study 2.

Repetition 1 (R 1). The purpose of this phase was to demonstrate that response variability could be suppressed if reinforcement was made contingent on response repetition. During R 1 sessions, reinforcement (edible item and praise) was delivered for each response that was the exact same as the immediately preceding response. Initially, this phase was alternated with the EXT phase (for control purposes). However, if I did not observe a decrease in both within-session variability and across-session novelty during yoked reinforcement phases (see below), this phase was again implemented and alternated with lag phases for control purposes.

Extinction (EXT). The purpose of this phase was to determine the direct effects of extinction on the induction of variable responses within session and novel responses across sessions. During EXT sessions, there were no programmed consequences for responses and the edible items were not present. No edible items or praise were delivered following a response. Immediately following a response, the therapist cleared the card and started the next trial.

Fixed-Lag reinforcement 1 (FL 1). The purpose of this phase was to determine the effects of a low-value lag schedule on the production of variable responses within session and novel responses across sessions. During FL 1 sessions, reinforcement (edible item and praise) was delivered for any response that differed from the immediately preceding response.

Fixed-lag reinforcement 4 (FL 4). The purpose of this phase was to determine the effects of a higher value lag schedule on the production of variable responses within session and novel responses across sessions. Prior to the start of each session, four pre-session trials were conducted. The purpose of these trials was to provide comparisons such that the lag 4 value on the first four trials would have sufficient preceding comparison trials for the lag value to truly be in effect. That is, without the inclusion of the pre-session trials, either the lag schedule during the first four trials would be different than the lag value on subsequent trials or, if reinforcement was provided for any response on the first four trials, responses repetition could be inadvertently reinforced. During FL 4 sessions, reinforcement (edible item and praise) was delivered for any response that differed from the immediately preceding four responses. For example, if the responses on trials 5, 6, 7, and 8 were RRRR, BBBB, GGGG, YYYY, to produce reinforcement, the response on trial 9 had to be any response other than all red, blue, green, or yellow.

To aid in the rapid delivery of reinforcement, a Microsoft Excel© spreadsheet was programmed to immediately tell the therapist if reinforcement should be delivered on that trial. In this spreadsheet, the therapist entered the response form emitted on a particular trial (e.g., R R R R) in one column and in the immediately adjacent column, either “REINFORCE,” shaded green, or, “NO REINFORCE,” shaded red, was displayed (according to the FL 4 schedule and previous responding by the child). “REINFORCE” was displayed if the response differed from the immediately preceding four trials, and “NO REINFORCE” was displayed if the response was the same as any of the responses on the immediately preceding four trials.

Variable-lag reinforcement 4 (VL 4). The purpose of this phase was to determine the effects of a higher value VL schedule on the production of variable responses within session and novel responses across sessions. During VL 4 sessions, reinforcement (edible item and praise) was delivered for any response that differed from, on average, the immediately preceding four responses. The range of values was 0 (response could be the same as the immediately preceding response) to 8 (response must be different that the preceding eight responses). To determine this range, prior to the start of each session, the therapist generated the lag schedule for each trial, about the mean of 4, using the Flescler and Hoffman (1962) variable-interval progression. This progression was then be randomized using a randomization function of Microsoft Excel©.

Prior to the start of each session, eight pre-session trials were conducted. The purpose of these trials was to provide comparisons such that any lag value on the first eight trials would have sufficient preceding comparison trials for the lag value to truly be in effect. For example, if 7 was the lag value for trial two, one would need six pre-session trials plus trial one for the lag 7 schedule to operate. To prevent conducting varying numbers of pre-session trials, eight trials were always conducted before each session.

To aid in the rapid delivery of reinforcement, a Microsoft Excel© spreadsheet was programmed to immediately tell the therapist if reinforcement should be delivered on that trial. In this spreadsheet, prior to the start of the session, the therapist entered the randomized sequence of lag values (calculated as described above) and adjusted the formula for each trial (i.e., to adjust the comparison range in the formula for each trial). Then, on each trial, as in the FL 4 phases, the therapist entered the response form (e.g., R R R R) in one column and in the immediately adjacent column, either “REINFORCE,” shaded green, or, “NO REINFORCE,” shaded red, was displayed. “REINFORCE” was displayed if the response differed from the relevant number of immediately preceding trials, and “NO REINFORCE” was displayed if the response was the same as any of the responses on the relevant number of immediately preceding trials.

Yoked Reinforcement (YOKE). The YOKE phase was initially alternated with lag reinforcement phases. The purpose of this phase was to serve as a control for conditions that produce greater within-session and across-session response variability by delivering response-independent edible items and praise on the same trials as those that had edible item and praise delivery in the last session of the immediately preceding lag-schedule phase. To determine the trials on which to deliver edible items and praise, the trials with reinforcement on the last session of the immediately preceding phase were recorded. For example, if the trials with reinforcement during the last session of the immediately preceding lag 1 phase were trials 1-15, 16, 18, and 20, in the subsequent YOKE phase, edible items and praise were delivered on those same trials, regardless of the responses on those trials.

Experimental Design

A reversal design was used to evaluate the effects of the various procedures. The R 1 and EXT phases were alternated before any of the lag schedule phases were conducted. The purpose of this initial alternation was to determine if EXT would induce response variability. If it did not, then subsequent lag schedules would likely not be effective. The YOKE phase was initially to be alternated with the various lag-schedule phases (FL 1, FL 4, VL 4) in order to demonstrate the effect of each procedure on the production of variable responses within session and novel responses across sessions. However, if decreases in variability and novelty during did not occur during YOKE phases, the R 1 phase was alternated with the lag-schedule phases for the purpose of control. The order of presentation of the fixed and variable procedures was FL 1, FL 4, and VL 4.

The EXT phase was presented before all other phases, due to concern that children may have stopped assenting to sessions due to the absence of reinforcement. This could have been particularly problematic if EXT phases were implemented after lag-schedule phases (i.e., the child may not contact high levels of reinforcement in lean lag-schedule phases). That is, if children contacted little to no reinforcement, the sessions may have become “aversive” and the children may have begun to refuse to come to sessions. In addition, if the children did not contact reinforcement for placing blocks on the card after they had a history of earning reinforcers for placing blocks on the card, they may have begun to emit incompatible responses during sessions (e.g., moving around the room, stacking blocks, etc.).

Results and Discussion

Arlo and Imilia. The results of the first two phases (R 1 and EXT) for Arlo (top panel) and Imilia (bottom panel) are depicted in Figure 11. During the R 1 phase, Arlo emitted four responses. These four responses were emitted during session 1 and continued to be the only four responses emitted during subsequent R 1 sessions. Therefore, to determine if exposure to extinction would increase variability, Arlo was exposed to the EXT condition. However, during the EXT phase, Arlo emitted the same four responses he emitted during the preceding R 1 phase. This pattern persisted for 7 sessions in the EXT phase. Because Arlo did not emit increased variability or novelty during exposure to the EXT phase, Arlo did not continue in Study 2. Imilia emitted a similar pattern of responding. During the R 1 phase, Imilia alternated between one and four different responses in each session. However, regardless of the number of responses she emitted in the session, she always emitted one of only four responses. During the EXT phase, she reliably emitted four responses in each session. However, these responses were identical to those she emitted in the preceding R 1 phase. Because Imilia did not emit increased variability or novelty during the EXT phase, she did not continue in Study 2.

Furio. The data for Furio in Study 2 are depicted in Figure 12. During the first R 1 phase, Furio emitted a decreasing number of different responses (from three to one) and did not emit any novel responses after session 1. During the first EXT phase, Furio emitted between two and five different responses and emitted at least one novel response, on both a phase and study basis, in each session. During the second R 1 phase, Furio emitted only one response in each session and only emitted one novel response after the first session, the single response he emitted in the third session. During the second EXT phase, Furio emitted a larger number of novel responses (between six and 12 responses) and emitted at least two novel responses, on both a phase and study basis, in each session. Because Furio emitted higher variability and novelty in EXT phases relative to R 1 phases, he was next exposed to the FL 1 phase.

During the first FL 1 phase, Furio emitted an increasing number of different responses within each session (from 10 to 17). In addition, he emitted a steady, increasing number of novel responses on both a phase and study basis across all sessions in this FL 1 phase. In an attempt to demonstrate the effects of contingent reinforcement on production of variability and novelty, Furio was next exposed to the YOKE phase. However, during this phase, Furio continued to emit a similar (relative to the immediately preceding FL 1 phase) number of different responses in each session (between 13 and 18) and an almost identical steady, increasing number of novel responses. These data provide possible evidence of unprogrammed stimulus control or adventitious reinforcement (see discussion below). Therefore, to demonstrate control over variability and novelty, Furio was next exposed to R 1. During this phase, Furio initially emitted a high number of different responses in each session (14 or more in the first eight sessions) and a similar steady, increasing number of novel responses. However, with additional exposure to this schedule, the number of different responses Furio emitted began to decrease, to the point that he emitted three and two different responses in the final two sessions. In addition, Furio stopped emitted novel responses, on both a phase and study basis, in each session, such that he only emitted novel responses in 3 of the final 12 sessions.

During the second FL 1 phase, Furio immediately began emitting a high number of different responses in each session (between seven and 17 in all sessions). In addition, Furio emitted novel responses, on a phase basis, in all but one session (the fourth session in this phase) and emitted novel responses, on a study basis, in 7 of 15 sessions. The majority of these responses occurred at the end of the phase (five of the last six sessions), a pattern that was the opposite of the cessation of novel responses in the final phases of the immediately preceding R 1 phase. To again demonstrate the effects of contingent reinforcement on variability and novelty, Furio was again exposed to R 1. During this phase, Furio initially emitted a high number of different responses in each session and novel responses, on both a phase and study basis. However, following the sixth session, Furio emitted only three and two different responses in the final two sessions. In addition, following session 5 in this phase, Furio stopped emitting novel responses on both a phase and study basis. This pattern of responding, a low number of different responses within each session and little to no novel responses, was identical to his pattern of responding in the previous R 1 phase; however, in the final R 1 phase, this pattern occurred much more quickly (after 18 sessions in the second to last implementation of R 1 and after six sessions in the last implementation of R 1). Because Furio emitted variability in the FL 1 phases that far-exceeded the expected level of variability and novelty that was likely to be produced by either a FL 4 or VL 4 schedule, he was not exposed to any additional lag schedules. Overall, after exposure to the FL 1 condition, Furio was emitting high levels of variability and novelty (based on the criteria from Study 1).

Matilda. The data for Matilda in Study 2 are depicted in Figure 13. During the first R 1 phase, Matilda emitted between one and three different responses. However, across sessions, she emitted the same three responses. During the subsequent extinction phase, there was a significant increase in the number of different responses she emitted in each session, with a range of 5 to 7 responses. In addition, in each session, Matilda emitted at least one response she had not emitted before. During the second R 1 phase, there was a rapid decrease in the number of responses she emitted in each session, ranging from 1 to 4. By the third session in this R 1 phase, she stopped emitting novel responses. During the second EXT phase, there was again an increase in the number of responses she emitted in each session; however, these responses were not entirely novel (she emitted them previously in the study). Because Matilda emitted increased variability and novelty during EXT, she was exposed to the lag phases.

Initially during the first FL 1 phase, Matilda did not display an increase (relative to the preceding EXT phase) in the number of different responses she emitted in each session. However, after the fourth session, she began emitting a higher number (range: 6-10) of different responses within each session. These responses were also novel responses on both a phase and study basis. To determine if the lag schedule in the FL 1 phase directly produced the increases in variability, Matilda was exposed to the YOKE phase. However, although I observed near cessation of novel responses, I did not observe decreases in the number of different responses she emitted in each session. Therefore, to re-establish the contingency between variable responding and reinforcement, I returned to the FL 1 phase. As in the first FL 1 phase, there were sustained levels of different responses in each session, with a range of 5 to 9 responses. However, during the FL 1 phase, I did not observe sustained emission of novel responses. Because the YOKE phase did not suppress variability, for the remainder of Study 2, Matilda was exposed to R 1 phases for demonstrations of control over variability.

During the subsequent R 1 phase, by the third session, the number of different responses emitted by Matilda decreased. In addition, novel responding within the phase stopped after the third session, and novel responding across the study stopped after the second session. Because FL 1 produced only moderate gains in variability and novelty (i.e., based on the criteria from

Study 1), Matilda was exposed to the FL 4 condition. During this condition, I observed an immediate increase in the number of different responses emitted in each session (either 9 or 10). In addition, Matilda continued to generate new responses, on a phase basis, in each session. She also generated novel responses, on a study basis, in every session except session 65 and session 68. When Matilda was then exposed to the R 1 condition, there was an immediate decrease in the number of responses she emitted in each session (to one) and she did not emit any novel responses (on a phase or study basis). During the second FL 4 phase, Matilda emitted an even higher number of responses in each session (range: 10 – 16) and emitted novel responses, on both phase and study bases, in each session. This variability and novelty quickly decreased during the subsequent R 1 phase. To determine if even higher levels of variability could be obtained with a variable lag schedule, Matilda was exposed to the VL 4 phase. During the initial VL 4 phase, Matilda emitted between 10 and 14 different responses in each session (a level seen only in the second FL 4 phase), and with the exception of the second session in the phase, emitted novel responses that had not been previously observed in the study. This high level of variability and novelty was suppressed (within one session) when Matilda was exposed to the final R 1 condition. Upon returning to the VL 4 condition, there was an immediate increase in the number of different responses she emitted in each session (range: 11 – 15), and she more immediately emitted a larger number of novel responses (range: 11-15) that had not been previously observed in the study. Overall, when exposed to either FL 4 or VL 4, Matilda was emitting high levels of variability and novelty (based on the criteria from Study 1).

Matteo. The data for Matteo in Study 2 are depicted in Figure 13. During the initial R 1 phase, Matteo emitted between two and four different responses in each session and stopped emitting novel responses after the fourth session. During the first EXT phase, he emitted a slight increase in the number of different responses in each session (range: 2-6); however, he only emitted novel responses in two sessions (sessions 10 and 15). In the subsequent R 1 phase, the number of different responses Matteo emitted in each session decreased to only two or one and in the following EXT phase, Matteo immediately emitted a higher number of different responses (range: 4-8). Interestingly, during this EXT phase, Matteo emitted at least one novel response, on both a phase and study basis, in each session. Because Matteo was alternating between four and eight responses in each session, the FL 1 schedule was likely to only decrease the number of different responses he emitted in each session (to a level similar to that he emitted in each R 1 session). Therefore, the FL 4 phase was implemented next.

During the first FL 4 phase, Matteo emitted an increasing number of different responses in each session (range: 4 – 9). In addition, in each session, he emitted at least two novel responses on a phase basis and at least one novel response on a study basis. In the subsequent R 1 phase, within one session, Matteo began emitting four or fewer different responses; however, he did not emit any novel responses after the first session in this phase. During the subsequent FL 4 phase, the number of different responses Matteo emitted in each session immediately increased (range: 8-11). He also emitted at least four novel responses on a phase basis and at least two novel responses in each session on a study basis. The eight novel responses he emitted in three sessions during this phase nearly equaled the number of novel responses (10) he emitted in six phases in the first implementation of FL 4.

During the subsequent R 1 phase, there was an immediate decrease in the number of different responses in each session (range: 2-8), and he emitted novel responses on a phase and study basis in only two sessions (sessions 38 and 39). To determine if the variable lag schedule would produce higher levels of variability and novelty than FL 4, Matteo was exposed to the VL

4 condition. During the first VL 4 condition, the number of different responses Matteo emitted in each session (range: 4-11) was similar to the number of different responses he emitted in the previous FL 4 phases. However, on both a phase and study basis, during the VL 4 phase, he emitted more novel responses (range: 0-5) and emitted at least one novel response in all but session 44. During the subsequent R 1 phase, as in previous R 1 phases, the number of different responses Matteo emitted in each session (range: 1-9) immediately decreased and he stopped emitting novel responses after the first session. In the second VL 4 phase, Matteo immediately began emitting a high number of different responses in each session (range: 9-15; based on the same criteria as in Study 1). In the three sessions of the second implementation of VL 4, Matteo emitted twice the number of novel responses on a phase basis (20) and the same number of novel responses on a study basis (8) that he emitted in the six sessions of the first implementation of the VL 4 phase. Overall, when exposed to either FL 4 or VL 4, Matteo was emitting high levels of variability and novelty (based on the criteria from Study 1).

Carlo. The data for Carlo are depicted in Figure 15. During the initial R 1 phase (top panel), Carlo emitted four responses in all sessions except session two. During the first EXT phase, as with other participants, there was an immediate increase in the number of different responses Carlo emitted in each session (range: 5-8) and an increase in the number of novel responses on both a phase and study basis (range: 1-4). During the second R 1 phase, there was an immediate decrease in the number of different responses Carlo emitted in each session (4; no range) and a cessation of novel responses. In the second EXT phase, there was an increase in the number of different responses emitted in the session (range: 2-6) and he emitted one novel response, on both a phase and study basis, in sessions 15 and 16. Although exposure to extinction increased the number of different responses Carlo emitted during each session, during the subsequent FL 1, YOKE, FL 1, and R 1 phases, Carlo emitted the same four responses (all red, all yellow, all green, and all blue) in each session. However, his pattern of responding within each session was quite different based on the phase. Therefore, to capture these differences in within-session patterns, the number of switches and his distribution of responses were analyzed. That is, although Carlo was emitting the same four responses regardless of phase, during R 1 phases, he would most often emit the four responses on the first 4 trials and then emit only one response on trials 5 – 20. However, during the lag phases, Carlo would continue to alternate between the four responses across all trials within each session. Therefore, although he always emitted four responses in each session, during lag phases, he sustained emission of these responses and the distribution of trials across the four responses would often change.

The middle panel of Figure 15 depicts the mean number of switches (i.e., the mean number of responses during a phase that were different from the response on the immediately preceding trial). The bottom panel of Figure 14 depicts Carlo's distribution of responses across his most common four sequences (all red, all yellow, all green, and all blue) and all other responses combined in each phase. During the initial R 1 and EXT phases, Carlo emitted a similar number of switches (M=10.3 & 12.7, respectively). During the second R 1 phase, Carlo emitted a mean of 2.3 switches; however, during the second EXT phase, Carlo emitted a mean of 8.3 switches. Carlo's switching pattern continued in the subsequent FL 1 phases. During the first FL 1 phase, Carlo emitted a mean of 14 switches; during the second FL 1 phase, Carlo emitted a mean of 12.7 switches. Although exposure to the YOKE schedule, as with Furio and Matilda, did not produce a reduction in switches (M=14), subsequent exposure to the R 1 condition produced a significant reduction in the number of switches to a mean of 4.7. With respect to response distribution, a similar pattern was observed (bottom panel of Figure 15). During the R 1 phases, Carlo emitted significantly more responses to the all red sequence (M=15.5) as compared to the other three sequences (all yellow, all green, all blue) combined. However, during both EXT and FL 1 conditions, Carlo alternated among his four most common responses (all red, all yellow, all green, all blue) at nearly identical frequencies (i.e., he emitted each response at least once in every session). In addition, during the second FL 1 phase, Carlo emitted responding nearly exclusively to two responses, all red and all blue (the most efficient pattern under FL 1). Upon returning to R 1 phases, Carlo quickly stopped these alternations and began emitting nearly exclusively, or exclusively in some sessions (e.g., sessions 32 and 35), the all red response.

Because there was a change in Carlo's within-session patterns of responding during exposure to FL 1 phases (e.g., he switched more often and responding was distributed more evenly across four responses) relative to the R 1 phase, Carlo was exposed to the FL 4 phase. During the first FL 4 phase, after session four, Carlo began emitting additional responses in each session (range: 4-8) relative to all the preceding phases. In addition, he emitted novel responses, on both a phase and study basis, in two sessions. With respect to switches and response distribution, Carlo emitted a mean of 16.1 switches during this phase and emitted a similar number of each of his top four response sequences in each session. Upon returning to R 1, Carlo immediately stopped emitting novel responses and only emitted four responses in three sessions. During the second FL 4 phase, he immediately began emitting a larger number of different responses in each session (range: 4-13) and emitted novel responses, on both a phase and study basis, in two of four sessions. In the subsequent R 1 phase, Carlo immediately began emitting fewer different responses in each session (range: 1-5) and did not emit novel responses in any session.

Because Carlo emitted a higher number of different responses and novel responses in the FL 4 phases relative to the R 1 phases, Carlo was exposed to the VL 4 phase to determine how the variable schedule would affect variability and novelty. Within one session, Carlo began emitting a high number of different responses in each session (range: 4-11). Carlo also emitted a high number of novel responses on both a phases and study basis (range: 4-7). In fact, the total novel responses in three sessions of this phase matched the total in four sessions in the second FL 4 phase. In the subsequent R 1 phase, Carlo emitted more than one response (all red) in only one session and stopped emitting novel responses. During the second VL 4 phase, Carlo immediately emitted a high number of different responses in each session (range: 10-13) and emitted the highest number of novel responses (i.e., 9 novel responses in only two sessions), on both a phase and study basis. With respect to switches, Carlo emitted a mean of 13 and 13.5, respectively, in the two VL 4 phases. Interestingly, in both VL 4 phases, the majority of Carlo's responses were not from his top four responses, likely because he was emitting a high number of different responses that were also entirely novel. Overall, when exposed to either FL 4 or VL 4, Carlo was emitting high levels of variability and novelty (based on the criteria from Study 1).

Nico. The results for Nico in Study 2 are depicted in Figure 16. During the first R 1 phase (top panel), Nico emitted two responses only in the second session. During the first EXT phase, unlike other participants, significant exposure to extinction (nine sessions) was required to produce increases in both within-session variability (range: 1-5) and across-session novelty. Upon returning to R 1, Nico continued to emit only one or two responses within each session but did emit one novel response, on a study basis, on session 20, and emitted three novel responses on a study basis. During the second EXT phase, however, Nico began emitting additional responses in each session by the fourth session (range: 1-5); however, during this phase, he did not emit any novel responses on either a phase or study basis. Interestingly, this pattern (four or five different responses in each session and no novel responses) continued across all subsequent phases including FL 1 and FL 4 phases. Nico was exposed to the FL 4 phase after the third implementation of R 1 because his responding in the R 1 phase was similar to his responding in the preceding FL 1 phase. That is, if Nico learned to emit a "base" level of variability of four or five responses, a lag 1 schedule would be insufficient (but a lag 4 schedule could be sufficient) to increase this level of variability relative to the level he emitted in the R 1 phase. Because (based on overall sessions values) Nico was not emitting differing levels of different responses and novel responses based on the phase, switching and response distribution was analyzed for Nico.

During the initial R 1 phase, Nico emitted a mean of 0.33 switches per session (middle panel of Figure 16). During the EXT phase, Nico's mean number of switches increased only slightly to a mean of 0.55 switches, which was less than the mean number of switches (0.8) in the following R 1 phase. However, in the second EXT phase, Nico emitted a significantly higher number of switches (M=3.13). During the FL 1 phase, the mean number of switches (M=11.3) again increased. However, the mean number of switches continued to increase in the subsequent R 1 phase (M=15.6) and FL 4 phase (M=17.8). With respect to response distribution (bottom panel of Figure 16), during the first and second R 1 and EXT phases, Nico emitted almost one response exclusively (RYGB), although he emitted a greater number of different responses (range: 1-5) during the EXT phases. However, during the first FL 1 phase, by the third session, Nico began emitting between three and five responses at relatively similar frequencies (R Y G B, all red, all yellow, all green, and all blue). This pattern continued, and became more equalized, during the subsequent R 1, FL 4, and VL 4 sessions. Interestingly, during these phases, the response that was most frequent during the initial R 1 and EXT phases, R Y G B, became the least frequent response of Nico's five most frequent responses. Overall, after exposure to FL 1, which produced maintained levels of variability, exposure to either R 1or leaner lag schedules (FL 4 or VL 4) did not produce systematic changes in Nico's levels of variability and novelty.

Taken together, three participants emitted low "base variability" that was not significantly increased during either EXT (Arlo and Imilia) or exposure to lag schedules (Nico). The remaining four participants emitted increased levels of variability under either FL 1 (Furio) or FL 4 and VL 4 (Matilda, Matteo, and Carlo). In addition, for each participant, the lag schedule that produced the highest levels of variability also produced the highest levels of novelty on both a phase and study basis. There are several important implications of these findings.

First, a somewhat surprising number of participants were insensitive to either EXT or lag schedules of reinforcement. However, interestingly, all three participants reliably emitted the same four response sequences in each session. There are several possible reasons why this may have occurred. First, from the participant's perspective, a response may have consisted of four sequences (i.e., because the participants emitted the same four responses in a reliable order, it is possible that four responses comprised a single response chain). This could account for the inability of the R 1 schedule to suppress this pattern. That is, if the four responses comprised a chain, the R 1 schedule was, in fact, producing repetition of the chain rather than individual responses. In this case, a participant was only exposed to five "trials" of extinction (five sets of four sequences were possible in the 20-trial session). This would approximate a fixed-ratio (FR) 6 schedule, which could be considered a dense schedule. Thus, the corresponding limited exposure to extinction under even a leaner lag schedule (e.g., FL 4 or VL 4) may have been insufficient to induce variability. A second possibility is that switching from one response to the other was more reinforcing than earning the edible item and praise. Occasionally, the children would make comments about the design made or what design he or she would do on the subsequent trial (e.g., "First, I'm going to make a red necklace, then I'm going to make a green necklace." "It is nice to share with all the blocks."). This provides evidence that some form of the response itself (the specific order of blocks or the alternation between blocks) or even self-generated rules about how to respond (i.e., based on a self-generated game or imaginary play and a strong history of reinforcement for following rules) was influencing responding.

The three children for whom there was not increased variability under EXT or lag schedules also reliably responded in a specific order. For example, children responded with all one color with the color changing based on the lateral position (left to right) of the blocks prior to each trial. Although this pattern is what is typically considered higher-order stereotypy, it may also be, in some ways, simpler for experimenters and clinicians to alter. For example, if a child reliably emits responses based on lateral position, the simplest intervention would be to randomize the position of blocks either after each instance of a response or on some intermittent schedule. In addition, this may be a better approximation of the natural environment. That is, in many environments (e.g., a school), little attention is directed to the exact placement of individual items (e.g., crayons placed on the table). Therefore, the simple, "behavior-as-usual" approach may be sufficient to induce and maintain variability in children who are less sensitive to EXT and lag schedules. In addition, if this simple environmental arrangement is sufficient to reduce or eliminate the higher order stereotypy, this may allow for lag schedules of reinforcement to "capture" this variable responding and thus produce increases in variability.

Although EXT and lag schedules were insufficient to increase variability for three children, lag schedules were sufficient to increase variability for four children. Interestingly, for one child (Furio), a lag 1 schedule was sufficient to produce very high variability (based on the criteria from Study 1) and the highest novelty, on both phase and study bases, of all participants in Study 2. This finding was particularly interesting because of the findings of previous researchers (e.g., Lee et al., 2002; Napolitano et al., 2010) of higher order stereotypy under lag 1 schedules. Although alternating between only two responses is the most efficient method to maximize reinforcement under a lag 1 schedule, this response pattern requires significant attention to how each response produces differential consequences. If a child is unable (or unwilling) to attend to these response-consequences relations, another, potentially simpler method to maximize reinforcement is to "randomly" select responses. This "random" responding may account for the high number of novel responses Furio emitted (over 50% of the total possible response, 256) across the entirety of Study 2. Furthermore, random responding is suggested in that there was no detectable pattern to his responding across trials (e.g., he was not selecting or placing blocks based on lateral position, pairs of colors, etc).

Despite the lack of higher order stereotypy under lag 1 with Furio, Matilda and Carlo did emit mild higher order stereotypy during FL 1 (i.e., they alternated among a small number of responses, particularly compared to responding under FL 4 and VL 4). First, Matilda emitted a mean of 6.3 (range: 3-10) different responses across both FL 1 phases and a total of 21 novel responses across both FL 1 phases. However, she only emitted four novel responses in the second FL 1 phases. Therefore, particularly in the second FL 1 phase, although she emitted, on average, 6.3 different responses in each session, she was reliably emitting the same responses in all sessions during this phase (i.e., she varied, moment-to-moment, but one could predict which response would occur based on the preceding responses). Second, by the end of the second FL 1 phase, Carlo was reliably alternating between two responses, all red and all blue (see middle panel of Figure 14). However, interestingly, for Matilda and Carlo, both FL 4 and VL 4 reduced or eliminated this higher order stereotypy (i.e., a high number of different responses within session and generation of a high number of novel responses, based on the criteria from Study 1, which precluded predicting subsequent responses based on past responses) by producing similar levels of novel responses, on both a phase and study basis. This was a particularly promising finding for a few reasons. First, although FL 4 and VL 4 are leaner schedules than FL 1, they are not significantly leaner than FL 1 and efficient responding that maximizes reinforcement under FL 4 would only require five different responses (which could still be considered higher order stereotypy). This provides some evidence that previous findings of higher order stereotypy may, in fact, be a more direct effect of the specific procedures used in other studies rather than a representation of natural response patterns. That is, individuals may learn to respond to meet the minimum schedule requirement. The minimum schedule requirement under a lag 1 schedule is only two responses (a small number), which may make higher order stereotypy more apparent. Because a lag 1 schedule is one of the most commonly used schedules, the repeated finding of higher order stereotypy may, in fact, be (predominately) due to the lag 1 schedule-value itself. If true, then simple adjustments (i.e., small increases in the lag value) could be made to increase variability. Further, if this finding is replicated across other studies, it could be useful to develop simple assessments to identify what lag value will eliminate this higher order stereotypy. This would present a powerful tool for teaching both creativity and problem solving and treating (or even preventing) the development of stereotypy across a wider number of responses classes in children with a higher (assumed) probability of emitting stereotyped responding (e.g., children with autism).

Second, because the FL 4 and VL 4 schedules produced similar levels of variability and novelty (although VL 4 schedules, such as with Carlo, sometimes produced more rapid increases), caregivers are given a wider variety of options for choosing what procedures to use. For example, in some settings, such as a classroom with one teacher and multiple children, it may be difficult to implement different lag schedules for different children. Therefore, it may be easier to arrange the classroom into teaching trials and implement a fixed lag schedule for all children. In this case, if the teacher use a leaner lag schedule, he or she could be reasonably certain that the schedule will produce little to no higher order stereotypy. However, in other settings, such as an art or music classroom where rapid generation of novelty is of the most importance, a teacher or trainer could choose to use a variable lag schedule to produce a more rapid generation of novel responses. Regardless, this was a promising finding that provides a useful tool for clinicians, particularly because it provides the option of selecting a simpler intervention (the FL 4 schedule) that still produces nearly the maximum result.

Third, for the three participants exposed to the YOKE schedule (Furio, Matilda, and Carlo), removing the contingency between response and consequence was insufficient to produce reductions in variability and novelty. Although this presented a potential threat to demonstration of control over variability and novelty, it did provide initial evidence that lag schedules, like other schedules of reinforcement, may exert some discriminative control over responding. That is, although responses did not differentially affect reinforcement, the delivery of reinforcement after variable responding may have signaled that reinforcement was available for variable responding. If true, this would suggest a very useful method for maintaining variable responding without requiring the (sometimes difficult) implementation of a lag schedule. For example, in a preschool center, where children may be exposed to learning areas specifically designed to teach creative behavior (e.g., an area for dramatic or imaginative play), a teacher may not be able to implement a lag schedule, in real-time, for all children or a different teacher may be unsure if the response he or she observes is different from preceding responses. However, if noncontingent delivery of reinforcement after a history of exposure to lag schedules maintains levels of variability, a teacher could briefly teach each child to behave variably using a lag schedule then switch to the noncontingent schedule to maintain this variability.

The possibility that some form of stimulus control accounted for maintenance of response variability under the YOKE schedule may also be supported by comments and other potentially adjunctive behavior emitted by some of the children across the various phases. For example, in lag conditions, before selecting different colored blocks, Matteo would often state he was making some food item, such as a fruit salad or a pizza. When he placed a specific color on the card, he would then state what ingredient he had included. However, in R 1 phases, Matteo would often make statements about using one color, such as, "green is my favorite color today," or, "yellow is for Cheeto." With respect to other potential forms of adjunctive behavior, during lag phases, Carlo would sometimes "cry" (i.e., make a crying sound without tears) when his response did not produce reinforcement. Likewise, during R 1 phases, when he was selecting the same blocks on each trial, he would often make a repetitive vocal statement (that could not easily be deciphered) or tap his finger near the blocks (most typically the red pile) he would eventually select. Taken together, these potential adjunctive responses may have been methods each child developed to signal what response(s) would produce reinforcement (e.g., make a food item when you need to vary and pick a favorite color when you need to repeat). Although these responses were not systematically investigated, they do provide evidence of potential antecedent sources of control over responding and provide some indirect evidence that maintenance of variability in the YOKE condition was due to stimulus control.

A second possible interpretation of the maintenance of variability under the YOKE schedule is the effects of adventitious reinforcement. If children emit similar patterns of responding across sessions, even if the reinforcement schedule changes, under YOKE conditions, the child would emit a previously established pattern, which would then “produce” reinforcement adventitiously rather than contingently. However, in this case, the contiguity could be sufficient to maintain this pattern. Although the three children did not emit strict patterns of across-trial responding in consecutive sessions and did emit some novel responses in the YOKE phase, they did often emit the same responses across the FL 1 and YOKE phases (i.e., the same responses in different orders). Thus, the contiguity between those specific responses and reinforcement may have been sufficient to maintain variable responding under the YOKE schedule. Regardless of the mechanism of maintenance of variability under noncontingent reinforcement schedules, this finding does present a promising advancement for many caregivers in applied settings interested in teaching creativity and problem solving.

Fourth, there were no detectable changes in other dimensions of responding during lag conditions or repetition conditions that were independent from changes in response variability. For example, there were no systematic increases or decreases in session duration – participants generally completed sessions in the same amount of time across phases (there was some variation that was not phase specific). During Carlo’s first 12 phases (first R 1 phase to second FL 4 phase), Carlo consistently completed sessions in just under 10 min. However, during the final four phases, Carlo consistently completed sessions in over 10 min, with some sessions lasting for just over 20 min. Data collectors noted that during these phases, Carlo had begun to emit a repeated pattern of behavior after he completed his sequence but while he was consuming his edible item (he would close his eyes 15 to 20 s while he was chewing, hum a 5- to 7-s song line once he had finished chewing, and say a phrase from the song). Because this pattern was not phase specific (i.e., he emitted this pattern during both R 1 and VL 4 phases), it is unclear what accounted for this pattern of responding between trials.

In addition, other dimensions of the response itself, such as selection order from the pile of blocks or placement order on the card did not systematically differ from what would be expected when emitting different responses. For example, during repetition phases, children reliably emitted a small number of responses, which limited the selection and placement options (e.g., if one emits the response RRRR, the selection and placement options are quite limited). However, during lag phases, these dimensions co-varied with the different responses. That is, if a child emitted RYGB on trial 1 and BBRR on trial 2, even though the child's placement order may have remained fixed (e.g., laterally from L to R), selection would necessarily vary because of the different colored blocks in the response. This was a particularly interesting finding for two reasons. First, this finding provides evidence that responding under leaner lag schedules is not necessarily more "difficult" that responding under dense lag schedules or repetition schedules. Second, the co-variation in variability between contingent and noncontingent dimensions of the response provides further evidence of the power of lag schedules in producing gains in novelty of responses (i.e., there is greater opportunity to generate truly "creative" responses based on combinations of even noncontingent dimensions).

Finally, rather unexpectedly, levels of novelty, on both a phase and study basis, closely matched levels of variability within session. That is, under FL 1, there were only moderate levels of variability and novelty but under FL 4 and VL 4, there were significantly higher levels of both variability and novelty. This was a particularly promising finding because it provides evidence that lag schedules may be beneficial in producing creativity. Although researchers have reliably found lag schedules to be effective for increasing variability, this variability typically occurs among a small number of responses (applied research) or with long sequences amongst a small number of responses options (basic research). However, in the current study, participants reliably generated novel responses out of a large potential response class (256 responses). Although creativity does have boundary conditions (a painting is limited by materials, canvas size or shape, location, etc.), most consider something creative if it differs from all productions before. In the current study, many of the children emitted response sequences that were not emitted by any other participant. Although the responses in this study were arbitrary and simple, some responses were different from all productions before and after, and thus, in a narrow sense, these responses could be considered creative. If leaner lag schedules continue to be shown to produce novel responses, particularly from response classes with more social significance, this would provide a valuable contribution to the comprehensive science of behavior and evidence that behavior analysis has valuable contributions to make to an area, creativity, traditionally considered to be the purview of other traditions, such as humanism and cognitive psychology (e.g., Hennessey & Amabile, 2010).

Study 3: Stimulus Control Evaluation

Purpose

The purpose of Study 3 was to determine if variable and repetitious responding could come under stimulus control when discriminative stimuli were programmed and rapidly alternated. That is, I wanted to evaluate whether a stimulus associated with reinforcement for response variability and a stimulus associated with reinforcement for response repetition could evoke response variability and response repetition, respectively. First, I replicated previous research (e.g., deSouza et al., 2010; Denney & Neuringer, 1998; Page & Neuringer, 1985; Ward et al., 2008) by bringing the response variability and repetition of young children under stimulus control. Second, I determined whether rapid alternation of these stimuli would allow us to turn variability "on and off." This could be particularly useful in some settings, such as social gatherings, where one must sometimes repeat responses (e.g., a greeting response) but then quickly begin varying responses (e.g., carrying on a conversation).

Participants

The children in Study 3 were three children (Matilda, Furio, and Carol) from Study 2. To participate in Study 3, a child had to display higher levels of variability (there was no fixed criterion) in one of the variability dependent phases as compared to the R 1 phases in Study 2.

Materials

The materials in Study 3 were identical to those used in Study 2 with the exception that a laminated card, with a different color and design, that showed the four squares on which the child would place the blocks was used across different conditions. That is, during variability components, the laminated card was one color and design (e.g., orange with horizontal waves; Figure 16), whereas during repetition components, the card was a different color and design (e.g., purple with vertical waves; Figure 17). The purpose of the different colors or designs was to program stimulus control of variable (i.e., must vary responding when the card is orange with horizontal waves) and non-variable (do not need to vary responding when the card is purple with vertical lines) responding.

Response Measurement, Interobserver Agreement, and Data Analysis

During the stimulus control evaluation, the dependent variables, response measurement, IOA calculations, and data analysis were identical to those in Study 2. IOA was calculated for at least 30% of training sessions and 30% of trials during the stimulus control test. IOA was 100% for all participants.

Procedure

As discussed above, I first paired stimuli with each of the conditions. To do so, I paired the presence of the orange with horizontal-waves card with sessions in which response variability was directly reinforced (using the FL 1 or FL 4 schedule) and purple with vertical-waves card with sessions in which response repetition was directly reinforced (identical to the R 1 sessions in Study 2). These sessions were conducted in the same way that they were in Study 2; however, at the beginning of the session I placed the corresponding stimulus card on the table and told the child that this is either the "purple up-and-down session" or an "orange side-to-side session." In addition, throughout the session, children were instructed to make the same responses (during R 1 sessions) or make a different response (during FL 1/FL 4 sessions), to facilitate rapid development of discriminated responding. Response repetition sessions and FL 1 or FL 4 conditions were implemented using a multielement design, to more rapidly establish the relationship between the antecedent stimulus (card color and design) and the schedule of reinforcement.

After I provided the child with a history (minimum of 6 sessions) of the stimuli associated with the two contingencies, I conducted the stimulus control test. Within the test session, there were 60 trials; however, the two stimuli were alternated within a three-component multiple schedule. That is, the stimulus associated with repetitious responding (the purple with vertical-waves card) was present for the first 20 trials, the stimulus associated with variable responding (the orange with horizontal-waves card) was present for the next 20 trials, and the stimulus associated with repetitious responding (the purple with vertical-waves card) was again presented for the last 20 trials. Across all components, edible items and praise were delivered independent of variability or repetition in responding. That is, reinforcers were delivered based on the trials in which reinforcement occurred during the final R 1 and FL 1 or FL 4 sessions during the stimulus control training (i.e., identical to how reinforcement was scheduled and delivered during YOKE phases in Study 2).

Results and Discussion

The results of Matilda's stimulus control training and evaluation are depicted in Figure 19. The results of the stimulus control training are depicted in the top panel. During training, Matilda emitted only one response in each of the R 1 sessions and emitted 12, 14, and 15 responses in the three FL 4 sessions. Because Matilda emitted discriminated responding (i.e., rapid alternation from repetition to variability on an across-session basis), the training was ended and the stimulus control test was implemented.

The total number of different responses (white circles) Matilda emitted in each component is depicted in the middle panel of Figure 19. During the first R 1 component, Matilda emitted only one response. During the FL 4 component, Matilda emitted a total of 17 different responses. During the subsequent R 1 component, Matilda emitted 7 different responses. To further demonstrate the rapid alternation from repetition to variation and back to repetition, the trial-by-trial data were also analyzed. Switching (i.e., a response is different from the response on the preceding trial) was analyzed to provide a further demonstration that the schedule-correlated stimuli were producing changes in responding (i.e., the R 1 schedule requires little to no switching but the FL 4 schedule requires switching on all trials). The cumulative number of different responses (white circles) and cumulative number of switches (black squares), across trials, in each component, are depicted in the bottom panel of Figure 19. During the first R 1 component, Matilda emitted the same response on each trial and did not switch. During the FL 4 component, on trial 1, Matilda emitted a different response than the response she emitted in the preceding R 1 component; this pattern continued until trial 35 (trial 15 of the FL 4 component), after which she emitted a different response on four of the five remaining trials in the FL 4 component. Matilda also switched on all but four trials (trials 15, 17, 19, & 20) during the FL 4 component. During the subsequent R 1 component, Matilda emitted the same response on the first three trials. She continued to emit a different response on four of the next five trials; however, following this, she only emitted a different response on 2 of the remaining 12 trials. Matilda also emitted a similar pattern of switching in this component – switches on trials 4, 6-8, 12, and 16.

The results of Furio's stimulus control training and evaluation are depicted in Figure 20. During training (top panel), Furio emitted two responses in the first R 1 session and only one response in the two other R 1 sessions. In the FL 1 sessions (FL 1 was used because FL 1 was the only lag schedule Furio was exposed to in Study 2), Furio emitted 4, 13, 5, and 5 different responses in the four sessions. Because Furio emitted discriminated responding, the training was ended and the stimulus control test was implemented.

The total number of responses Furio emitted in each component is depicted in the middle panel of Figure 20. During the first R 1 component, Furio emitted four responses; and during the FL 4 component, he emitted seven responses. During the second R 1 component, he emitted five responses. The cumulative number of different responses (white circles) across trials in each component is depicted in the bottom panel of Figure 20. During the first R 1 component, Furio emitted a novel response on trials 4, 11, and 15. During the FL 1 component, Furio emitted different responses on trials 2, 6, 7, 8, 16, and 18. During the final R 1 component, Furio emitted different responses on trials 3, 5, 12, and 17. Although there was not a large difference between the number of different responses in each component, Furio's switching more clearly demonstrated the effects of the schedule-correlated stimuli. That is, during the first R 1 component, Furio emitted a total of 10 switches that tended to happen in bursts. However, during the FL 1 component, Furio switched on all but 4 trials. Finally, in the second R 1 component, Furio again switched on all but two trials (similar to Matilda, this may have been due to some loss of stimulus control).

The results of Carlo's stimulus control evaluation are presented in Figure 21. During training (top panel), Carlo emitted two responses in the first R 1 session and one response in the other two R 1 sessions. In the FL 4 sessions, Carlo emitted 12, 9, 6, and 10 different responses in the four sessions. Because Carlo emitted discriminated responding, the training was ended and the stimulus control test was implemented.

The total number of responses Carlo emitted in each component is depicted in the middle panel of Figure 21. During the first R 1 component, Carlo emitted only one response; however, during the FL 4 component, he emitted eight different responses and then during the final R 1 component, he emitted only one response. The cumulative number of different responses (white circles) and the cumulative number of switches (black squares) across trials in each component is depicted in the bottom panel of Figure 21. During the first R 1 component, Carlo emitted only one response and did not switch. During the FL 4 component, Carlo begin emitting different responses and switching on the third trial. He emitted different responses on trials 4-6, 9-10, 16, and 20 and switched on trials 4-6, 9-11, 14, 16-18, and 20. During the second R 1 component, Carlo emitted only one response and did not switch.

Taken together, arbitrary stimuli were quickly paired (within seven sessions for all participants) with R 1 and FL 4 or FL 1 schedules. These schedule-correlated stimuli were then effective in producing relatively rapid changes in response repetition and variation.

There are several important implications to the results of Study 3. First, for all participants, repetition and variation were quickly brought under stimulus control. Based on the training results, providing a history of variable and repetitious responding may facilitate subsequent stimulus control of the response patterns. That is, had each participant been initially exposed to the different schedules (lag and repetition) and novel stimuli at the same time, he or she would have to learn two "skills" – responding that meets the schedule requirement and responding in the presence of the correct stimulus. By having a history of each response pattern, each participant had to simply learn when to vary and when to repeat (i.e., turn variability "on and off"). This may be beneficial because it could allow clinicians to address one issue at a time (e.g., alleviating problematic variation or repetition then teach when to vary and repeat), which could result in overall faster terminal performance.

Second, during the test phase, all three participants (relatively) rapidly alternated between variation and repetition. Although Furio did not emit the same level of different responses (which may have been the result of the FL 1 schedule, which does not require as many different responses as a lag 4 schedule) as Matilda and Carlo in the lag component, his behavior (both different responses and switching) did systematically change to "match" the schedule-correlated stimuli. This change, even with different lag values, is particularly useful because in most situations, one is required to quickly "move" back and forth between repetition and variation.

For example, in a jazz performance, one may have an improvised solo that is preceded and followed by playing the set notes of the song. In addition, this solo may be part of a "trading-four" (i.e., one player improvises for four bars and then another player repeats the improvisation) sequence. In both these situations, one must quickly alternate between playing what is written or heard and playing new (or very infrequent) notes. Furthermore, some individuals may have difficulty recognizing social cues as to when to vary and when to repeat. If arbitrary, simple stimuli can be used to produce rapid changes in responding, clinicians would have a very powerful tool to teach the variation/repetition discrimination to these individuals in a manner that is very unobtrusive or even undetectable by others (e.g., a small pager that buzzes, the presence or absence of a bracelet or necklace, etc.).

Finally, although there were an increased number of different responses in the second R 1 component relative to the first R 1 component for Matilda and Furio, the overall total in the second R 1 component was still much lower than the overall total in the FL 4 component. Further, because reinforcement delivery was yoked and reinforcers were delivered on every trial in the final training session, variation sometimes produced what could be considered adventitious reinforcement (i.e., although the food item and praise were delivered noncontingently, they reliably followed the response). Despite this, there were still fewer responses in the second R 1 component than in the FL 4 component. This would suggest that even as stimulus control deteriorates and in the presence of potential adventitious reinforcement, a past history with the stimuli and correlated schedule still exerted the most control over responding. This is a potentially useful finding for many applied settings. For example, in a classroom where a teacher is attempting to increase (or decrease) the variability of responding of multiple students, the teacher may be unable, in a given moment, to accurately record and differentially reinforce repetition. Based on the results of Study 3, however, this noncontingent reinforcement may not entirely disrupt the pattern of responding (i.e., variation may persist even when reinforcement becomes noncontingent), which could be very useful in maintain variable responding while limiting the burden to caregivers. Further, in situations where a teacher is required to implement lag schedules with multiple students, a teacher might inadvertently reinforce a non-variable response (e.g., the teacher observed a response by child A but recorded the response for child B). However, based on the results of Study 3, this reinforcement error may not entirely disrupt the longer-term emission of variable responding, which again, allows the maintenance of discriminated variable and repetitious response less resource-intensive for caregivers.

General Discussion

Based on the results of all three studies, there were three general findings. First, across a large local sample of young children, the majority of children emitted either low variability and novelty or high variability and novelty with only a small number of children emitting moderate variability and novelty. Second, when exposed to a variety of procedures that have been found to affect variability, the variability of three children remained unchanged; however, for four children, leaner lag schedules produced significantly higher levels of variability and novelty than dense lag schedules or extinction. Finally, when one stimulus was correlated with a repetition schedule and a second stimulus was correlated with a variability schedule, rapid presentation and removal of these stimuli produced rapid, real-time changes in variability of responding. Taken together, there are several important implications of these findings and several areas worthy of future research.

First, although a variety of procedures can be effective at increasing both variability and novelty of responding, for some children these procedures may not be necessary. Interestingly, there were no systematic differences, based on age, gender, or developmental status, between children who emitted low variability and novelty and children who emitted high variability and novelty. This was a surprising finding, particularly given that the majority of applied research on variability has focused on evaluating procedures with children with autism (e.g., Esch et al., 2009; Heldt & Schlinger, 2012; Miller & Neuringer, 2000; Napolitano et al., 2010; Susa & Schlinger, 2012). Based on the results of this study, it might be important to place more emphasis on identifying features of the environment that promote either repetition or variability rather than specific features of developmental characteristics of individuals. For example, many children with autism who receive early intensive behavioral intervention are exposed to a consistent schedule and high repetition of teaching procedures and skill programs. It may be the case that this repeated pattern, rather than some inherent property of autism, promotes the development of repetition across a wide-variety of response classes. If true, it may be more effective to focus on preventing the development of this repetition by systematically varying the teaching environment. In addition, in may preschool environments, zone-based instruction may be used much more frequently than tabletop, one-to-one instruction and this form of instruction, with opportunities for both teacher-directed and incidental teaching (e.g., Heal, Hanley, & Layer, 2009), may facilitate (e.g., LeLaurin & Risley, 1972) the development of variable behavior. By investigating the role of specific teaching procedures and arrangements, future researchers may be able to identify specific features of the environment that promote or hinder the development of response variability regardless of the developmental status of an individual. This could lead to environments that promote the long-term development of creativity and problem solving for all children.

A second implication of this series of studies is the moment-to-moment changes between repetition and variation. Across both Study 1 and Study 2, there were periods in which children would repeat responses. This even occurred during FL 4 and VL 4 phases in which children were, on the basis of responding in the entire session, generating many novel responses. Truly random or stochastic responding involves both repetition and variation (e.g., Neuringer, 1986) with responding at one extreme (e.g., repetition in individuals with autism) or the other (i.e., variability in individuals with schizophrenia; Shin et al., 2013; Taylor, Rosenthal, & Snyder, 1967) being problematic. If this is true, reducing repetition and increasing variability in children with autism may simply move responding from one extreme to the other without solving the underlying problem of aberrant response patterns. Therefore, future researchers should investigate the parameters of repetition and variability that occur in typically developing individuals. For example, researchers could conduct large-scale descriptive studies to identify levels of variability on repeated tasks in a variety of settings, including preschools, elementary and middle schools, art and music schools, and the like. By doing so, a general "normal" or typical levels of variability could be identified and clinicians could then more precisely target their interventions to match this level, ensuring that individuals emit responding that is relevant to the environment that they are in and thus likely to be maintained by that environment. For example, if higher levels of variability are required in an art class but lower levels of variability are required in a reading class, clinicians could teach individuals to emit the relevant levels of variability through the use of schedule-correlated stimuli that signal the amount of variability required in that environment. However, until future researchers have identified what levels of variability are appropriate to a variety of environments, a cautious approach should be taken in the development of variability to ensure that one extreme pattern of responding is not simply replaced with another extreme pattern of responding.

A third implication of these studies is the strong antecedent control over repetition and variation that produced both maintenance of variable responding under a yoked schedule in Study 2 and rapid alternation between repetition and variation in Study 3. This was a very promising result for two reasons. First, lag schedules can be difficult to implement, particularly if one is required to implement lag schedules for multiple individuals at the same time. However, if variable responding will maintain under noncontingent reinforcement after a history of exposure to lag schedules, much of the difficulty of reinforcing variability is removed because caregivers can simply deliver reinforcers on a time-based schedule. Second, the rapid development of control over repetition and variation by simple arbitrary stimuli provides a unique tool for teaching an individual to match his or her response pattern to real-time changes in the environment without having to attend to more subtle cues (which for some individuals, such as those with autism, may be difficult to identify). For example, in many social settings, one typically emits the same, or very similar, greeting response to each person he or she meets. However, after this greeting response, one must engage in conversation or activity that can change at any moment (e.g., a new person is added to the group, one moves from one group to another, etc.). Therefore, future researchers should identify how schedule-correlated stimuli can be embedded in a variety of environments to facilitate the rapid alternation between repetition and variation. For example, given the ubiquity of small electronic devices, it may be possible to use these devices to signal when to repeat and when to vary (e.g., one vibration pattern for repetition and another for variation). Future researchers may also consider investigating how these stimuli can be used to promote more molar variation. For example, although one may emit variation in one context, remaining in the context at all times may be considered repetitious (e.g., a person at a party who only plays one game). Therefore, it may be useful to determine if and how these stimuli can promote both molecular (within-the-moment or context) and molar (across moments or contexts) variability.

A final implication of these studies deals with the conceptual status of response variability. First, there has been some debate as to whether variability is a fundamental dimension of operant behavior (e.g., de Souza Barba, 2012; Machado, 1989; Machado & Tonneau 2012; Neuringer, 2012). Based, in particular, on the results of Study 2, in which generation of novel responses occurred most frequently under FL 4 and VL 4, it is apparent that it is the operant itself (i.e., reinforcement is contingent on any response from the functional response class) that is reinforced. This finding also addresses the problem of "response infinity" (e.g., as raised by Machado and Machado & Tonneau), namely, responses will vary based on the already determined size of the response class (in the current study, 256). This could also account for the lack of development of higher order stereotypy under FL 4 and VL 4. That is, because the response class in this study was large (256), the number of responses that can be varied was quite large, and thus a pattern is not as likely to develop because a large number of novel responses can be emitted. Also, the demonstration in Study 3 that response variability and repetition could be (relatively) rapidly alternated provides another demonstration that it is the operant itself, and specifically "sets" of responses in the operant, that is being altered by lag and repetition schedules of reinforcement.

Second, with respect to the relationship between variability and creativity, as was initially made apparent by Goetz and Baer (1973), to teach variable responses and creative behavior, one must monitor responding both in the moment and over time. However, interestingly, in Study 2, levels of novelty closely matched levels of variability – under FL 1, there were low levels of variability and novelty and under FL 4 and VL 4, there were high levels of variability and novelty. Although, in a narrow sense, these novel responses could be considered creative (particularly because children often emitted responses that no other child emitted), creative behavior (and, in may situations, problem solving) is often considered a novel combination of previously learned responses (e.g., Arieti, 1976; Hennessey & Amabile, 2010). However, the development of large, and increasing, response classes may be a critical component in teaching creativity. That is, development of large and increasing response classes will reduce the probability that two responses will be emitted together multiple times, thus allowing for a more continual development of creativity. This could be particularly important in teaching new skills. Often, once a sufficient number of instances of an operant class are emitted, one may conclude that the operant is now in the individual's repertoire and begin to teach other operants (e.g., in early intervention for autism). However, based on the results of the current three studies, to prevent response repetition (which may be more probable with certain populations), researchers should investigate the utility of continuing to teach new instances of the operant once the operant is established in facilitating the subsequent development of creative behavior and problem solving. Additionally, especially with populations prone to response repetition, future researchers should investigate the utility of, from the start of training, bringing variable responses under stimulus control. By doing this, one could further train the individual when to vary and when to repeat, which could limit the need to develop an ever-increasing response class.

Despite these promising results and implications, there were several limitations of this series of studies. First, the response in the series of studies, a series of four blocks, was very simple, arbitrary, and of little social significance. In addition, there were a large number of potential members of the response class (256). Therefore, it is possible that the generation of novel responses was the result of the large potential size of the response class rather than the direct effect of the lean lag schedules. Future researchers should continue to investigate the role of various lag schedules, both dense and lean, on the production of both variability and novelty. Further, future researchers should also investigate how dense and lean lag schedules affect variability and novelty within smaller response classes (e.g., mands, simple art designs, etc.). If the effects of dense and lean lag schedules are, in part, based on the size of the response class, having this information could allow clinicians to select the lag schedule that best matches the response class size and contextually relevant level of variability and novelty needed.

Second, because the response in these series of studies was simple, determining whether the response different from preceding responses (even under the maximum schedule of lag 8) was relatively simple. However, many socially relevant responses are difficult to transduce (e.g., various topographies of verbal behavior), comprised of many components that can change individually (e.g., a sentence in which a change in one word alters the "meaning" of the sentence), or both. Therefore, future researchers should investigate methods to facilitate the transduction of responses and methods to allow for automatic determination of meeting lag requirements. For example, for responses that are relatively simple to transduce, a simple mobile application could be developed to signal to a clinician when a lag requirement has been met (or this capability could be incorporated into currently available mobile applications for behavioral data collection). Further, for responses that have multiple components that are difficult to transduce, such as a variety of forms of verbal behavior, researchers should investigate how automatic transduction can be used to quickly and easily “score” responses (e.g., computerized dictation or computerized photo comparisons).

Third, although the sample of children in Study 1 was relatively large, all children were from the same geographical area and attended the same preschool. Therefore, the levels of variability and repetition may have been idiosyncratic to the similar ages of the children, the specific community, or the type of teaching and social environment the children shared. Future researchers should continue to conduct large-scale evaluations of variability across a wide variety of participants and contexts, such as older individuals, individuals in different types of schools (e.g., public schools, private schools, Montessori schools, etc.), and individuals in different content contexts (e.g., art studios, music classrooms, dance studios, etc.). By doing so, future researchers could develop a fuller understanding of typical levels of variation and repetition in these contexts, which could greatly enhance the social validity of programs designed to increase response variability.

Fourth, in Study 2, for three children, consequent-based procedures were ineffective in altering the child’s variability. Further, there was no analysis of why this occurred or if and how other environmental manipulations could have altered each child’s variability. Because of this lack of control (and analysis) occurred for three out of seven participants in Study 2, the overall conclusions of the primary role of consequent-based strategies in the modification of variability must be tempered. Although this lack of effect may have been idiosyncratic to the design of this study, it is likely that other individuals may also be insensitive to modification of variability by consequences alone. Therefore, future researchers should investigate other methods, such as simple manipulation of response materials (i.e., altering the position of materials), general rules about responding, or specific response feedback, to produce changes in response variability. If other procedures are found to be effective, they could be combined with lag schedules to more rapidly produce changes in variability, produce longer maintenances of these changes, or both.

Finally, although the development of increased variability and novelty in the FL 4 and VL 4 conditions was relatively rapid, exposure to these contingencies was also relatively brief. Therefore, it is unclear if these leaner lag schedules would continue to produce novel responses. This is particularly problematic with smaller responses classes (i.e., in which all possible response may be quickly emitted and thus preclude the emission of novel responses) or highly complex responses (i.e., in which subtle changes may be particularly important but difficult to quantify). First, future researchers should investigate how lean lag schedules affect the long-term emission of novel responses within large response classes (or response classes that can be easily expanded). If these lag schedules do promote long-term emission of novel responses, researchers should then investigate how these lean lag schedules affect the long-term emission of novel responses within smaller response classes. Second, future researchers should investigate how lean lag schedules can be used to promote combinations of responses across response classes in the development of creativity and problem solving. For example, if a response class is small (e.g., using a hammer and nail to join pieces of wood), dense or lean lag schedules may have no differential effect on the production of novel responses. However, it may be the case that lean lag schedules are more effective at promoting the long-term emission of novel combinations of responses (e.g., using a hammer and nail to join different sizes and pieces of wood to create different objects). This would provide strong evidence that lag schedules, which are programmed for variability, can also promote the development of creativity and problem solving.

References

Antonitis, J. J. (1951). Response variability in the white rat during conditioning, extinction, and reconditioning. *Journal of Experimental Psychology, 42*, 273-281.

Arieti, S. (1976). *Creativity: The Magic Synthesis.* New York, NY: Basic Books, Inc.

Cammilleri, A. P., & Hanley, G. P. (2005). Use of a lag differential reinforcement contingency to increase varied selections of classroom activities. *Journal of Applied Behavior Analysis, 38*, 111-115.

Denney, J., & Neuringer, A. (1998). Behavioral variability is controlled by discriminative stimuli. *Animal Learning and Behavior, 26*, 154-162.

de Souza Barbara, L. (2012a). Operant variability: A conceptual analysis. *The Behavior Analyst, 35*, 213-227.

de Souza Barbara, L. (2012b). Variability as a subject matter in a science of behavior: Reply to commentaries. *The Behavior Analyst, 35*, 257-263.

Duker, P. C., & van Lent, C. (1991). Inducing cariability in communicative gesture used by severely retarded individuals. *Journal of Applied Behavior Analysis, 24*, 379-386.

Esch, J. W., Esch, B. E., & Love, J. R. (2009). Increasing vocal variability in children with autism using a lag schedule of reinforcement. *The Analysis of Verbal Behavior, 25*, 73-78.

Ferraro, D. P., & Branch, K. H. (1968). Variability of response location during regular and partial reinforcement. *Psychological Reports, 23*, 1023-1031.

Fleshler, M., & Hoffman, H. S. (1962). A progression for generating variable-interval schedules. *Journal of the Experimental Analysis of Behavior, 5*, 529-530.

Glover, J., & Gary, A. L. (1976). Procedures to increase some aspects of creativity. *Journal of Applied Behavior Analysis, 9*, 79-84.

Goetz, E. M., & Baer, D. M. (1973). Social control of form diversity and the emergence of new forms in children's blockbuilding. *Journal of Applied Behavior Analysis, 6*, 209-217.

Grow, L. L., Kelley, M. E., Roane, H. S., & Shillingsburg, M. A. (2008). Utility of extinction-induced response variability for the selection of mands. *Journal of Applied Behavior Analysis, 41*, 15-24.

Harding, J. W., Wacker, D. P., Berg, W. K., Rick, G., & Lee, J. F. (2004). Promoting response variability and stimulus generalization in martial arts training. *Journal of Applied Behavior Analysis, 37*, 185-195.

Heal, N. A., Hanley, G. P., & Layer, S. A. (2009). An evaluation of the relative efficacy of and children's preferences for teaching strategies that differ in amount of teacher directedness. *Journal of Applied Behavior Analysis, 42*, 123-143.

Hennessey, B. A., & Amabile, T. M. (2010). Creativity. *Annual Review of Psychology, 61*, 569-598.

Holman, J., Goetz, E. M., & Baer, D. M. (1977). The training of creativity as an operant and an examination of its generalization characteristics. In B. C. Etzel, J. M. LeBlanc, & D. M. Baer (Eds.), *New developments in behavioral research, theory, methods, and application: In honor of Sidney W. Bijou* (441-472). Hillsdale, NJ: Erlbaum.

Holz, W. C., & Azrin, N. H. (1961). Discriminative properties of punishment. *Journal of the Experimental Analysis of Behavior, 11*, 285-292.

Joyce, J. H., & Chase, P. N. (1990). Effects of response variability on the sensitivity of rule-governed behavior. *Journal of the Experimental Analysis of Behavior, 54*, 251-262.

Kinloch, J. M., Foster, T. M., & McEwan, J. S. A. (2009). Extinction-induced variability in human behavior. *The Psychological Record, 59*, 347-370.

Lalli, J. S., Zanolli, K., & Wohn, T. (1994). Using extinction to promote response variability in toy play. *Journal of Applied Behavior Analysis, 27*, 735-736.

Lattal, K. A. (1975). Reinforcement contingencies as discriminative stimuli. *Journal of the Experimental Analysis of Behavior, 23*, 241-246.

Lee, R., McComas, J. J., & Jawor, J. (2002). The effects of differential and lag reinforcement schedules on varied verbal responding by individuals with autism. *Journal of Applied Behavior Analysis, 35*, 391-402.

Lee, R., & Sturmey, P. (2006). The effects of lag schedules and preferred materials on variable responding in students with autism. *Journal of Autism and Developmental Disorder, 36*, 421-428.

Lee, R., Sturmey, P., & Fields, L. (2007). Schedule-induced and operant mechanisms that influence response variability: A review and implication for future investigations. *The Psychological Record, 57*, 429-455.

LeLaurin, K., & Risley, T. R. (1972). The organization of day-care environments: “Zone” versus “man-to-man” staff assignments. *Journal of Applied Behavior Analysis, 5*, 225-232.

Lerman, D. C., Iwata, B. A., & Wallace, M. D. (1999). Side effects of extinction: Prevalence of bursting and aggression during the treatment of self-injurious behavior. *Journal of Applied Behavior Analysis, 32*, 1-8.

Machado, A. (1989). Operant conditioning of beahvioral variability using a percentile reinforcement schedule. *Journal of the Experimental Analysis of Behavior, 52*, 155-166.

Machado, A. (1992). Behavioral variability and frequency-dependent selection. *Journal of the Experimental Analysis of Behavior, 58*, 241-263.

Machado, A. (1997). Increasing the variability of response sequences in pigeons by adjusting the frequency of switching between two keys. *Journal of the Experimental Analysis of Behavior, 68*, 1-25.

Machado, A., & Tonneau, F. (2012). Operant variability: Procedures and processes. *The Behavior Analyst, 35*, 249-255.

Maltzman, I., Bogartz, W., & Berger, L. (1958). A procedure for increasing word association originality and its transfer effects. *Journal of Experimental Psychology, 56*, 392-398.

McElroy, E., Neuringer, A. (1990). Effects of alcohol on reinforced repetitions and reinforced variations in rats. *Psychopharmacology, 102*, 49-55.

Miller, N., & Neuringer, A. (2000). Reinforcing variability in adolescents with autism. *Journal of Applied Behavior Analysis, 33*, 151-165.

Morgan, D. L., & Lee, K. (1996). Extinction-induced response variability in humans. *The Psychological Record, 46*, 145-159.

Morris, C. J. (1987). The operant conditioning of response variability: Free-operant versus discrete-response procedures. *Journal of the Experimental Analysis of Behavior, 47*, 273-277.

Morris, C. J. (1989). The effects of lag value on the operant control of response variability under free-operant and discrete-response procedures. *The Psychological Record, 39*, 263-270.

Napolitano, D. A., Smith, T., Zarcone, J. R., Goodkin, K., & McAdam, D. B. (2010). Increasing response diversity in children with autism. *Journal of Applied Behavior Analysis, 43*, 265-271.

Neuringer, A. (1986). Can people behave "randomly?": The role of feedback. *Journal of Experimental Psychology: General, 115*, 62-75.

Neuringer, A. (1991). Operant variability and repetition as functions of interresponse time. *Journal of Experimental Psychology: Animal Behavior Processes, 17*, 3-12.

Neuringer, A. (2012). Reinforcement and induction of operant variability. *The Behavior Analyst, 35*, 229-235.

Neuringer, A., Deiss, C., & Imig, S. (2000). Comparing choices and variations in people and rats: Two teaching experiments. *Behavior Research Methods, Instruments, and Computers, 32*, 407-416.

Odum, A. L., Ward, R. D., Barnes, C. A., & Burke, K. A. (2006). The effects of delayed reinforcement on variability and repetition of response sequences. *Journal of the Experimental Analysis of Behavior, 86*, 159-179.

Page, S., & Neuringer, A. (1985). Variability is an operant. *Journal of Experimental Psychology: Animal Behavior Processes, 11*, 429-452.

Pryor, K. W., Haag, R., & O'Reilly, J. (1969). The creative porpoise: Training for novel behavior. *Journal of the Experimental Analysis of Behavior, 12*, 653-661.

Schwartz, B. (1980). Development of complex stereotyped behavior in pigeons. *Journal of the Experimental Analysis of Behavior, 33*, 153-166.

Schwartz, B. (1982). Failure to produce response variability with reinforcement. *Journal of the Experimental Analysis of Behavior, 37*, 171-181.

Shin, Y. S., Kim, S. N., Shin, N. Y., Jung, W. H., Hur, J., Byun, M. S., … Kwon, J. S. (2013). Increased intra-individual variability of cognitive processing in subjects at risk mental state and schizophrenia patients. *PLOS ONE, 8*, e78354.

Skinner, B. F. (1933). “Resistance to extinction” in the process of conditioning. *Journal of General Psychology, 9*, 420-429.

Souza, A. S., Abreu-Rodrigues, J. (2010). Discriminative properties of Vary and Repeat contingencies. *Behavioural Processes, 85*, 116-125.

Susa, C, & Schlinger, H. D. (2012). Using a lag schedule to increase variability of verbal responding in an individual with autism. *The Analysis of Verbal Behavior, 28*, 125-130.

Tatham, T. A., Wanchisen, B. A., Hineline, P. N. (1993). Effects of fixed and variable ratios on human behavioral variability. *Journal of the Experimental Analysis of Behavior, 59*, 349-359.

Taylor, I. A., Rosenthal, D., & Snyder, S. (1963). Variability in schizophrenia. *Archives of General Psychiatry, 8*, 163-168.

Turner, M. (1999). Annotation: Repetitive behavior in autism: A review of psychological research. *Journal of Child Psychology and Psychiatry. 40*, 839–849.

Valentino, A. L., Chillingsburg, M. A., Call, N. A., Burton, B., & Bowen, C. N. (2011). An investigation of extinction-induced vocalizations. *Behavior Modification, 35*, 284-298.

Ward, R. D., Kynaston, A. D., Bailey, E. M., & Odum, A. L. (2008). Discriminative control of variability: Effects of successive stimulus reversals. *Behavioural Processes, 78*, 17-24.

Tables

Participant	Age	Diagnosis	Variability Mean	Novelty Mean	Variability/ Novelty Level
Nico*	6.5	Autism	1	0.17	Low
Miles*	4	None	2.6	0.25	Low
Bruto	3.8	Autism	2.7	0.86	Low
Imilia*	4.5	None	3.9	0.50	Low
Luigi	4.3	None	4	0.71	Low
Arlo*	4.5	None	4	0.57	Low
Olga	2.7	None	4	0.57	Low
Cassio	4.3	None	4	0.44	Low
Mona	3.1	None	4.3	1	Low
Gastone	3.3	None	4.7	1.1	Low
Matilda*	4.5	None	3.2	1.3 (M)	Low
Carlo*	5.6	Autism	5.1 (M)	1	Low
Nadia	4.4	None	6.1	3.1	Moderate
Armondo	3.7	None	7.5	3	Moderate
Furio*	3.1	None	6.7	2.9	High
Manlio	3.1	None	6.7	3	High
Horatio	3.1	None	6.9	3.2	High
Biagio	4.4	Autism	8	3.7	High
Bea	4.2	None	8.1	4.1	High
Evangelina	3.0	None	9.4	2.8	High
Mordecai	4.3	None	10.4	7	High
Lorena	4.6	None	11.5	7.5	High
Kajetan	3.7	None	12	7.2	High
Maria	2.9	None	13.7	10.2	High

Table 1. Participant characteristics ordered by level of variability in Study 1. The table depicts the age in years of each participant at the start of their participation, any known diagnosis of intellectual or developmental disability, and their level of variability in Study 1. An asterisk (*) denotes participation in Study 2. An "M" next to either a variability or novelty mean denotes the mean value meeting the moderate criteria.

Figures

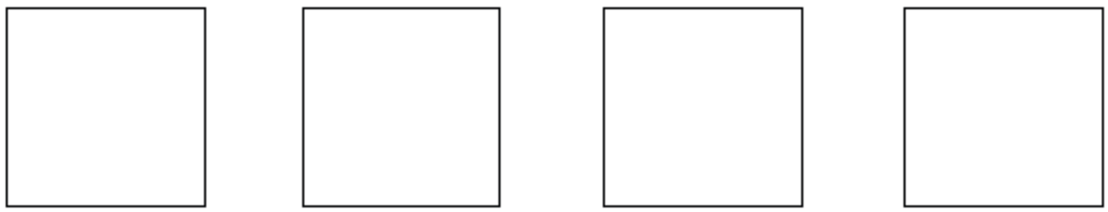


Figure 1. A depiction of the card used in all sessions in Studies 1, 2, and 3.

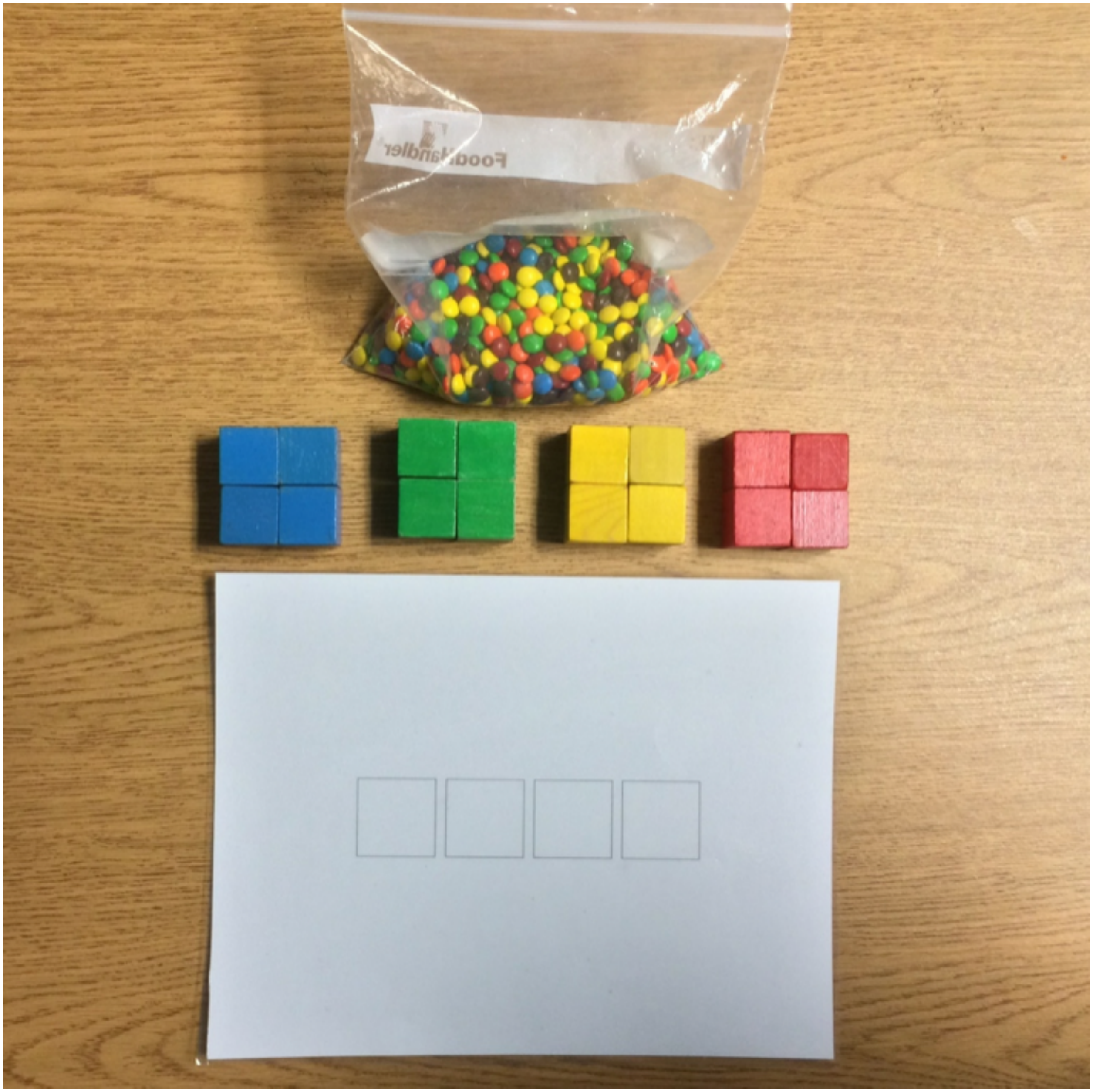


Figure 2. A depiction of session materials from Study 1, 2, and 3.

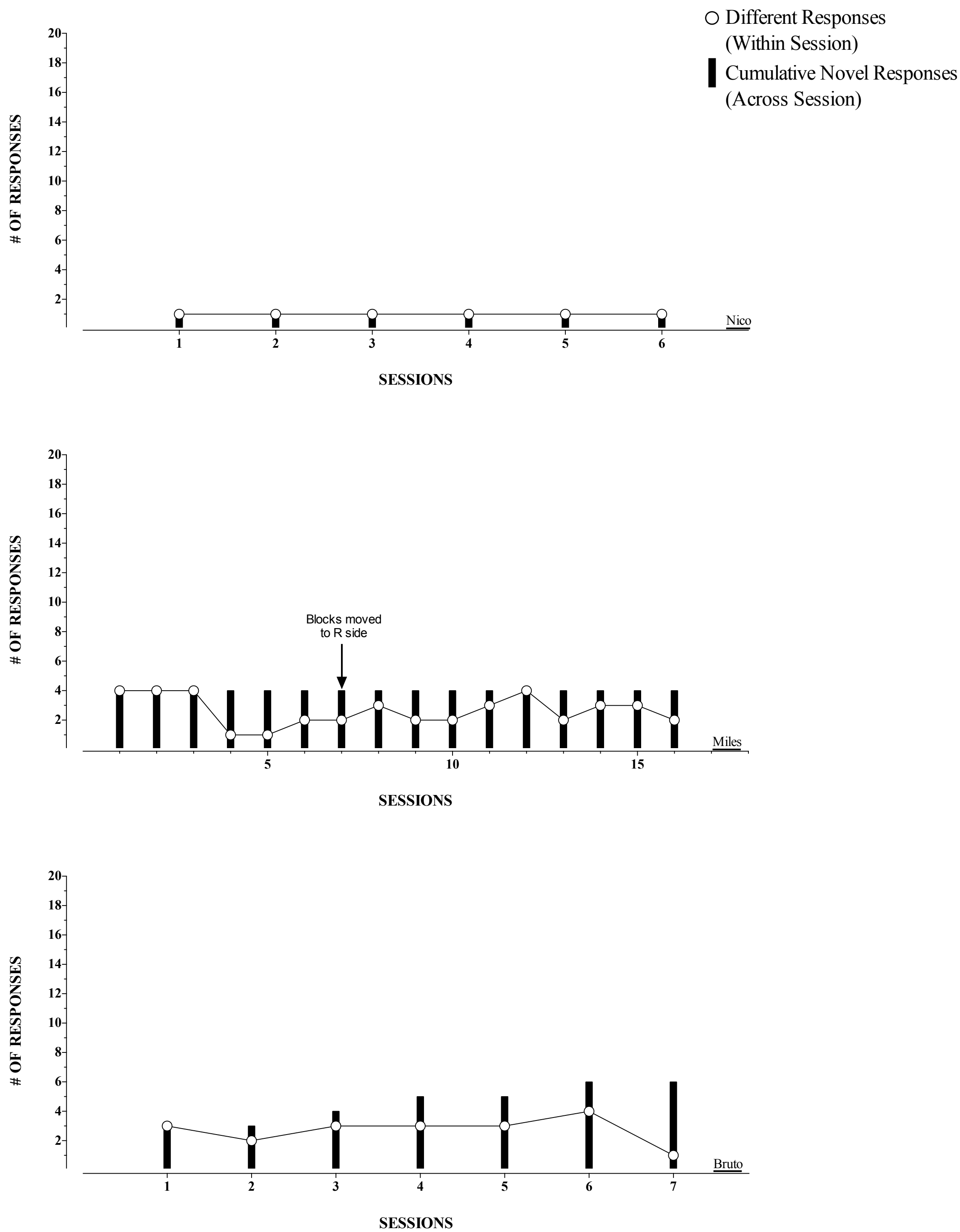


Figure 3. The results for Nico (top panel), Miles (middle panel), and Bruto (bottom panel) in Study 1. Sessions are depicted along the x-axis. The number of different responses (within session) and the cumulative number of novel responses (across sessions) are depicted on the left y-axis.

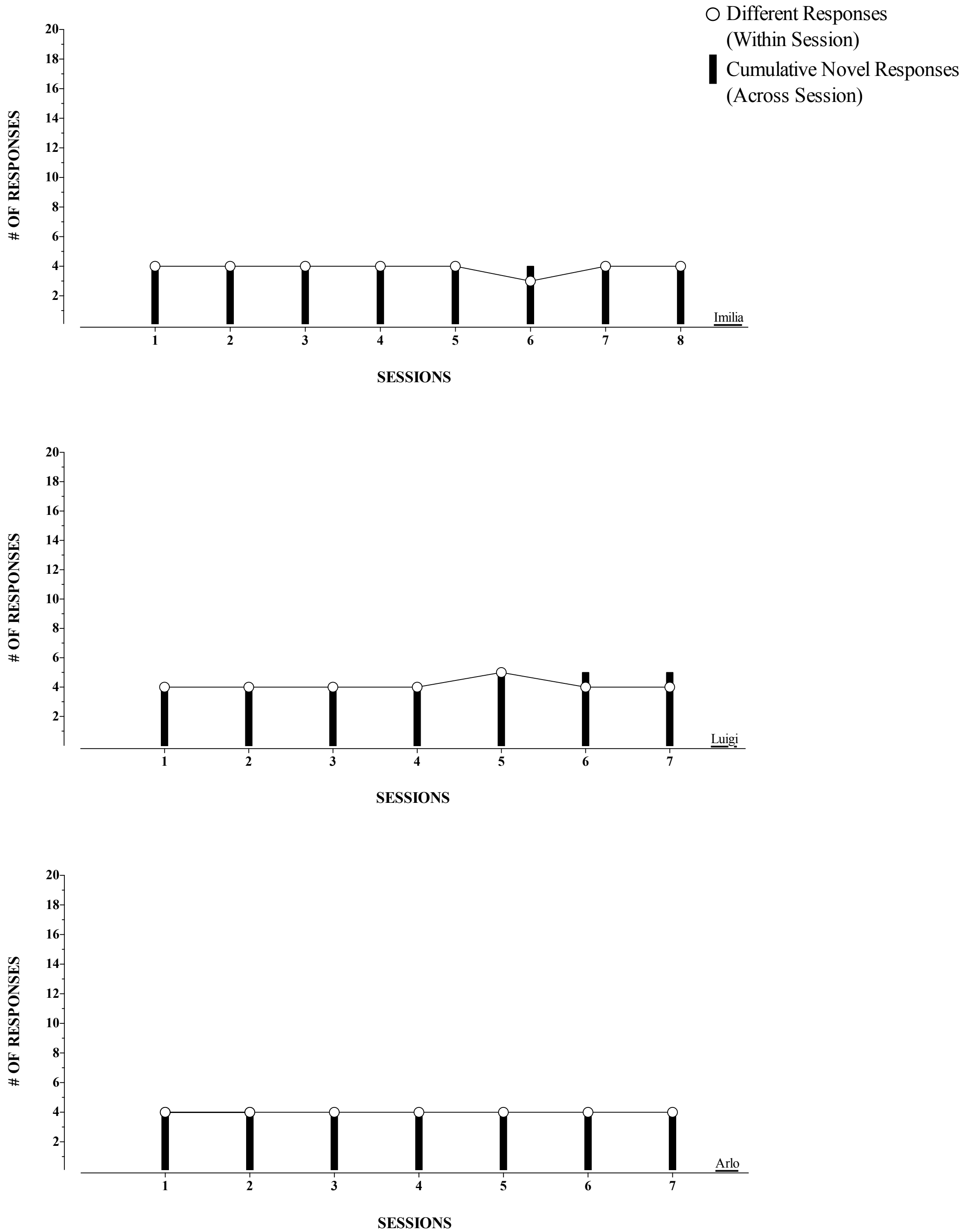


Figure 4. The results for Imilia (top panel), Luigi (middle panel), and Arlo (bottom panel) in Study 1. Sessions are depicted along the x-axis. The number of different responses (within session) the cumulative number of novel responses (across sessions) are depicted on the left y-axis.

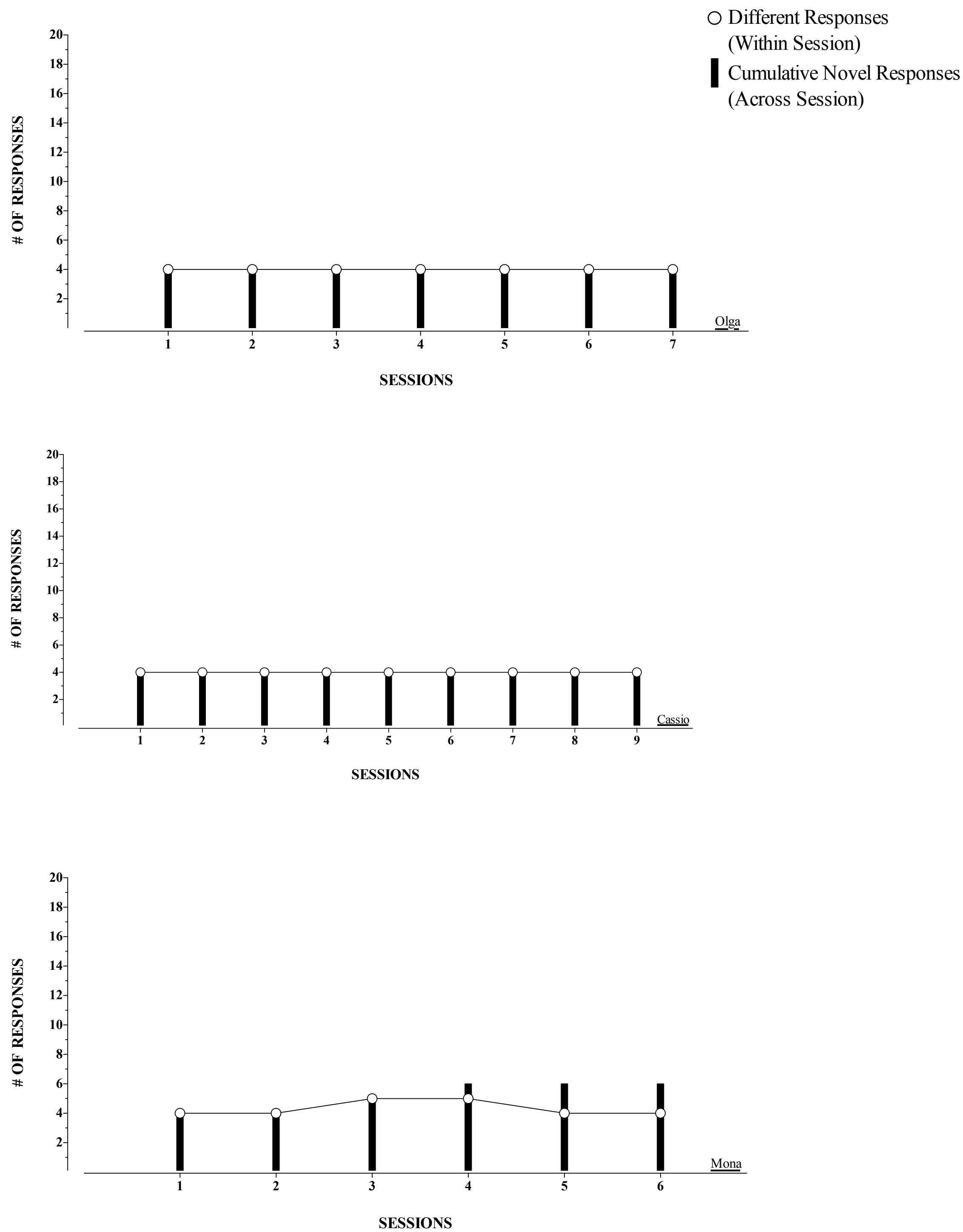
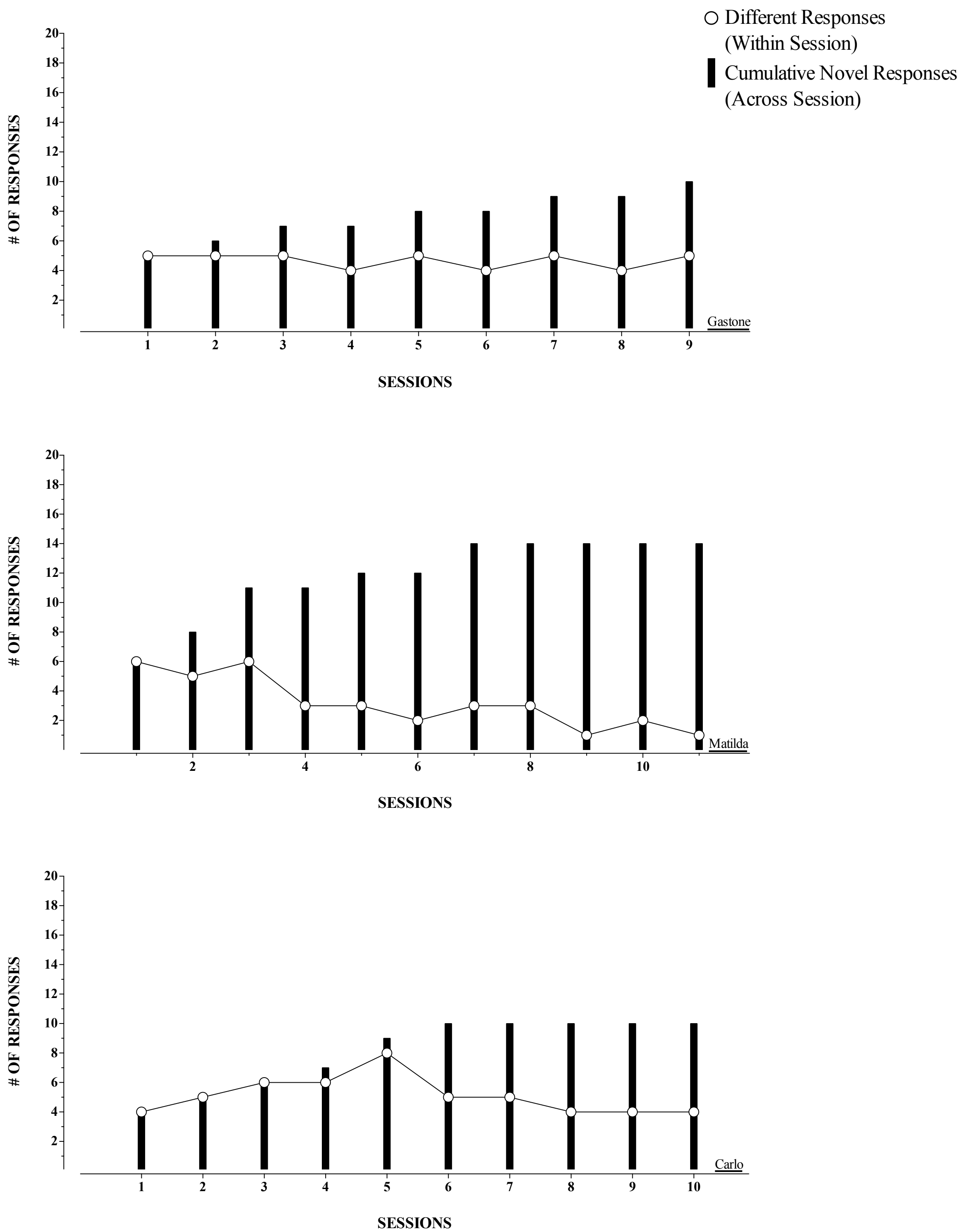


Figure 5. The results for Olga (top panel), Cassio (middle panel), and Mona (bottom panel) in Study 1. Sessions are depicted along the x-axis. The number of different responses (within session) cumulative number of novel responses (across sessions) are depicted on the left y-axis.



*F*igure 6. The results for Gastone (top panel), Matilda (middle panel), and Carlo (bottom panel) in Study 1. Sessions are depicted along the x-axis. The number of different responses (within session) and the cumulative number of novel responses (across sessions) are depicted on the left y-axis.

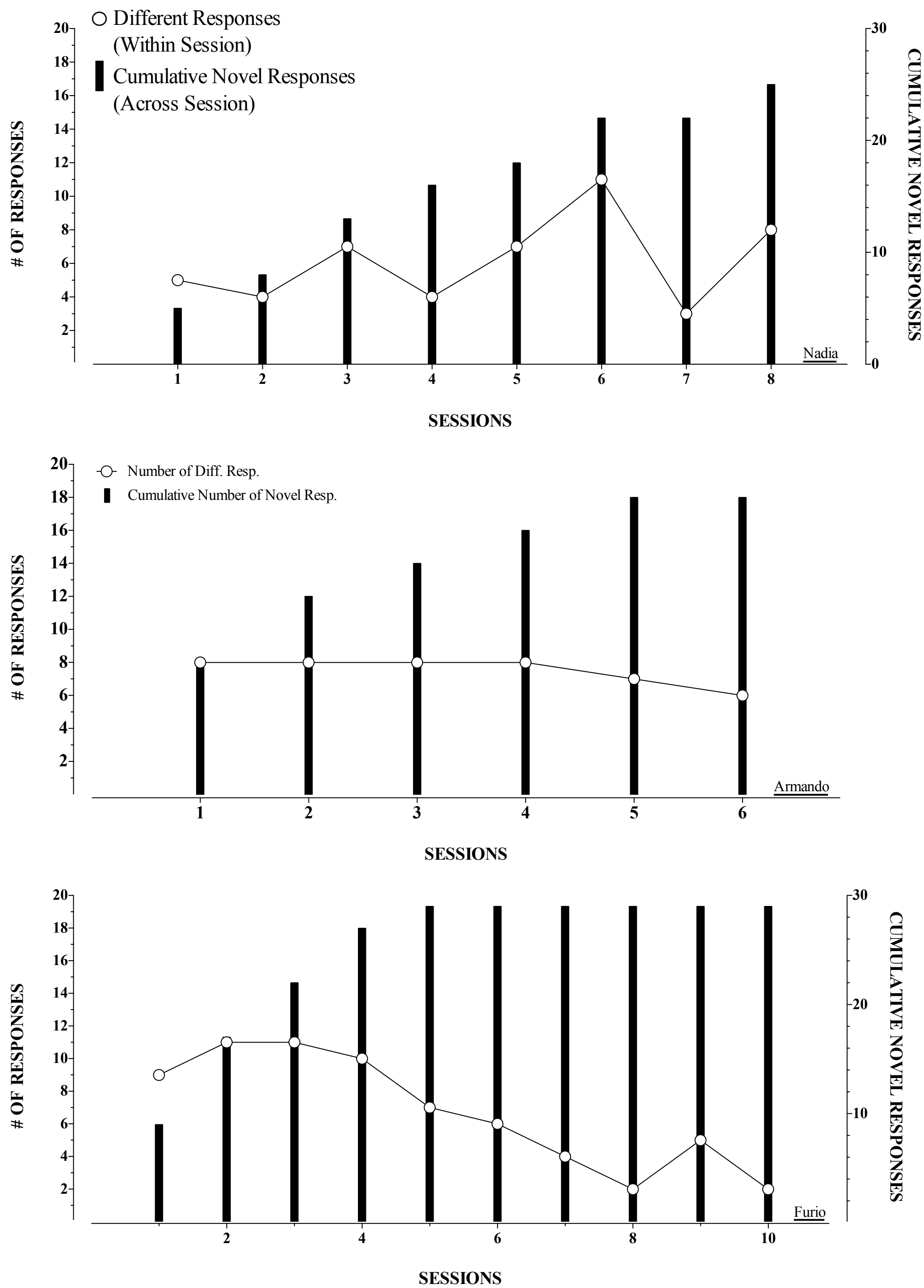


Figure 7. The results for Nadia (top panel), Armando (middle panel), and Furio (bottom panel) in Study 1. Sessions are depicted along the x-axis. In themiddle panel, the number of different responses (within session) and the cumulative number of novel responses (across sessions) are depicted on the left y-axis, In the top and bottom pane, the number of different responses (within session) is depicted on the left y-axis and the cumulative number of novel responses (across sessions) is depicted on the right y-axis.

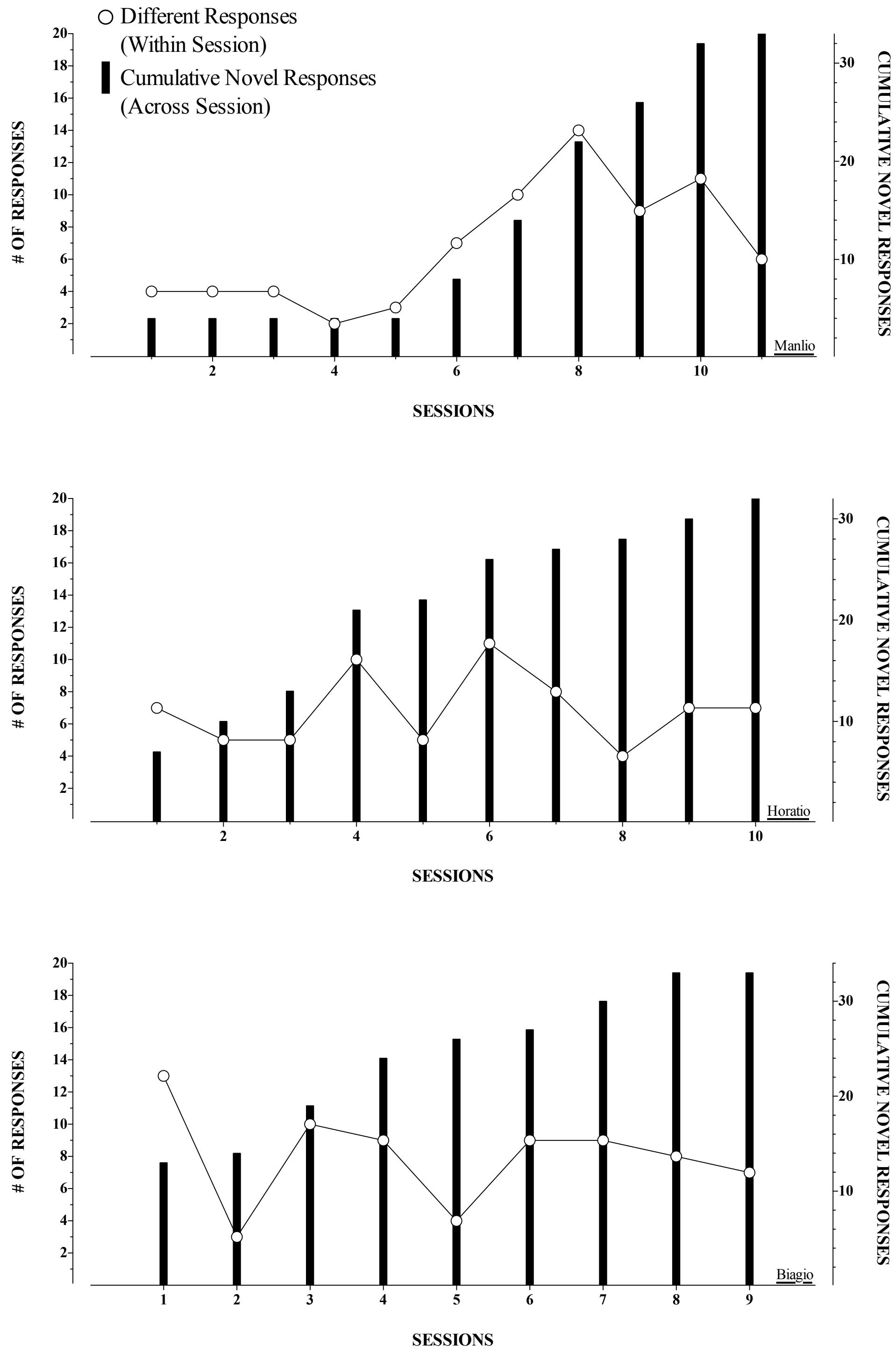


Figure 8. The results for Manlio (top panel), Horatio (middle panel), and Biagio (bottom panel) in Study 1. Sessions are depicted along the x-axis. The number of different responses (within session) is depicted on the left y-axis and the cumulative number of novel responses (across sessions) are depicted on the right y-axis.

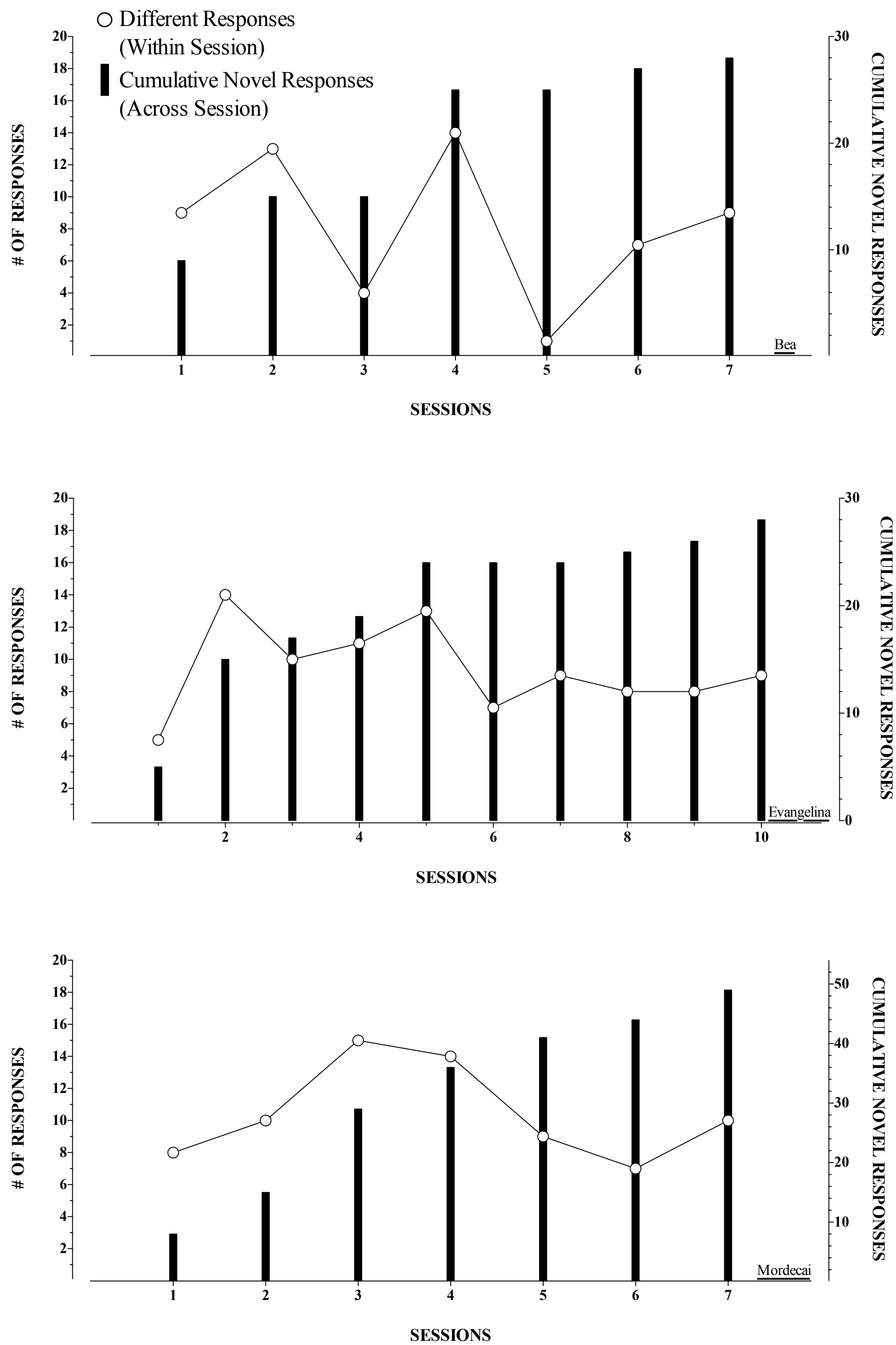


Figure 9. The results for Bea (top panel), Evangelina (middle panel), and Mordecai (bottom panel) in Study 1. Sessions are depicted along the x-axis. The number of different responses (within session) is depicted along the left y-axis and the cumulative number of novel responses (across sessions) is depicted on the right y-axis.

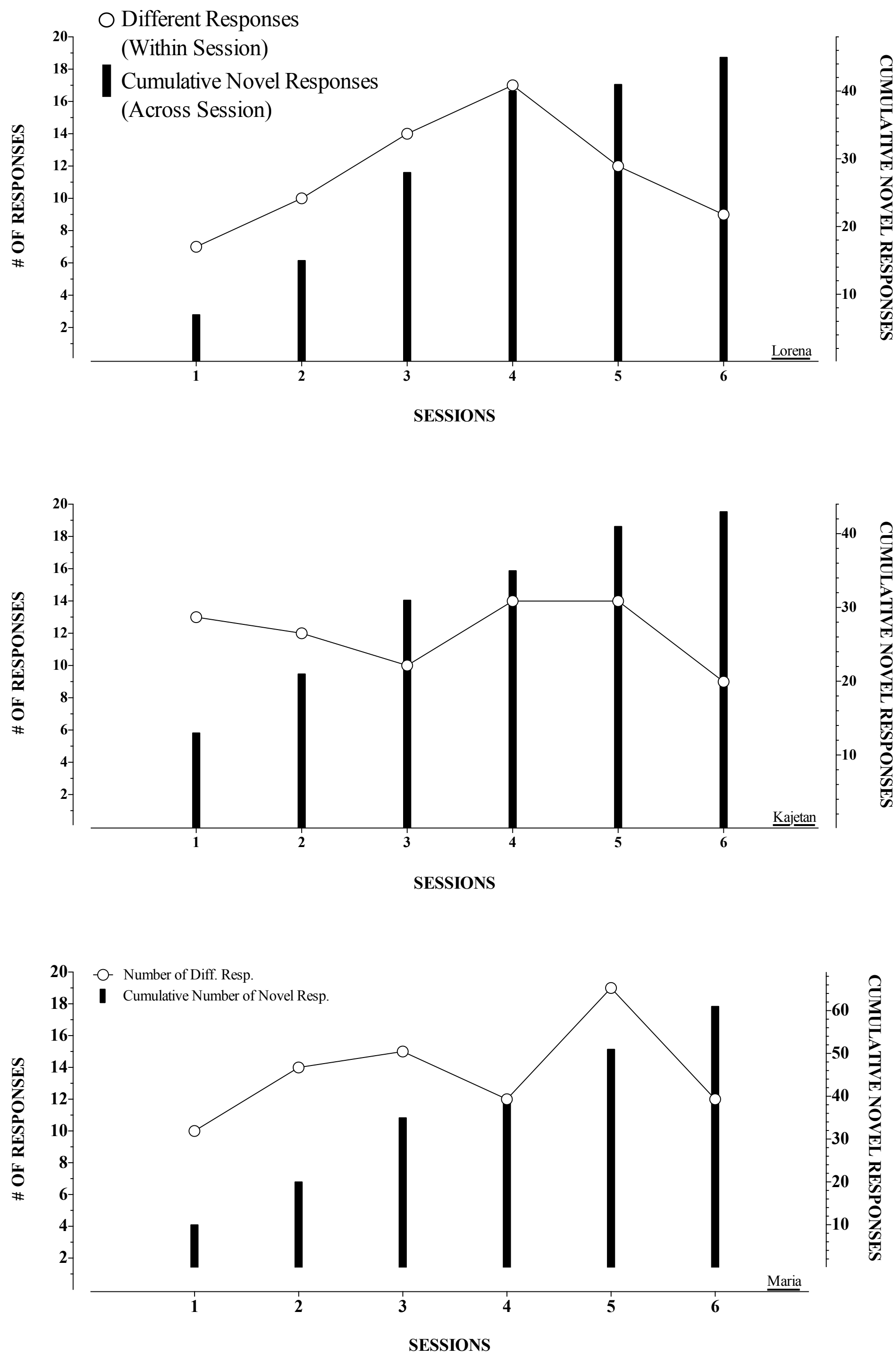


Figure 10. The results for Lorena (top panel), Kajetan (middle panel), and Maria (bottom panel) in Study 1. Sessions are depicted along the x-axis. The number of different responses (within session) are depicted on the left y-axis and the cumulative number of novel responses (across sessions) is depicted on the right y-axis.

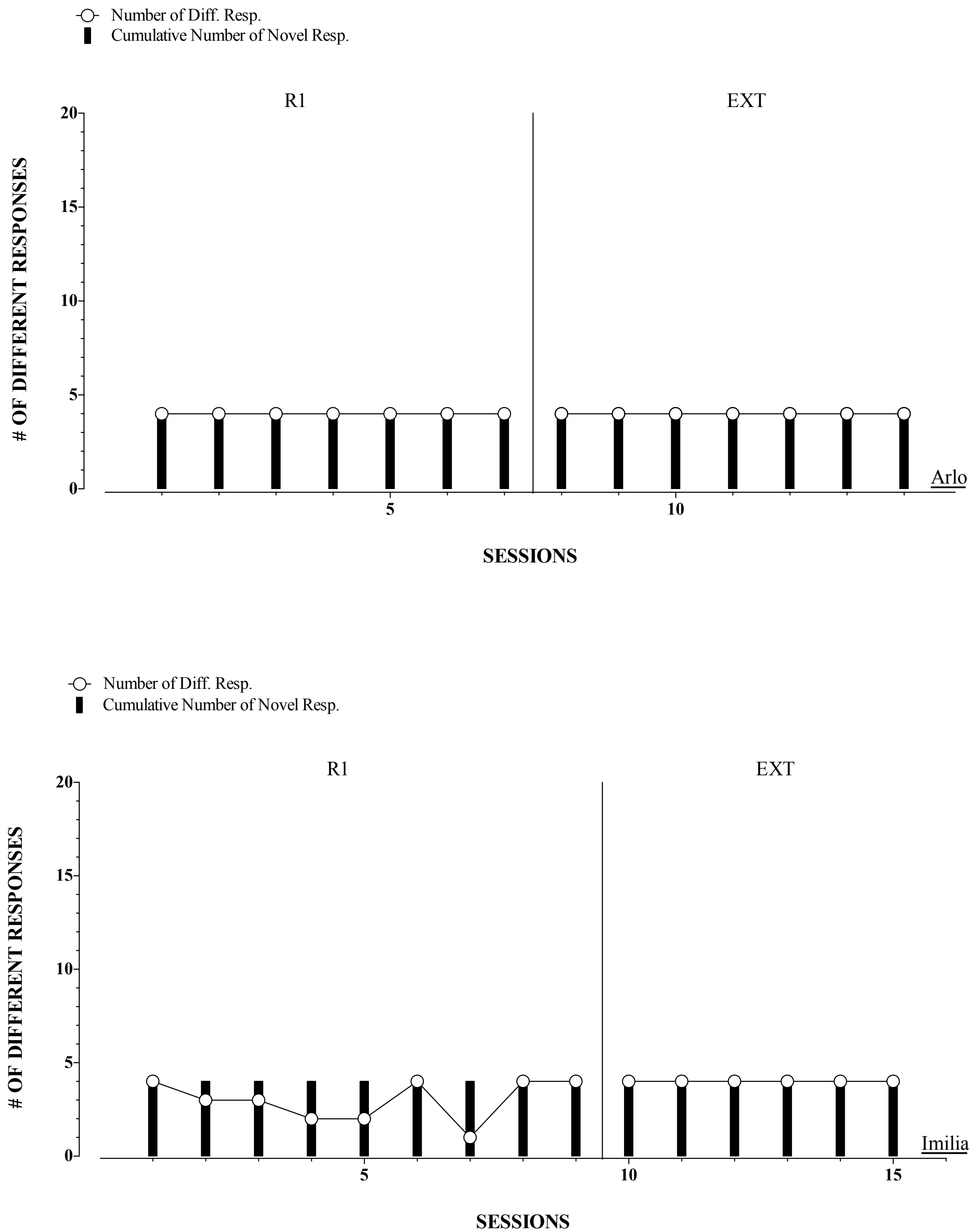


Figure 11. The results of Study 2 for Arlo (top panel) and Imilia (bottom panel). Sessions are depicted along the x-axis. The cumulative number of novel responses is depicted on the y-axis.

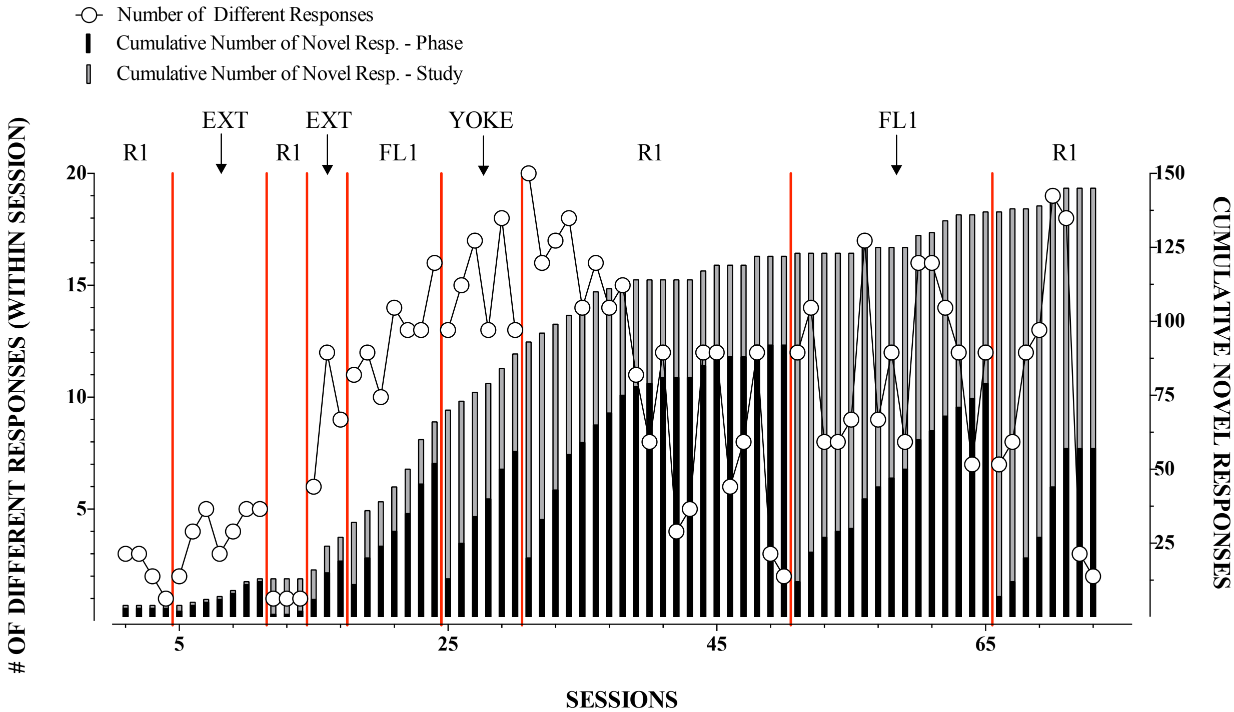


Figure 12. The results of Study 2 for Furio. Sessions are depicted along the x-axis. The number of different responses (within session) is depicted along the left y-axis and the cumulative number of novel responses (phase and study) is depicted on the right y-axis.

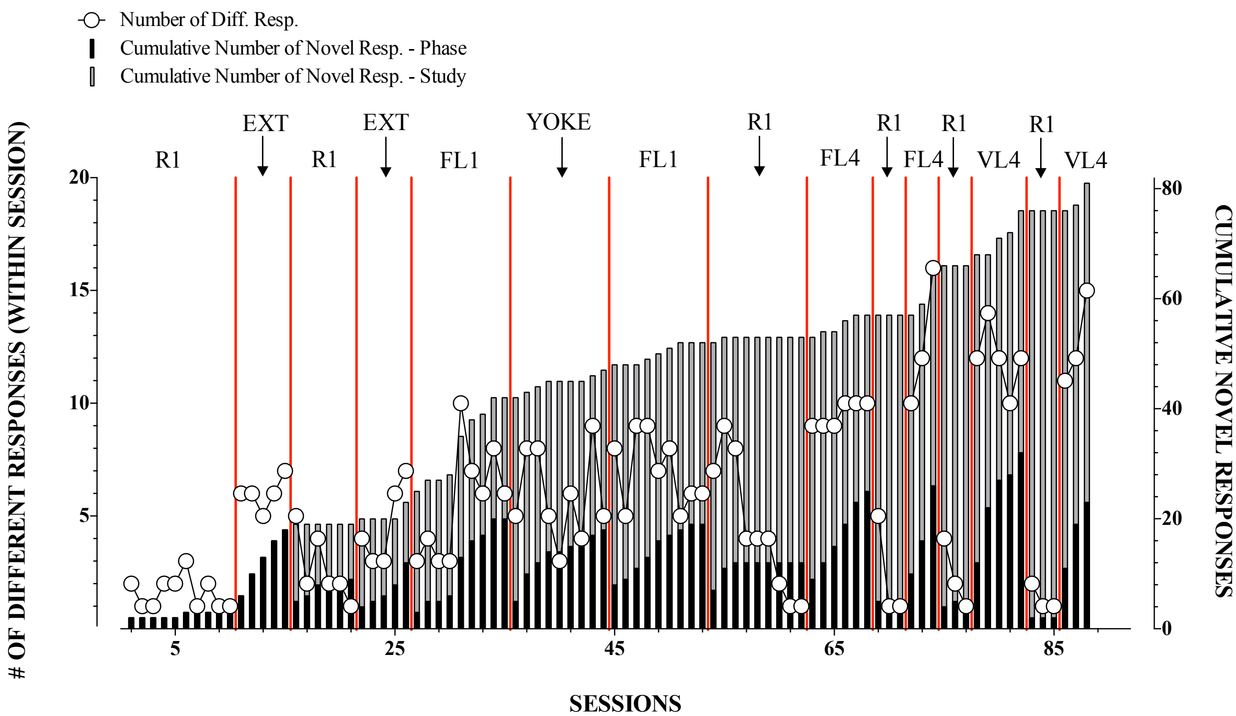


Figure 13. The results of Study 2 for Matilda. Sessions are depicted along the x-axis. The number of different responses (within session) is depicted on the left y-axis and the cumulative number of novel responses (phase and study) is depicted on the right y-axis.

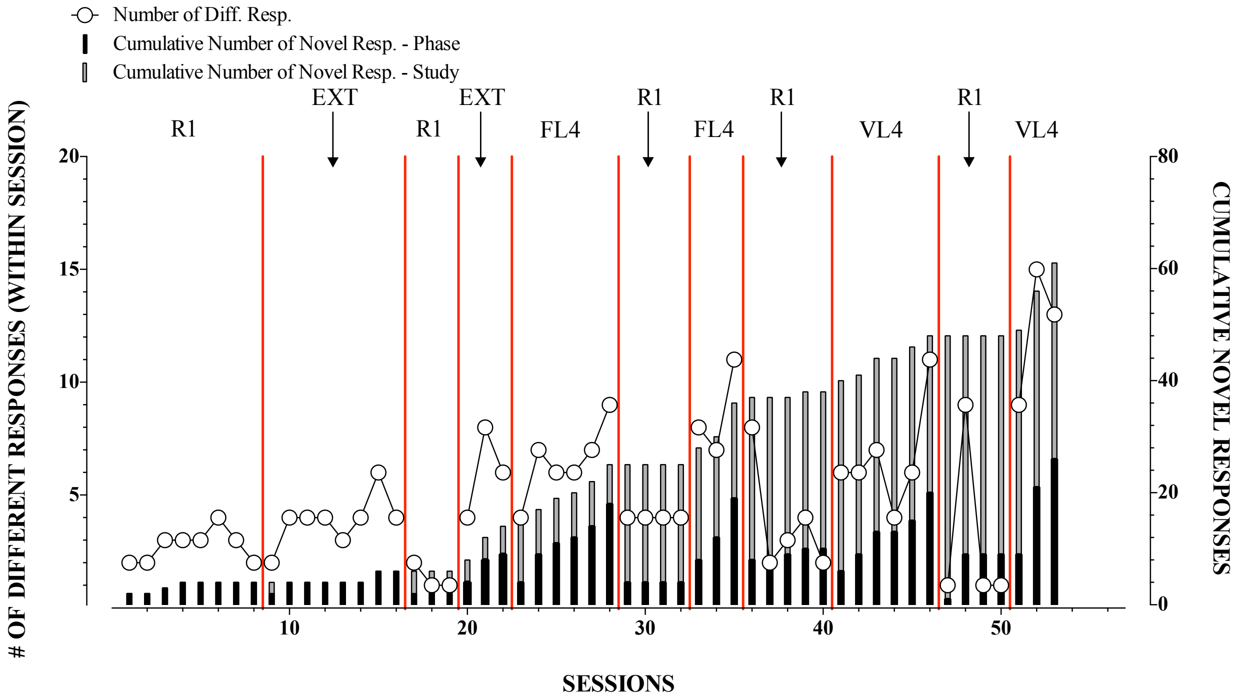


Figure 14. The results of Study 2 for Matteo. Sessions are depicted along the x-axis. The number of different responses (within session) is depicted on the left y-axis and the cumulative number of novel responses (phase and study) is depicted on the right y-axis.

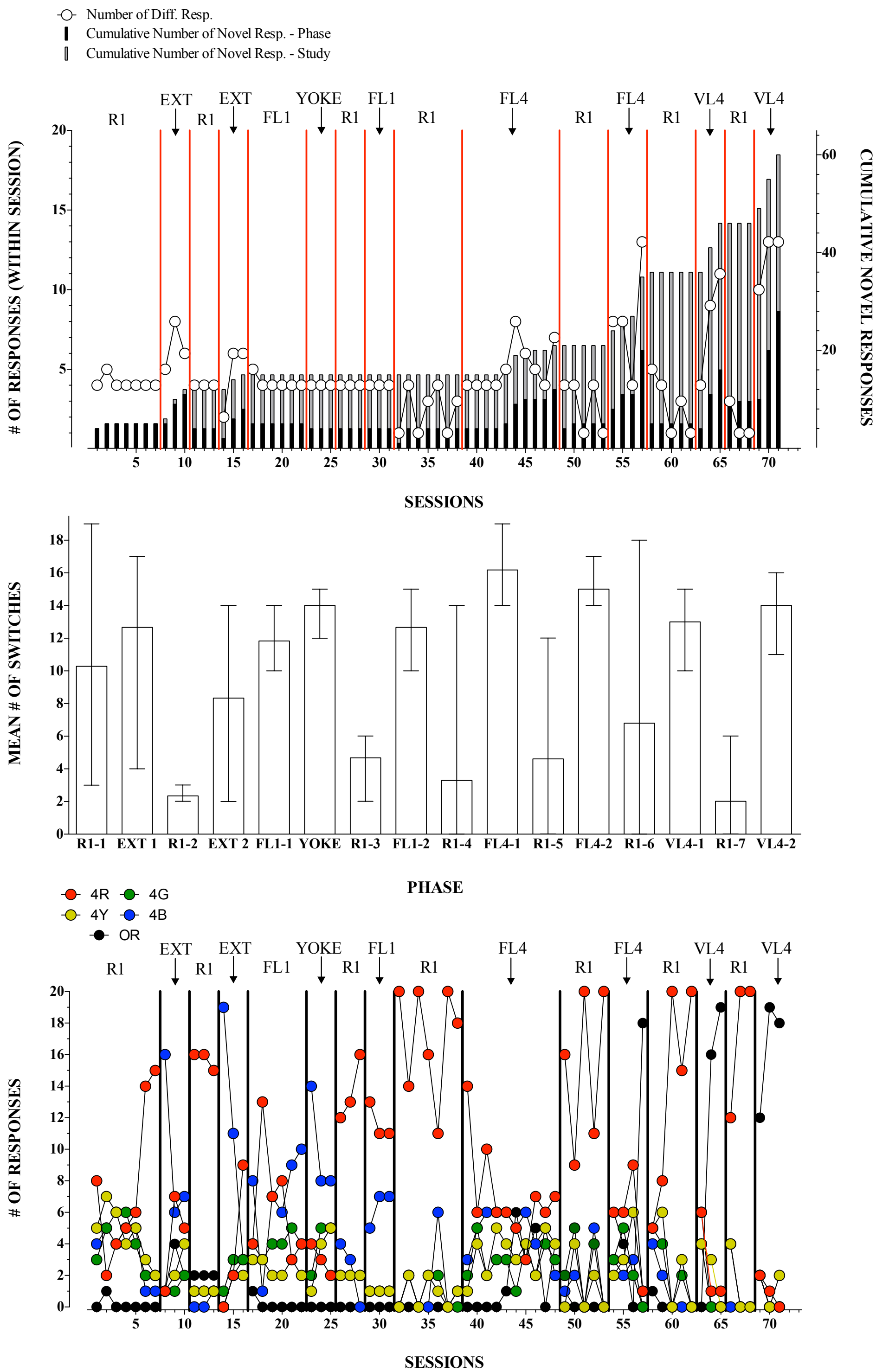


Figure 15. The results of Study 2 for Carlo. In the top panel, sessions are depicted along the x-axis. The number of different responses (within session) is depicted on the left y-axis and the cumulative number of novel responses (phase and study) are depicted on the right y-axis. In the middle panel, each phase is depicted along the x-axis. The mean number switches is depicted along the y-axis. In the bottom panel, sessions are depicted along the x-axis. The number of responses is depicted along the y-axis.

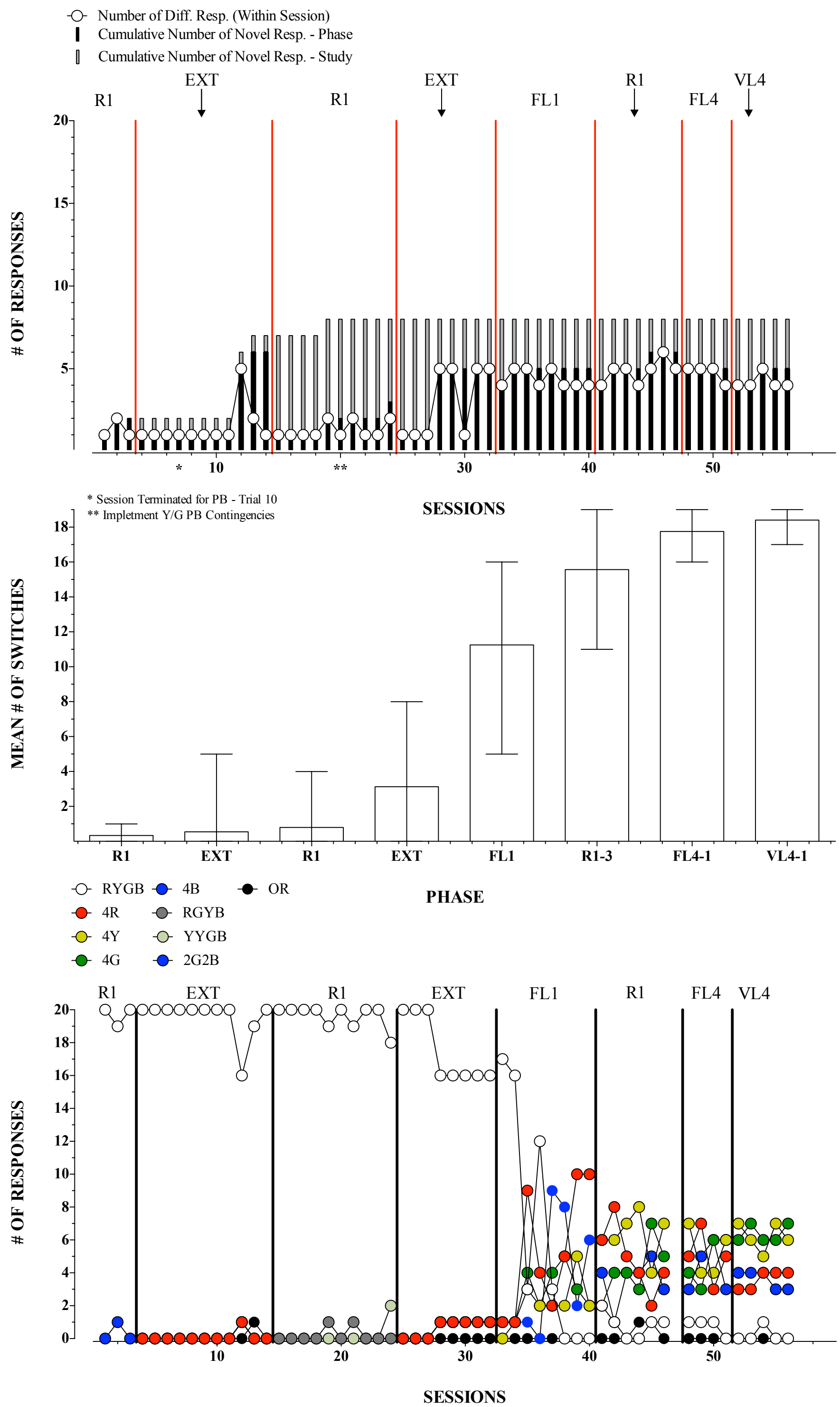


Figure 16. The results of Study 2 for Nico. In the top panel, sessions are depicted along the x-axis. The number of different responses (within session) is depicted on the left y-axis and the cumulative number of novel responses (phase and study) is depicted on the right y-axis. In the middle panel, each phase is depicted along the x-axis. The mean number of switches is depicted along the y-axis. In the bottom panel, sessions are depicted along the x-axis. The number of responses is depicted along the y-axis.

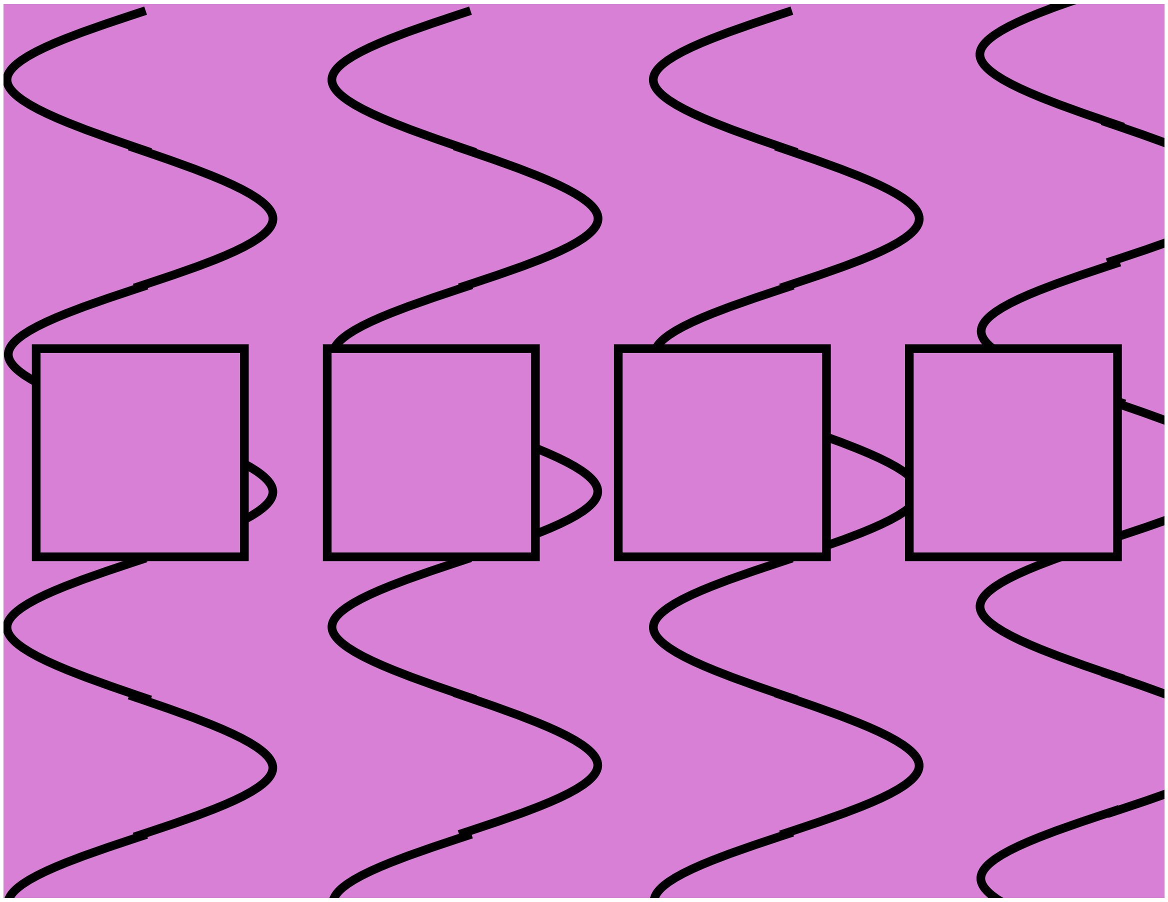


Figure 17. A depiction of the card used in the R 1 training sessions and R 1 components in Study 3.

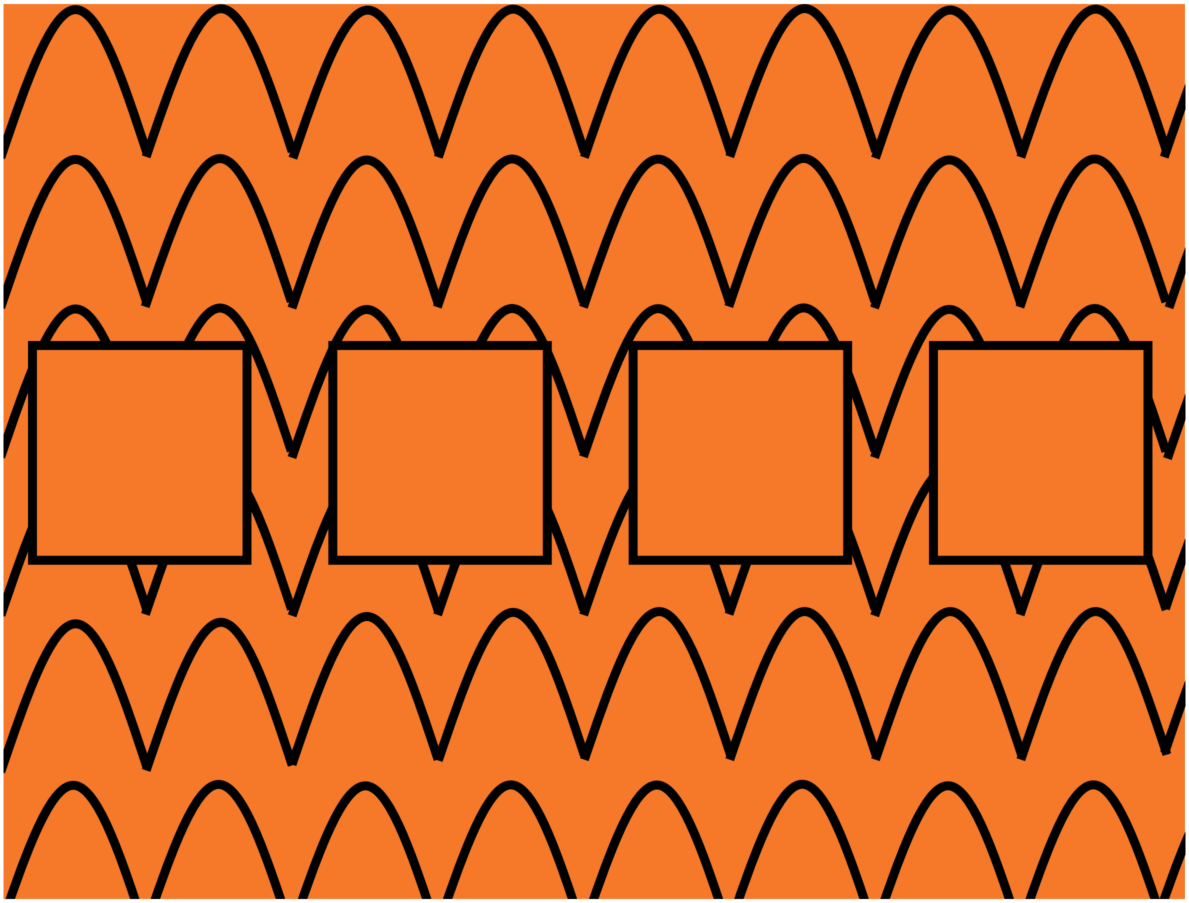


Figure 18. A depiction of the card used in the FL 4 training sessions and FL 4 components in Study 3.

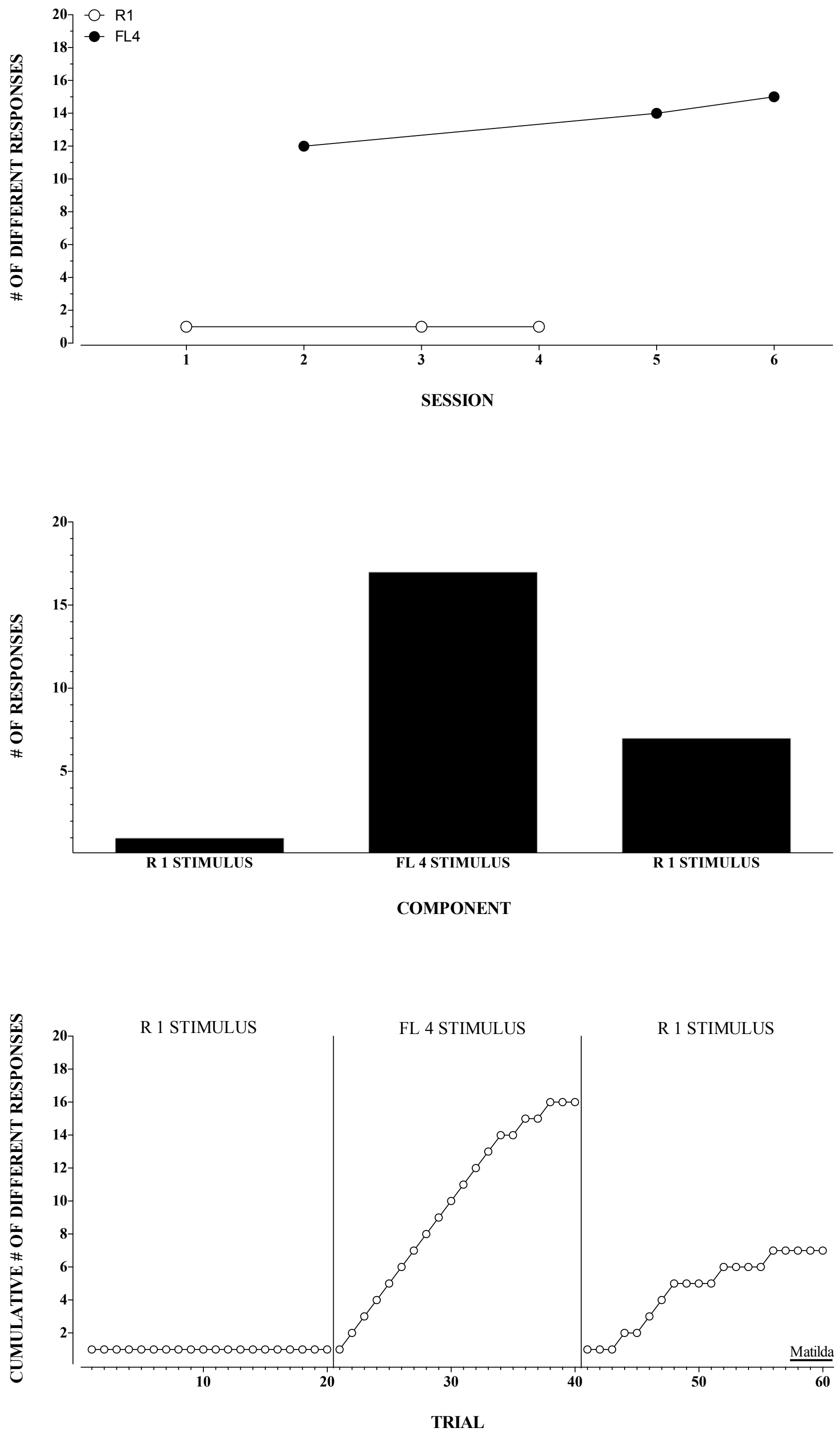


Figure 19. The results of Study 3 for Matilda. The top panel depicts the results of the stimulus-control training. Sessions are depicted on the x-axis. The number of different responses is depicted on the y-axis. The middle panel depicts the total number of responses emitted in each component of the stimulus control test. Components are depicted along the x-axis. The total number of different responses is depicted on the y-axis. The bottom panel depicts the trial-by-trial results of the stimulus-control evaluation. Components are depicted on the x-axis. The cumulative number of different responses and switches are depicted on the y-axis.

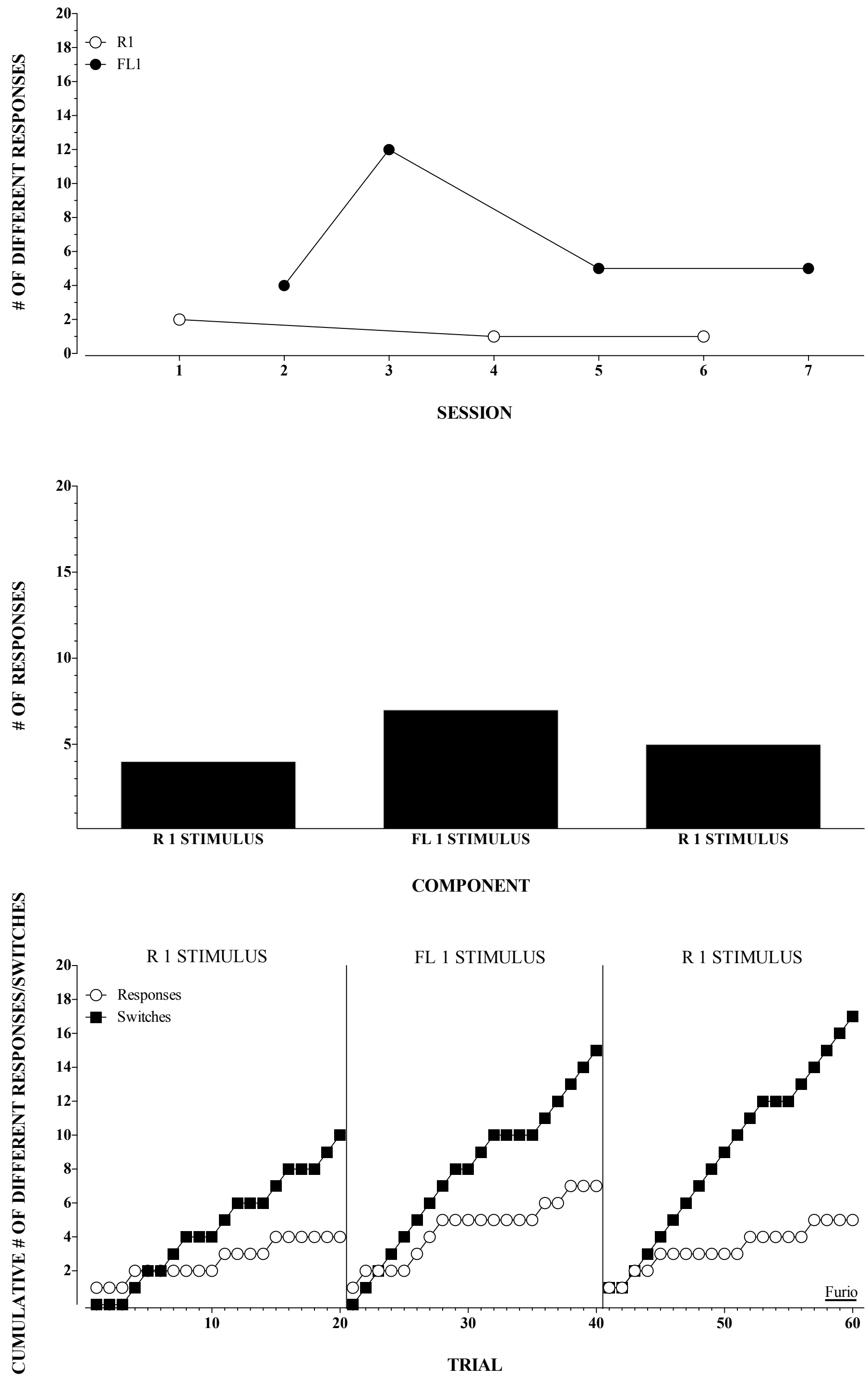


Figure 20. The results of Study 3 for Furio. The top panel depicts the results of the stimulus-control training. Sessions are depicted on the x-axis. The number of different responses is depicted on the y-axis. The middle panel depicts the total number of responses emitted in each component of the stimulus control test. Components are depicted along the x-axis. The total number of different responses is depicted on the y-axis. The bottom panel depicts the trial-by-trial results of the stimulus-control evaluation. Components are depicted on the x-axis. The cumulative number of different responses and switches are depicted on the y-axis.

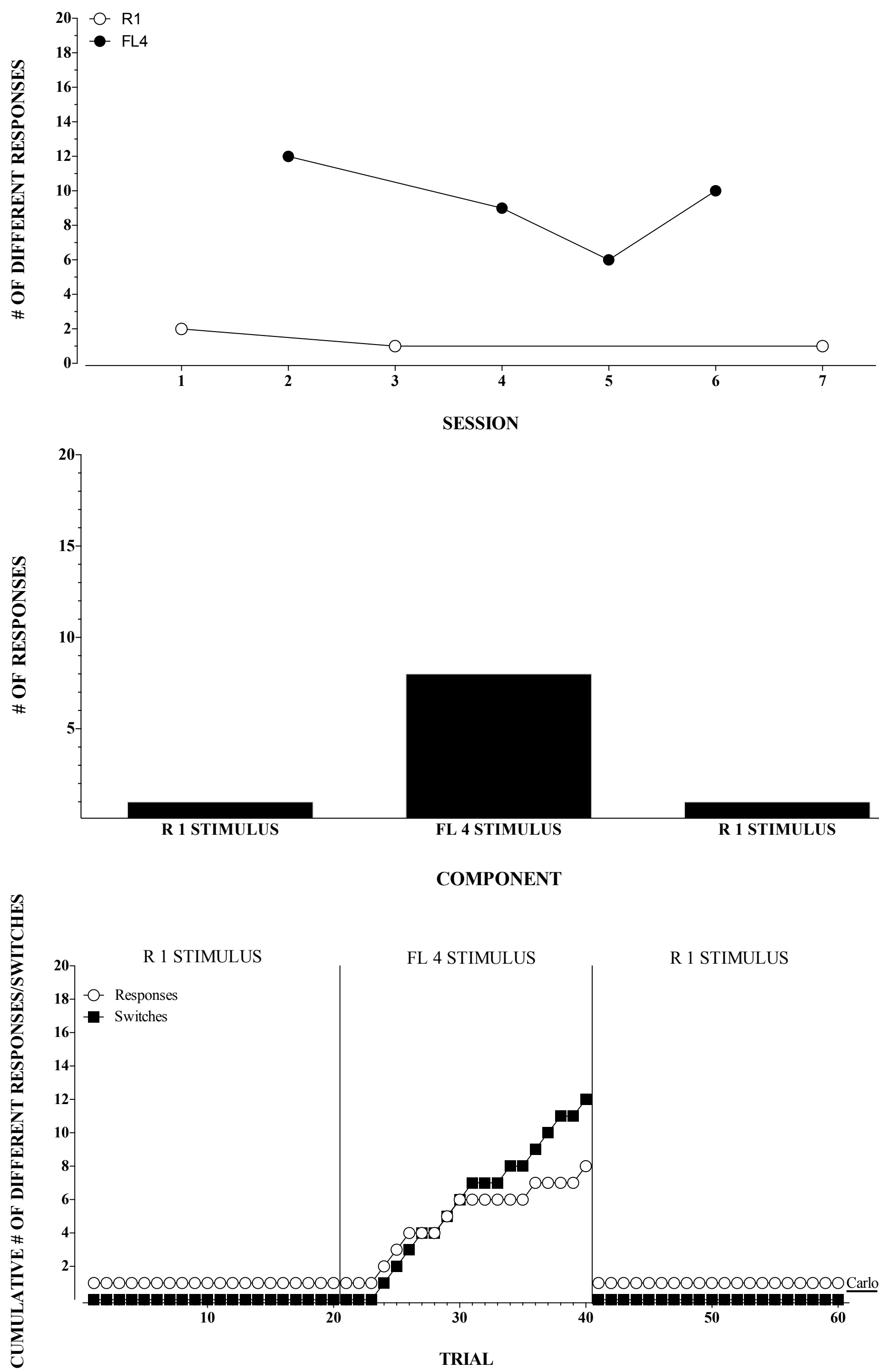


Figure 21. The results of Study 3 for Carlo. The top panel depicts the results of the stimulus-control training. Sessions are depicted on the x-axis. The number of different responses is depicted on the y-axis. The middle panel depicts the total number of responses emitted in each component of the stimulus control test. Components are depicted along the x-axis. The total number of different responses is depicted on the y-axis. The bottom panel depicts the trial-by-trial results of the stimulus-control evaluation. Components are depicted on the x-axis. The cumulative number of different responses and switches are depicted on the y-axis.