An evaluation of lag schedules and prompting methods to increase variability of naming category items in children with autism spectrum disorder

Katie M. Wiskow¹ · Ashley L. Matter² · Jeanne M. Donaldson³

Published online: 12 November 2018

© Association for Behavior Analysis International 2018

Abstract

A lag schedule of reinforcement is one way to increase response variability; however, previous research has been mixed with regard to the necessary parameters to increase variability. For some individuals, low schedule requirements (e.g., Lag 1) are sufficient to increase variability. For other individuals, higher lag schedules (e.g., Lag 3) or a lag schedule in combination with prompting is needed to increase variability. We evaluated the efficiency of different within-session progressive lag schedules to increase response variability with 2 children with autism. Results showed that increasing the lag criterion across sessions increased variability to levels similar to beginning with a high lag schedule for one participant. When lag schedules did not increase variability for the second participant, we compared a variety of prompting procedures. Results of the prompting evaluation showed that a tact-priming procedure was effective to increase varied responding.

Keywords Autism · Lag schedule · Intraverbals · Priming · Prompting · Variability

Neuringer (2002) describes variability as a set of responses that are distributed along a continuum, from repetitive to stochastic. Variability is a dimension of operant behavior, like frequency, duration, and force. This means that we can arrange reinforcement contingencies to directly reinforce or extinguish varied responding. Variability is an

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s40616-018-0102-5>) contains supplementary material, which is available to authorized users.

✉ Katie M. Wiskow
kwiskow@csustan.edu

¹ Department of Psychology and Child Development, College of Science, California State University, Stanislaus, One University Circle, Turlock, CA 95382, USA

² Trumpet Behavioral Health, Dublin, CA, USA

³ Department of Psychology, Louisiana State University, Baton Rouge, LA, USA

important dimension of behavior; however, individuals with autism spectrum disorders (ASD) may exhibit repetitive behaviors, including repetitive speech (American Psychiatric Association, 2013). When a specific response does not result in reinforcement, varied responding may increase the likelihood that an individual can access reinforcement in the novel situation. In addition, varied responding may help develop positive social interactions between the individual with ASD and his or her peers. For example, an individual may be more likely to be accepted by peers if he or she engages in more variable responding during conversation than rote or scripted responding.

Implementing a lag schedule of reinforcement is one way we can directly reinforce varied responses. In a lag schedule, reinforcement is arranged only if the response emitted is different from a specified number of previous responses, along a specific dimension. In applied settings, we often measure variability in the context of whether responses are different in form. For example, in a Lag 2 schedule, the experimenter only provides reinforcement if the response emitted is different from the previous two responses. The mechanisms involved in a lag schedule that likely contribute to its effectiveness to increase varied responding are reinforcement and extinction. Repeat responses are not reinforced, thus we may see extinction-induced variability. Then, when a varied response occurs that is sufficiently different (i.e., meets the lag criterion), the varied response produces reinforcement. In addition, over time, variability can come under discriminative control (e.g., Page & Neuringer, 1985). Researchers have successfully increased response variability using different procedures, including but not limited to the use of lag schedules of reinforcement across a variety of behaviors (e.g., greetings, toy play, vocal utterances) with children with ASD in applied settings (see Rodriguez & Thompson, 2015, for a review). In many of these studies, prompting methods were used in combination with lag schedules. Previous prompting methods have involved prompting a specific response that will be reinforced. For example, script fading (Brodhead, Higbee, Gerencser, & Akers, 2016; Lee & Sturmey, 2014), modeling (e.g., Napolitano, Smith, Zarcone, Goodkin, & McAdam, 2010), and gestural prompting (e.g., Heldt & Schlinger, 2012) methods were used either during the first lag schedule session or following a phase in which the lag schedule alone did not increase variability.

Although direct prompting methods (e.g., modeling) have been effective to teach participants to vary their responses, it is possible that indirect methods of prompting (e.g., response priming) may also be effective to promote varied responses and reduce the amount of within-session prompting. Furthermore, indirect prompting methods may be preferred because they may naturally generate more varied responding by prompting a range of responses, as opposed to prompting imitation of a single response. That is, an experimenter could present multiple discriminative stimuli from which the individual could choose in order to emit a response, instead of the experimenter modeling one response for the individual to imitate. Indirect prompting methods could also be accomplished by including additional antecedent stimuli designed to signal which responses will or will not be reinforced. For example, Brodhead et al. (2016) used script training and lag schedules of reinforcement to establish stimulus control of varied mand frames by a colored mat. During training, the experimenter attached scripts containing different mand frames to the colored mat and implemented a lag schedule in the presence of the green mat and a repetition contingency in the presence of the red mat. Results showed that two of the three participants continued to engage in varied

responding in the presence of the green mat and repetitive responding in the presence of the red mat after the experimenter faded all scripts and did not provide differential consequences.

Overall, invariant responding may occur because (a) the individual has not learned a sufficient number of responses; (b) the individual contacts reinforcement for invariant responses, and therefore, reinforced responses are more likely to occur relative to other responses (Lee, Sturmey, & Fields, 2007); or (c) both. For example, Betz, Higbee, Kelley, Sellers, and Pollard (2011) showed that teaching three mand-frame exemplars did not increase varied responding until the experimenters implemented extinction. Prompting may be a necessary component in cases in which the individual has not learned a sufficient number of exemplars or varied responses are not sensitive to the reinforcement contingencies. Although prompting methods may need to be incorporated into treatment to promote varied responding, evaluating the effects of lag schedules prior to adding additional treatment components may be beneficial for several reasons. First, researchers can determine the effectiveness of using lag schedules to increase varied responding, which may limit the potential intrusiveness of an intervention. Second, implementing lag schedules could serve as a potential baseline assessment to determine the current repertoire of responses the individual is able to emit. Finally, lag schedules can provide independent opportunities for an individual to emit creative (i.e., novel) responses, as well as help experimenters avoid teaching a sequence of responses.

Two recent studies evaluated the effects of Lag 0 and Lag 1 schedules of reinforcement on varied responding to naming category items without direct prompting (Contreras & Betz, 2016; Wiskow & Donaldson, 2016). A Lag 0 in these studies is functionally equivalent to a fixed-ratio (FR) 1 schedule of reinforcement. That is, the experimenter reinforced any correct response, whether or not it was varied. Contreras and Betz (2016) evaluated the effects of Lag 0 and 1 schedules of reinforcement with three children with ASD during individual teaching sessions, whereas Wiskow and Donaldson (2016) evaluated the effects of Lag 0 and Lag 1 schedules of reinforcement in a group context. Results of both studies showed that a Lag 1 schedule was effective to increase varied responding (with two of the three participants in Contreras & Betz, 2016; varied responding only increased with the third participant in Contreras & Betz when the experimenter provided variability training and increased the lag schedule to a Lag 3).

To our knowledge, two studies to date have evaluated the effects of systematically increasing the lag schedule requirement across phases (Falcomata et al., 2018; Susa & Schlinger, 2012). Susa and Schlinger (2012) demonstrated that when the lag criterion increased (from a Lag 1 to a Lag 2 to a terminal Lag 3), varied responding to the discriminative stimulus (SD) “How are you?” increased to meet the criterion. Similarly, Falcomata et al. (2018) implemented increasing lag schedules during functional communication training. When the lag schedule increased (from a Lag 1 to, eventually, a Lag 5), variable mands increased. Although the experimenters in both of these studies systematically increased the schedule requirement, it is unknown whether variability would have increased more quickly or whether the intervention would have produced more variability overall by starting with the larger lag requirement. In fact, Stokes (1999) suggested that arranging larger lag schedules at the onset of variability training may produce higher overall levels of variability compared to increasing the lag schedule over time.

The purpose of the present case studies was to explore variables related to response variability to naming category items with two children with ASD. In a classroom context, a teacher may pose questions (e.g., “Who can name a community helper?”) that require students to vary their responses from each other to receive social approval. Thus, teaching children with ASD to generate varied and novel responses to category items may have social significance. In the first case study, we compared the efficiency of two methods to increase response variability: a progressive lag schedule condition (Lag 1, Lag 2, Lag 4, Lag 8, Lag 9) versus a terminal lag schedule condition (Lag 9). The second case study consisted of the comparison of the progressive lag schedule condition and terminal lag condition followed by a series of evaluations of indirect prompting methods to increase response variability. Across all conditions in both case studies, we provided a description of the schedule contingency prior to each session, presented a visual contingency-specifying stimulus during the session, and delivered feedback within sessions.

General method

Participants, setting, and materials

Two children participated in the study. Randall was a 6-year-old boy diagnosed with autism and pervasive developmental disorder—not otherwise specified. Sophie was a 6-year-old girl diagnosed with autism and a speech disorder. Both participants spoke using full sentences, followed multistep directions, could read short stories, and were able to answer a variety of questions (e.g., “What’s your name?” “What do you want to do?”). In addition, both participants engaged in some stereotypy. Randall often engaged in stereotypy with his hands (i.e., bouncing his hands on the table) and Sophie frequently scripted portions of preferred videos. Parents enrolled their children in the present study and reported that their child was able to name some items in categories; however, parents were interested in improving this skill. We chose the categories of vehicles, bathroom items, and zoo animals for Randall and fruits, vehicles, and instruments for Sophie. Informed consent was obtained from all individual participants included in the study.

We conducted all sessions in a university-based laboratory designed to look like a classroom. During all sessions, the experimenter sat next or adjacent to the participant at a table. Materials included a token board, picture cards (Sophie), a letter strip, and reinforcers (i.e., iPad®, toys, and snacks). The letter strip contained letter combinations to signal the contingency for reinforcement. We used AAAAA for FR 1, ABABA for progressive Lag 1, ABCAB for progressive Lag 2, and ABCDE for progressive Lags 4 to 9. We chose potential reinforcers based on parent and child interviews. In addition, participants could choose from all available toys and games in the classroom and request new games be downloaded on the iPad®.

Dependent variables

We scored all dependent variables the same across phases. We scored a response as *correct* if the participant emitted the response within 10 s of the instruction and the

response matched the category stated by the experimenter, regardless of whether it met the progressive lag criterion. We scored a response as *incorrect* if the participant emitted a response within 10 s of the instruction but the response did not match the category stated by the experimenter. We scored *no response* if the participant did not emit a response within 10 s of the instruction. We scored a response as *different* if it (a) was a correct response and (b) it was the first trial that the participant emitted that response during the current session. Thus, a correct response on the first trial was always scored as *different*. We scored the first correct response as *different* so that the number of different responses during a session was accurate. We did not score a response as *different* if the participant added an adjective to a noun and the noun was functionally the same. For example, we did not score “truck” and “super-duper truck” as *different* because they are functionally the same, but we scored “truck” and “garbage truck” as *different*. We scored a correct response as *novel* if the participant did not emit that response during any previous sessions.

We measured response variability by calculating an *average variability score* (Susa & Schlinger, 2012) for each session. First, we assigned a variability score to each correct response, excluding the first correct response because there were no previous responses with which to compare it. We scored each response the same regardless of the progressive lag schedule in place. For each correct response, we counted the number of different responses that preceded that trial until (a) we had counted all previous trials or (b) we encountered a repeat response. The first score we assigned (for the second trial) was either a 0 (if the second correct response was not different from the first) or a 1 (if it was different). Subsequently, the variability score could be (a) 0 if the response was not different from the last, (b) 1 if the response was different from the last, (c) the number of previous consecutive different responses until the response in the current trial is repeated, or (d) $N + 1$, with N as the highest lag score recorded that session, if a response was scored as *different* during the current trial. We did not include trials containing an incorrect response. Second, we added the variability scores across trials. Finally, we divided the sum of the variability scores by the total number of responses given a variability score (i.e., the number of correct responses minus 1). The maximum variability score (i.e., if all 10 responses within a session were different) was 5. The average variability score measure provides a within-session analysis of how responding conformed to the progressive lag schedule in effect that cannot be determined from the number of different responses. See the Appendix for scoring examples.

Interobserver agreement

Trained graduate students served as the experimenter and data collectors. We compared the response recorded by the primary data collector to the response recorded by the reliability data collector on a trial-by-trial basis. We calculated interobserver agreement (IOA) by scoring each trial as 1 if both observers recorded the same response or 0 if the observers recorded different responses, adding the scores from each trial, dividing the sum of the scores by the total number of trials, and multiplying the result by 100 to yield a percentage. We collected IOA data during 51% and 52% of sessions across all phases with Randall and Sophie, respectively. Average agreement was 99% (range 90%–100%) for both Randall and Sophie.

General procedure

All sessions consisted of 10 trials of a single category. We selected categories of common items that contained more than 10 common exemplars and were absent from the current physical environment (i.e., items that could not be tacted) and with which parents reported their child was familiar; however, we did not conduct any formal preassessments. All sessions included rules, contingency-specifying stimuli, and feedback. At the beginning of each session, the experimenter showed the participant the letter strip and stated the letters with a brief description. For example, during progressive Lag 1 the experimenter said, “Look, this says A-B-A-B-A, so you have to tell me at least two different [category] and you can’t say the same thing in a row.” Then, the experimenter initiated the session by stating, “Tell me a [category].” The experimenter delivered a token and praise immediately following each correct response that met or exceeded the schedule in effect. We implemented a within-session progressive lag schedule during all lag schedule sessions. Each session consisted of 10 trials, regardless of the lag criterion; therefore, the number of opportunities to meet the lag criterion was different for each lag schedule. For example, sessions in which we implemented a progressive Lag 6 schedule of reinforcement included reinforcement for responses meeting a Lag 1 criterion (during Trial 2), a Lag 2 criterion (during Trial 3), a Lag 3 criterion (during Trial 4), and so on. Therefore, there were four total trials (Trials 7 through 10) in a Lag 6 session in which responding could meet a Lag 6 criterion. We implemented the lag schedules in this manner because we included several lag schedules and wanted to keep the total number of trials consistent across phases. If a response was correct but did not meet the lag criterion (Lag 1 through 9 conditions), the experimenter recorded the response; said, “You already told me [item]”; and initiated the next trial by saying, “Tell me a different [category].” If a response was incorrect, the experimenter provided corrective feedback (e.g., “Nice try! [Item] is not a [category].”) and initiated the next trial. If the participant did not respond within 10 s of the SP, the experimenter initiated the next trial with the initial instruction, “Tell me a [category].” We only varied the instruction following correct responses that did not meet or exceed the lag criterion.

Participants traded in their tokens after each session (DeLeon et al., 2014). We set the value of each token at 15 s of access to the chosen backup reinforcer. Participants chose the backup reinforcer following each session. If a participant earned all 10 tokens within a session, he or she was able to access the backup reinforcer for 2 min 30 s. Participants could select any toy or activity available in the room (e.g., various toy sets, iPad®).

FR 1 There were two main purposes of the FR 1 phase. First, the FR 1 schedule of reinforcement served as a control condition. During this condition, we reinforced all correct responses, regardless of whether the response was different. Second, this phase provided evidence that participants were able to emit at least one correct response for each category.

FR 1 + varied instruction Prior to implementing progressive lag schedules, we evaluated whether varying the instruction alone would increase variability. Similar to the instruction test conducted by Phillips and Vollmer (2012) to determine if an

experimenter's vocal prompt could evoke correct responding to pictorial prompts, we sought to determine if stimulus control could be transferred from repeat responding to varied responding by simply providing a prompt to do so. If varying the instruction produced an increase in varied responding, then no other interventions would be needed. Contingent on a correct response that was not different from the previous response, the experimenter implemented an error-correction procedure by saying a variation of "Can you think of a different [category]?" The experimenter provided a token and praise contingent on any correct response (i.e., FR 1) following the experimenter's feedback. Therefore, each trial included up to two responses. We did not vary the instruction within a trial if the first response met the criteria for a different response. We only implemented this error-correction procedure during this phase. We calculated the variability score in this phase similarly to the other phases except that in trials in which we implemented the error-correction procedure, we only included each second response the participant emitted. Therefore, we maintained 10 responses in the session analyses during this phase.

Lag schedule comparison We compared the effects of a progressive lag schedule condition to a terminal lag schedule condition on variability in responding across sessions. In the progressive lag schedule condition, we started with a progressive Lag 1 schedule of reinforcement and increased the criterion geometrically by two (i.e., progressive Lag 1, progressive Lag 2, progressive Lag 4, progressive Lag 8) except for the final increase from progressive Lag 8 to progressive Lag 9. In the terminal lag schedule condition, we started with a progressive Lag 9 schedule of reinforcement immediately after the FR 1 + varied instruction condition. We set the terminal lag schedule such that we only reinforced different responses within a session. In both conditions, we implemented a within-session progressive lag schedule. That is, the lag schedule increased within session. For example, during progressive Lag 4 sessions, we reinforced the first four different responses with a token and praise and then reinforced each subsequent response that met or exceeded the Lag 4 criterion. During progressive Lag 9 sessions, the participant could earn 10 tokens only when he or she emitted 10 different responses in 10 trials.

Case study 1: Randall

Procedure

We used a multiple-baseline across-categories design to replicate the effect of the terminal lag schedule condition, a changing-criterion design to evaluate the effects of the progressive lag schedule condition, and a multielement design to directly compare the progressive lag schedule condition and the terminal lag schedule condition on response variability. We never conducted more than three consecutive sessions of a single category except for the final four sessions with Randall. The criterion to increase the lag schedule was two sessions in which the average variability score was within 0.5 of the average variability score indicating conformation to the lag schedule in effect (see Table 1). One exception to this was the change between progressive Lag 8 and

progressive Lag 9 with Randall. The average variability score for Randall was 4.22 (4.89 max) in the final progressive Lag 8 session prior to increasing the criterion to progressive Lag 9. We increased the criterion to progressive Lag 9 because it only required one additional response to meet the criterion and Randall had already emitted eight different responses during progressive Lag 8. In addition, the contingencies appeared identical until the final two trials (progressive Lag 8) or last (progressive Lag 9) trial of the session. We conducted all conditions as described previously.

Results and discussion

The top panel of each figure shows the comparison of the progressive lag schedule condition to the terminal lag schedule condition using a multielement design. In addition, we replicated the effects of the terminal lag schedule condition using a multiple-baseline design. Figure 1 shows the number of different responses Randall emitted during each session across the categories of vehicles, items found in the bathroom, and zoo animals. In FR 1, Randall emitted two different vehicles, one item found in the bathroom, and four or five zoo animals. When we implemented the FR 1 + varied instruction condition, we observed no difference in the vehicles category, one additional response in the in-the-bathroom category, and similar variability in the zoo animals category. Thus, varying the instruction alone did not produce an increase in the number of different responses. When we implemented the progressive lag schedule condition in the vehicles category, the number of different responses increased with each increase in the schedule requirement. The number of different responses Randall emitted immediately increased when we implemented the terminal lag schedule condition with the in-the-bathroom category. Randall met the terminal criterion of nine different responses in 21 sessions in the vehicles category and in 6 sessions in the in-the-bathroom category. The number of different responses decreased across sessions during FR 1 in the zoo animals category. When we implemented the terminal lag schedule condition, we observed an immediate increase in the number of different responses and reached the terminal criterion of nine in two sessions. Interestingly, errors were more likely to occur across both conditions when the criterion for reinforcement initially increased. Errors primarily consisted of no response or of items that did not fit the category. For example, Randall said, “robots” in response to naming a vehicle, “mustache” in response to naming a bathroom item, and “alien” in response to naming a zoo animal. Also, on a few occasions Randall emitted a response that shared features of a previously emitted response that was reinforced. For example, following reinforcement for the response “spaceship,” he said “planets,” and following reinforcement for the response “fire truck,” he said “tires.”

Table 1 Average variability scores indicating conformation with each lag schedule

Schedule	Score
Lag 1	1.00
Lag 2	1.89
Lag 4	3.33
Lag 8	4.89
Lag 9	5.00

Figure 2 shows the average of the average variability scores for each schedule across categories for Randall. In FR 1 we observed low variability with vehicles ($M = 0.78$), zero variability with items found in the bathroom, and moderate variability with zoo animals ($M = 2.25$). When we implemented the FR 1 + varied instruction condition, we observed no change in variability with vehicles ($M = 0.72$) and a small increase with items found in the bathroom ($M = 0.69$). Subsequently, the average variability score increased with each schedule change for vehicles (Lag 1, $M = 0.74$; Lag 2, $M = 1.29$; Lag 4, $M = 2.06$; Lag 8, $M = 3.39$; Lag 9, $M = 3.55$). If we included responses to which Randall attached adjectives (e.g., super car, super-duper car, toy car, sonic car), then Randall engaged in maximum variability in three of five sessions during progressive Lag 9. When we implemented progressive Lag 9 for items found in the bathroom, variability immediately increased and remained high across sessions ($M = 3.40$). We observed an immediate increase in the average variability score when we implemented progressive Lag 9 with zoo animals ($M = 3.88$). The maximum variability score Randall emitted was 4.22 (vehicles) in the progressive lag schedule condition and 4.50 (bathroom items) and 5.00 (zoo animals) in the terminal lag schedule condition.

Figure 3 shows the cumulative number of novel responses across categories for Randall. These data show a similar response pattern as was observed in Figs. 1 and 2. That is, in the FR 1 phase, Randall emitted the same one or two responses across sessions. The cumulative number of novel responses increased with each increase in the lag criterion but increased more quickly in the terminal lag schedule condition. At the end of this study (excluding those responses with adjectives), Randall emitted 20

Fig. 1 Number of different responses emitted during each session across categories for Randall

Fig. 2 Average of the average variability scores during each schedule across categories for Randall. Adj. represents adjectives and NT represents not tested

unique vehicles, 17 unique items found in the bathroom, and 23 unique zoo animals (Table 2). Randall emitted several creative responses that varied by the adjective but not the noun and thus were not counted as novel. For example, Randall said, “super-duper car,” “fast car,” “super truck,” and “two shampoos.”

Randall’s data are consistent with previous research demonstrating that response variability increased with an increase in the lag schedule requirement (Falcomata et al.,

Fig. 3 Cumulative number of novel responses emitted during each session across categories for Randall

Table 2 Novel responses emitted across categories for randall

Schedule	Vehicles	In the bathroom	Zoo animals
FR 1	Car Airplane	Toilet paper Soap Toilet	Zebra Lion Penguin Giraffe Panda bear Polar bear Gorilla Tiger
Lag 2	Truck Bus	NT	NT
Lag 4	Fire truck Garbage truck Bicycle Construction truck	NT	NT
Lag 8	Spaceship Boat Ship	NT	NT
Lag 9	Skateboard Doctor truck Hot-air balloon Mail truck Cheese (dairy) truck Zoo animal truck Pirate ship Cage (horse and buggy) Racing car	Sink Paper towel Shower Bath Counter Drawers Shampoo Body wash Towel Curtain Trash Makeup Floating toys Lid (drain)	Hippo Monkey Cheetah Wolves Bears Alligator Baby long neck (turtle) Leopard Lizard Elephant Seal Fish Dolphin Whale Ducks

NT not tested; responses in parentheses are the therapist's interpretation of an ambiguous response

2018; Susa & Schlinger, 2012). In addition, the results of Randall's case study showed that response variability increased more rapidly when we implemented a higher lag schedule (i.e., progressive Lag 9) first compared to progressively increasing the lag schedule across sessions. However, these results should be interpreted with caution because we did not replicate these effects. Furthermore, both conditions eventually produced a similar level of variability and novel responses.

An additional finding from the present study was that despite producing more rapid increases in variability, starting with a higher lag schedule may have aversive properties. Randall engaged in social collateral behaviors, similar to behaviors observed during extinction (e.g., emotional responses), at the beginning of the terminal lag schedule condition suggesting that it was aversive. For example, when we displayed the letter strip indicating a progressive Lag 9 criterion, Randall emitted verbal statements such as "Ugh! No! It is too hard!" and often manded for the other category associated with the progressive lag schedule condition. Therefore, the letter strip may

have acquired aversive properties because we paired it with the terminal lag schedule condition. Subsequently, it may have functioned as an S^Δ for repeated responses. In addition, during the terminal lag schedule condition sessions, Randall would often say things such as “Think, think, think!” while tapping his head and “I can’t think!” while groaning. Although these social collateral behaviors decreased across sessions and eventually ceased, they suggest that the response effort involved in emitting multiple vocal responses to a single vocal S^D was aversive. Randall did not engage in similar social collateral behaviors during the progressive lag schedule condition when the criterion reached progressive Lag 9. Thus, although the terminal lag schedule condition was more efficient than the progressive lag schedule condition for Randall, practitioners should be cautious with its use and monitor for potential negative side effects. In addition, the total number of novel responses emitted in each condition was similar; therefore, the overall benefit of the terminal lag schedule condition may be limited. Future research could also investigate using a progressive lag schedule to avoid negative social collateral behaviors but probe the terminal lag schedule when increasing the lag schedule requirement to increase the efficiency of the procedure.

Case study 2: Sophie

Procedure

Both progressive Lag 2 and progressive Lag 9 marginally increased varied responding for Sophie. We hypothesized that the low levels of variability may have been related to the effort involved in generating multiple vocal responses to a single vocal S^D without the presence of additional stimuli that may evoke correct responses. Therefore, we conducted a series of indirect prompting modifications to reduce the effort involved with emitting these responses and subsequently faded out the prompts.

We used a multiple-baseline across-categories design and reversal design to replicate the effect of the tact-priming procedure. We determined the order of sessions by selecting for each one of the three categories in random order. With Sophie, we repeated intervention sessions more frequently than FR 1 sessions (instruments). We conducted sessions with categories for which the participant’s progress was occurring more slowly to provide multiple opportunities to contact the session contingencies; however, we never conducted more than three consecutive sessions of a single category.

Dynamic visual S^Δ prompts When Sophie emitted a correct response, the experimenter placed a picture corresponding to the item she named on the table. Pictures on the table were programmed to function as S^Δ s. That is, if Sophie tacted a picture on the table, then the experimenter withheld reinforcement. We provided a token and praise contingent on any correct response emitted of a category item not on the table. The experimenter removed picture cards when the response corresponding with that card became eligible for reinforcement. For example, in the progressive Lag 2 + dynamic visual S^Δ prompt sessions, if Sophie said, “apple,” “banana,” and “orange” in the first three trials, the fourth trial would be initiated with a picture of a banana and an orange on the table because the response “apple” (and any fruit except banana and orange) was eligible for reinforcement.

Tact priming The dynamic visual S^Δ prompts were ineffective to increase varied responding; therefore, we implemented a tact-priming procedure. Prior to each session, the experimenter prompted Sophie to tact nine different pictures of category items (one at a time). Picture cards were selected based on (a) responses Sophie previously emitted and (b) experimenter discretion of common items in the category. The nine picture cards selected remained the same across sessions. The experimenter held up the first picture card and asked, “What is it?” The experimenter only repeated this prompt if Sophie did not tact the picture card within 5 s of its presentation. Following each correct tact, the experimenter echoed the response. The experimenter did not deliver any other feedback or reinforcement during the tact-priming procedure. After Sophie tacted all nine pictures, the experimenter removed the pictures from Sophie’s field of vision and began the session. The experimenter shuffled the picture cards before each session so that the experimenter did not present the pictures in the same order.

Tact priming + visual S^D s With the instruments category, Sophie had emitted a sufficient number of category items across sessions to meet the lag criterion (progressive Lag 9) but had not emitted some responses for numerous sessions. Therefore, we added visual S^D prompts, in the form of picture cards, for the instruments category. First, we conducted tact priming as described previously. Second, we placed six picture cards of instruments that Sophie emitted in prior sessions, but had not emitted within the previous three sessions before implementing this phase, on the table. After Sophie tacted one of the instruments on a picture card, the picture was removed from her sight. Therefore, we programmed the presence of the picture cards on the table to function as S^D s. That is, if Sophie tacted a picture card on the table, the experimenter delivered a token and praise and removed the picture card. After four sessions in this condition, we faded out the visual S^D prompts by including three pictures on the table for two sessions while maintaining the tact-priming prompts.

Results and discussion

Figure 4 shows the number of different responses emitted during each session across the fruits, vehicles, and instruments categories for Sophie. In FR 1 Sophie emitted one or two different responses during each session across categories. We implemented one session with a varied instruction and observed no change in the number of different responses with any category. Then, we implemented the progressive lag schedule condition with the fruits category and the terminal lag schedule condition with the vehicles category while instruments remained in FR 1. Sophie began emitting three different responses in both targeted categories; however, this number stabilized in the progressive Lag 9 phase. Therefore, it appeared that the lag schedules alone were insufficient to increase varied responding to a desirable level.

We maintained the current lag schedules (i.e., progressive Lag 2 for fruits and progressive Lag 9 for vehicles) and implemented the dynamic visual S^Δ prompts. After three sessions with no change we removed dynamic visual S^Δ prompts and implemented tact priming. With the implementation of tact priming in the fruits and vehicles categories, we observed an immediate increase in the number of different responses in both categories, regardless of the lag schedule. Therefore, we increased the

Fig. 4 Number of different responses emitted during each session across categories for Sophie. V represents varied instruction

criterion to progressive Lag 9 with the fruits category and conducted a reversal with and without tact priming. Little difference was observed between the two conditions, and Sophie continued to emit 10 different fruits in the absence of tact priming in the final phase. When we removed tact priming with the vehicles category, we observed an increase in variability in the number of different responses emitted across sessions, and the number of different responses remained high overall.

Because we observed the greatest increase in the number of different responses when tact priming was implemented, we implemented tact priming alone during FR 1 with the instruments category. When we implemented tact priming with the instruments category, we observed an immediate increase in the number of different responses; however, Sophie never exceeded eight different responses within a session although she had emitted 10 different responses across sessions. We implemented progressive Lag 9 with tact priming, but the number of different responses remained unchanged. Therefore, we added visual S^D prompts with tact priming to prompt responses that were

weakened across sessions—that is, responses that had not been emitted for several sessions and therefore had not been reinforced (i.e., strengthened). Tact priming plus visual S^D prompts increased the number of different responses emitted within session. The number of different responses maintained at high levels when we removed visual S^D prompts, as well as when we removed tact priming in the final phase.

Figure 5 shows the average of the average variability scores for each schedule across categories for Sophie. We observed the average variability score increase the most when we introduced tact priming with the fruits and vehicles categories. Similarly, tact priming alone was sufficient to increase the average variability score with the instruments category. Moreover, the addition of visual S^D prompts further increased the average variability score with the instruments. Similar to Randall, Sophie emitted some responses that varied based on the adjective but not the noun or said a portion of the noun. For example, Sophie said, “purple grapes” and “grapes,” “green pear” and “pear,” “squashy lemons” and “lemons,” “plane” and “airplane,” “truck box” and “truck,” “xylophone sticks” and “xylophone,” and “circle drums” and “drums.” In each of these examples, we did not count the second response of the pair as a different response. The average variability scores remained high across all categories when we withdrew tact priming in the final phase with each category.

Figure 6 shows the cumulative number of novel responses across categories for Sophie. Prior to implementing tact priming, Sophie had emitted four unique fruits, four unique vehicles, and three unique instruments across sessions. When tact-priming prompts were introduced, Sophie emitted six new fruits, six new vehicles, and ten new instruments. In the final phase, when we removed all prompts, Sophie emitted two new fruits, seven new vehicles, and one new instrument. See Table 3 for a summary of novel responses Sophie emitted across categories and conditions.

The results of Sophie’s case study showed that, similar to previous research (e.g., Contreras & Betz, 2016), lag schedules alone were insufficient to increase variability in responding. There are several potential explanations for the ineffectiveness of lag schedules to increase variability in responding for Sophie. First, the reinforcers provided for varied responding may not have been sufficient to compete with invariant responding. Second, it is possible that the research design also influenced the effectiveness of the lag schedule. For example, we alternated FR 1 sessions with the other conditions (e.g., progressive Lag 9). Thus, although Sophie may have earned a small magnitude of reinforcement (e.g., 30 s) in one session (e.g., progressive Lag 9), she had the opportunity to earn more of the reinforcer in subsequent sessions (e.g., FR 1). Finally, the history of reinforcement for invariant responding may have strengthened the responses that persisted in the absence of reinforcement (i.e., extinction) within sessions after the response was first emitted and reinforced. We reset the lag schedule requirement each session; therefore, stereotyped behaviors were still reinforced on an intermittent schedule both within (e.g., progressive Lag 1 schedule) and across (e.g., progressive Lag 9 schedule) sessions. For example, prior to tact priming, Sophie consistently emitted the responses “orange,” “apple,” and “banana” in the fruits category, and we reinforced each of those responses the first time Sophie emitted the response in a session. A similar finding on the influence of history of reinforcement was observed in Brodhead et al. (2016) with one participant. Experimenters reported that after the participant emitted a varied mand frame in a no-vary session and the experimenter implemented the error-correction procedure, the participant emitted the

Fig. 5 Average of the average variability scores during each schedule across categories for Sophie. NT represents not tested and TP represents tact priming

Fig. 6 Cumulative number of novel responses emitted during each session across categories for Sophie. V represents varied instruction

correct repeated mand frame for the remainder of the session. Thus, experimenters reported that the participant's invariant responding in the presence of the red mat may have been controlled by the experimenter's error-correction procedure rather than the colored mat paired with the vary (green matt) or no-vary (red mat) contingency. Future research may increase the number of trials within a session or extend a lag schedule across sessions. After a few sessions with no change in responding, the experimenter could (a) increase the number of trials within a session or (b) provide some type of stimulus (e.g., rule, pictures) to signal those responses will not result in reinforcement

Table 3 Novel responses emitted across categories for sophie

Schedule	Fruits	Vehicles	Instruments
FR 1	Orange Apple	Van Motorcycle	Trumpet Xylophone Violin
FR 1 + V	0	0	0
FR 1 + TP	NT	NT	Guitar Flute Harp Drums Piano Keyboard Maracas Tuba
Lag 1	Banana	NT	NT
Lag 2	Tomato	NT	NT
Lag 2 + D	0	NT	NT
Lag 2 + TP	Strawberry Grapes Blueberry Pineapple Pear	NT	NT
Lag 9	Lemon Lime	Car Truck Ship Garbage truck Cement truck Bulldozer Dump truck Tractor Bike	Harmonica
Lag 9 + D	NT	0	NT
Lag 9 + TP	Plum	Helicopter Airplane Bus Fire truck Boat Jeep	Horn
Lag 9 + TP + S ^D	NT	NT	Triangle

NT not tested, *0* zero novel responses, *D* dynamic visual S^A prompts, *TP* tact priming, *V* varied instruction, S^D visual S^D prompts

during the session. Utilizing the latter approach would eliminate a frequent response from meeting the reinforcement criterion on an intermittent schedule.

One potential concern with the use of lag schedules is the development of higher order stereotypy (Schwartz, 1982). Higher order stereotypy occurs when an organism repeats a specific pattern of responses. For example, in a Lag 1, higher order stereotypy would consist of alternating between two responses (apple, orange, apple, orange, etc.). Although we did not specifically measure higher order stereotypy, Sophie often alternated between a few responses (e.g., orange, orange, banana, orange, orange,

banana) during sessions with low lag requirements (i.e., Lag 1 and Lag 2). We also observed that some responses in the fruits category were more likely to occur during the first half of the trials (strawberry, orange, banana, tomato, apple) and some were more likely to occur during the second half of the trials (pineapple, pear, blueberries, grapes, lemon) within a session during progressive Lag 9. However, these responses did not always occur in the same order. Therefore, it is unlikely that Sophie learned a specific chain of responses, but it is possible that the responses she emitted in the beginning of the session had a stronger history of reinforcement and thus were strengthened in her repertoire.

Interestingly, the tact-priming procedure was more effective than the lag schedules across all categories for Sophie. This type of response priming may be effective because it decreases the time between emitting the exemplars as tacts and emitting the exemplars as intraverbal responses that may be reinforced. Furthermore, this is potentially promising for clinicians, as tact priming may be easier to implement as an antecedent intervention than implementing lag schedules within session.

General discussion

There are several findings from the present study that contribute to the literature on response variability. First, this is the first study to directly compare initial lag schedule values on levels of response variability. Although these results should be considered preliminary, given that we only demonstrated this comparison with Randall, future research should continue to evaluate parametric features of our procedures that produce efficient intervention results. Second, this is the first study to evaluate a tact-priming procedure within the context of lag schedules. Finally, both participants began to independently emit more creative responses toward the end of the study when the lag requirements were increased in both conditions (e.g., beginning with progressive Lag 4 for Randall) by adding adjectives to previously emitted responses. One of the purposes of teaching individuals to vary their responses is so that when one response is placed on extinction, the individual is able to access reinforcement by emitting a different, functionally equivalent response. Therefore, lag schedules of reinforcement may be one method to evoke and reinforce these creative responses. In the present study, we chose not to count these responses as different; however, future research should explore the effects of lag schedules on creative novel responses.

There are several limitations of this study that researchers should address in future studies. First, we did not thoroughly assess each participant's current repertoire prior to implementing the lag schedules. We chose categories that did not have visual aids readily available in the session environment (e.g., clothing, colors) and categories with which the parents reported participants were familiar. Contreras and Betz (2016) conducted a preexperimental assessment to determine the participants' existing repertoires prior to implementing lag schedules. During the preexperimental assessment, experimenters asked participants to name category items and provided verbal prompts such as "What else?" or "Tell me more" on each trial. However, our varied prompt condition was similar to this and was ineffective to evoke different responses. During bathroom-items category sessions, Randall would sometimes request to go to the bathroom (presumably to look at what items were in the bathroom). Therefore, it is

difficult to determine whether the participants' responding (Sophie, in particular) was influenced by a skill deficit rather than insufficient reinforcers.

Future research may also explore preference for the categories themselves. For example, Sophie engaged in positive social behaviors (e.g., shouting, "Hooray!") when we stated we were going to name fruits and negative social behaviors (e.g., groaning, "No, I don't want to!") when we stated we were going to name instruments. Therefore, it is possible that emitting a response in the fruits category required less response effort or was associated with other reinforcers that then negated the need for additional prompting procedures. Future researchers may use preference data to determine an initial response requirement for a lag schedule that may increase varied responding and avoid negative social collateral behaviors.

Another potential limitation is that we implemented several different prompting methods but only demonstrated experimental control with tact priming. Furthermore, the behavioral mechanism responsible for the effectiveness of tact priming is unknown. Tact priming may have been effective because it reduced the response effort involved in emitting multiple vocal responses to a single vocal S^D . Tact priming reduced the delay between observing stimuli and emitting a response. Although Sophie was able to independently tact all picture cards we presented during tact priming, it is possible that Sophie learned to emit these responses as intraverbals. Many previous studies utilizing lag schedules of reinforcement have included prompting procedures (e.g., Heldt & Schlinger, 2012; Susa & Schlinger, 2012). Thus, future research is needed to directly compare these prompting methods to determine the most efficient way to promote varied responding for individuals with ASD.

Readers should attend to two important features of the lag schedule conditions in the present study. First, we combined the lag schedules with several other components (i.e., rules, contingency-specifying stimuli, corrective feedback). Therefore, we cannot determine the individual influence of these components on response variability. Second, the way that we defined lag schedules may be more accurately described as within-session progressive lag schedules than a single value (e.g., Lag 6), which may make these data difficult to compare to previous research. In contrast, previous experimenters included primer trials prior to the session trials (e.g., Contreras & Betz, 2016). The primer trials consist of the same S^D used during session trials (e.g., "Tell me a [category]."), but the experimenter reinforced all correct responses (FR 1). The experimenter uses the responses the participant emits during the primer trials to determine whether the first response during session meets the lag schedule. Therefore, the number of primer trials varies depending on the lag schedule in effect, and it is possible for each session trial to meet the lag schedule.

Finally, researchers should explore the procedures implemented in the present study in classrooms with other students present, similar to Wiskow and Donaldson (2016). For example, teachers can arrange opportunities for students to take turns answering a question and only reinforce an individual's response if it is different than a response emitted by his or her peer (or a number of peers). Wiskow and Donaldson (2016) found that a Lag 1 schedule implemented in a group context increased varied responses but the number of different responses each participant emitted did not change or decreased. Thus, it would be interesting to compare lag schedules applied to the group (i.e., the contingency requires responses to vary from peers) to lag schedules applied to

individual students within a group. The latter may result in repeat responses across students but could also promote observational learning.

These preliminary data suggest that there may clinical benefits to implementing a large lag schedule at the onset of teaching (e.g., rapid increase in variability); however, there are also potential disadvantages (e.g., creating an establishing operation for escape behaviors). Furthermore, the same amount of variability likely can be achieved by systematically increasing the criterion. When additional prompting methods are needed, the tact-priming procedure evaluated in the present study may be an easy, low-effort option for clinicians to program. Clinicians and researchers should continue to explore the best ways to arrange lag schedules of reinforcement and prompting methods to promote varied and novel responding across a range of response topographies (e.g., academic, social, play) for individuals with ASD.

Acknowledgements We thank Batool Alsayedhassan for her assistance in collecting data.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflicts of interest.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Appendix

Tables 4, 5 and 6 are examples of sessions that consist of 3 different responses (Tables 7 and 8 consist of 6 and 10 responses, respectively); however, the average variability score is different based on the within-session response patterns.

Table 4 Three different responses with a lag sum of 3 and an average variability score of 0.33

Trial	Response	Variability Score
1	Orange	None
2	Apple	1
3	Banana	2
4	Banana	0
5	Banana	0
6	Banana	0
7	Banana	0
8	Banana	0
9	Banana	0
10	Banana	0
	<i>Sum</i>	3
	<i>Average</i>	0.33

Table 5 Three different responses with a lag sum of 9 and an average variability score of 1

Trial	Response	Variability SCORE
1	Orange	None
2	Apple	1
3	Orange	1
4	Apple	1
5	Apple	0
6	Orange	1
7	Banana	2
8	Orange	1
9	Banana	1
10	Banana	1
	<i>Sum</i>	<i>9</i>
	<i>Average</i>	<i>1</i>

Table 6 Three different responses with a lag sum of 17 and an average variability score of 1.89

Trial	Response	Variability score
1	Orange	None
2	Apple	1
3	Banana	2
4	Orange	2
5	Apple	2
6	Banana	2
7	Orange	2
8	Apple	2
9	Banana	2
10	Orange	2
	<i>Sum</i>	<i>17</i>
	<i>Average</i>	<i>1.89</i>

Table 7 Six different responses, one repeat, and one incorrect response

Trial	Response	Variability score
1	Orange	None
2	Apple	1
3	Banana	2
4	Lemon	3
5	Truck	None
6	Lemon	0
7	Lemon	0
8	Orange	3
9	Lime	4
10	Pear	5
	<i>Sum</i>	<i>18</i>
	<i>Average</i>	<i>2.25</i>

Table 8 Ten Different responses (i.e., maximum variability)

Trial	Response	Variability score
1	Orange	None
2	Apple	1
3	Banana	2
4	Lemon	3
5	Pineapple	4
6	Blueberry	5
7	Kiwi	6
8	Strawberry	7
9	Lime	8
10	Pear	9
	<i>Sum</i>	45
	<i>Average</i>	5

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: Author.
- Betz, A. M., Higbee, T. S., Kelley, K. N., Sellers, T. P., & Pollard, J. S. (2011). Increasing response variability of mand frames with script training and extinction. *Journal of Applied Behavior Analysis*, 44, 357–362. <https://doi.org/10.1901/jaba.2011.44-357>
- Brodhead, M. T., Higbee, T. S., Gerencser, K. R., & Akers, J. S. (2016). The use of a discrimination training procedure to teach mand variability to children with autism. *Journal of Applied Behavior Analysis*, 49, 34–48. <https://doi.org/10.1002/jaba.280>
- Contreras, B. P., & Betz, A. M. (2016). Using lag schedules to strengthen the intraverbal repertoires of children with autism. *Journal of Applied Behavior Analysis*, 49, 3–16. <https://doi.org/10.1002/jaba.271>
- DeLeon, I. G., Chase, J. A., Frank-Crawford, M. A., Carreau-Webster, A. B., Triggs, M. M., Bullock, C. E., & Jennett, H. K. (2014). Distributed and accumulated reinforcement arrangements: Evaluations of efficacy and preferences. *Journal of Applied Behavior Analysis*, 47, 293–313. <https://doi.org/10.1002/jaba.116>
- Falcomata, T. S., Muething, C. S., Silbaugh, B. C., Adami, S., Hoffman, K., Shpall, C., & Ringdahl, J. E. (2018). Lag schedules and functional communication training: Persistence of mands and relapse of problem behavior. *Behavior Modification*, 42, 1–21. <https://doi.org/10.1177/0145445517741475>
- Heldt, J., & Schlinger, H. D. (2012). Increased variability in tacting under a lag 3 schedule of reinforcement. *The Analysis of Verbal Behavior*, 28, 131–136. <https://doi.org/10.1007/BF03393114>
- Lee, R., & Sturmey, P. (2014). The effects of script-fading and a lag-1 schedule on varied social responding in children with autism. *Research in Autism Spectrum Disorders*, 8, 440–448. <https://doi.org/10.1016/j.rasd.2014.01.003>
- Lee, R., Sturmey, P., & Fields, L. (2007). Schedule-induced and operant mechanisms that influence response variability: A review and implications for future investigations. *The Psychological Record*, 57, 429–455. <https://doi.org/10.1007/BF03395586>
- 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. <https://doi.org/10.1901/jaba.2010.43-265>
- Neuringer, A. (2002). Operant variability: Evidence, functions, and theory. *Psychonomic Bulletin & Review*, 9, 672–705. <https://doi.org/10.3758/BF03196324>

- Page, S., & Neuringer, A. (1985). Variability is an operant. *Journal of Experimental Psychology: Animal Behavior Processes*, 11, 429–452. <https://doi.org/10.1037/0097-7403.11.3.429>
- Phillips, C. L., & Vollmer, T. R. (2012). Generalized instruction following with pictorial prompts. *Journal of Applied Behavior Analysis*, 45, 37–54. <https://doi.org/10.1901/jaba.2012.45-37>
- Rodriguez, N. M., & Thompson, R. H. (2015). Behavioral variability and autism spectrum disorder. *Journal of Applied Behavior Analysis*, 48, 167–187. <https://doi.org/10.1002/jaba.164>
- Schwartz, B. (1982). Failure to produce response variability with reinforcement. *Journal of the Experimental Analysis of Behavior*, 37, 171–181. <https://doi.org/10.1901/jeab.1982.37-171>
- Stokes, P. D. (1999). Learned variability levels: Implications for creativity. *Creativity Research Journal*, 12, 37–45. https://doi.org/10.1207/s15326934crj1201_5
- 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. <https://doi.org/10.1007/BF03393113>
- Wiskow, K. M., & Donaldson, J. M. (2016). Evaluation of a lag schedule of reinforcement in a group contingency to promote varied naming of categories items with children. *Journal of Applied Behavior Analysis*, 49, 1–13. <https://doi.org/10.1002/jaba.307>