Behavior Analysis in Practice
https://doi.org/10.1007/s40617-026-01182-0

BRIEF PRACTICE

Assessment and Treatment of Multiply Maintained Complex Motor Stereotypy

Matthew L. Edelstein[1,2] · **Thomas K. Franzmann**[1,2] · **Joshua A. Mellott**[1]

Received: 24 December 2025 / Accepted: 1 April 2026

Abstract

Previous studies have demonstrated the utility of multi-component interventions for the treatment of interfering motor stereotypy. However, less is known about contraindications to established treatment procedures when stereotypy is multiply maintained. The current study evaluated the effects of an empirically derived, caregiver-implemented response reduction procedure on motor stereotypy in a child with autism using a withdrawal design. Results suggested reductions in motor stereotypy from baseline ($M = 93\%$), with caregivers reporting high acceptability regarding the efficacy and feasibility of the approach.

- Stereotypic behavior deemed as functionally impairing may be considered as a target for behavior analytic treatment. Caregivers can be important partners for making this determination.
- Motor stereotypy is often assumed to be maintained by automatic reinforcement; however, the use of functional analysis is essential to confirm function and determine the most appropriate course of treatment.
- Incorporating caregiver input about treatment procedures prior to their being programmed is as useful for collaboration around family values as it is for ensuring the generality and durability of outcomes.

Keywords Complex motor stereotypy · Automatic reinforcement · Functional behavior assessment · Social validity

Motor stereotypy involves repetitive, rhythmic body movements and is one response form that falls within the restrictive and repetitive behavior core feature of autism spectrum disorder. Complex motor stereotypy, typically involving bilateral, coordinated, and sustained movements involving multiple muscle groups, can profoundly interfere with individuals' daily functioning, including reduced engagement with academic or community-based activities (Akers et al., 2020). In addition, engagement in high rates of stereotypy can pose safety risks when the individual is engaging in the response in the community to the exclusion of safety awareness (e.g., wandering away from caregivers while engaging in the behavior, running into people, objects, or traffic). While not all topographies of motor stereotypy are indicated for assessment and treatment (e.g., brief fidgeting, foot tapping; Cook & Rapp, 2018; Leaf et al., 2022), close collaboration with stakeholders (e.g., teachers, parents, same-aged peers) can provide meaningful information about whether the intrusiveness and impact of the behavior warrants intervention (Cunningham & Schreibman, 2008).

While motor stereotypy is often assumed to be maintained by sensory consequences (i.e., automatic positive reinforcement; Cook & Rapp, 2018), more than half of interventions are reported in the extant literature without confirmation of function via functional analysis. This pattern may stem from a long-standing, but incorrect assumption that stereotypy lacks a socially mediated function, reducing the perceived need to conduct a functional analysis. For instance, in their systematic review, Akers and colleagues (2020) identified that among the 41% of participants who received any type of functional behavioral assessment prior to intervention (e.g., indirect assessment, descriptive assessment), only about half had received an experimental functional analysis. The results of Akers et al. are troubling, as research on behavior-analytic treatment efficacy strongly supports the use of functional analysis to guide treatment planning (e.g., Heyvaert et al.,

✉ Matthew L. Edelstein
Edelstein@kennedykrieger.org

[1] Department of Behavioral Psychology, Kennedy Krieger Institute, 1750 E. Fairmount Avenue, Baltimore, MD 21231, USA

[2] Johns Hopkins University School of Medicine, Baltimore, MD, USA

Published online: 15 April 2026

2014). The finding that motor stereotypy may be influenced by social reinforcers has been reported in only a few publications (e.g., Kennedy et al., 2000; Roantree & Kennedy, 2006). Given our understanding of stereotypy as an operant behavior, it seems reasonable that some individuals may engage in stereotypy to access or maintain socially mediated reinforcers, rather than to exclusively access sensory consequences.

In cases where stereotypy interferes with safety and persists despite the programming of reinforcement-based procedures, consideration of punishment-based procedures may be clinically and ethically appropriate (BACB, 2020; Cook & Rapp, 2018; Leaf et al., 2022). Systematic methods have been identified to empirically evaluate the effects of punishment procedures (e.g., Response Interruption and Redirection [RIRD], Ahearn et al., 2007). Verriden and Roscoe (2019) evaluated the use of a punisher assessment wherein potentially punitive consequences were delivered contingent on the occurrence of target behavior. Specifically, after reinforcement-based procedures failed to sufficiently decrease stereotypy, a multielement design was used to evaluate several punishers. Following the analysis, the most effective punisher (determined by the lowest engagement level) was then implemented for treatment. The evaluation suggested that punishers, like reinforcers, may be idiosyncratic to individuals and require individualized assessment. However, the procedures of Verriden and Roscoe focused only on problem behavior maintained by automatic reinforcement.

There is still a significant gap between research and clinical practice regarding the use of empirical methods to assess and treat complex motor stereotypy. Thus, the purpose of the current investigation was to utilize procedures described by Verriden and Roscoe (2019) to evaluate the systematic use of consequent-based procedures in the treatment of socially mediated motor stereotypy.

Method

Participant and Setting

Toby was a 7-year-old male diagnosed with autism spectrum disorder and attention-deficit hyperactivity disorder, combined type, admitted to an intensive outpatient hospital unit for the assessment and treatment of complex motor stereotypy. His motor stereotypy occurred across contexts at reportedly high rates, preventing him from engaging in community and academic activities. In addition, he became so engrossed in his full-body stereotypic behavior that he would wander away from his caregivers in the community, creating imminent safety risks in areas such as parking lots and crowded places. Toby's functional communication skills included 2–3-word vocal utterances, and he could follow multiple-step directions. Appointments were conducted in a 5 m × 5 m treatment room with a one-way observation window and occurred for 3 h per day for 5 consecutive days. Additional materials in the room included two chairs, a desk, and relevant session materials (e.g., iPad, toys, and discriminative stimuli). Generalization sessions occurred first in the lobby of the clinic (Fig. 3, sessions 11, 28–30) and then in an adjacent hotel lobby (Fig. 3, sessions 31–33, 36–38), as these were places where caregivers reported that Toby would be likely to engage in the highest rates of motor stereotypy. All sessions (except those in which the clinician modeled noncontingent reinforcement procedures) were conducted by caregivers trained by the first author in all procedures. This study was deemed exempt by the Johns Hopkins Medicine Institutional Review Board (IRB), and caregivers provided informed consent before data collection began.

Procedures

Functional Analysis

The clinician trained caregivers to implement a functional analysis (FA; Iwata et al., 1994) to identify environmental variables thought to evoke and maintain Toby's motor stereotypy. Data were collected on the duration of stereotypy across all FA conditions and series; sessions were 5-min and relevant putative reinforcement was delivered contingent on stereotypy. The following conditions were evaluated in a randomized order using a multielement design and continued until responding showed a stable, differentiated pattern: divided attention, tangible, demand, toy play, and alone. Tests for positive reinforcement contingencies involved the removal of preferred access (i.e., attention, tangibles) at the start of the condition; the divided attention involved caregivers conversing with the clinician without responding to any of Toby's appropriate bids for attention. Tests for negative reinforcement contingencies involved the delivery of work by a caregiver (i.e., demand). Evocative stimuli in divided attention, tangible, and demand conditions were removed contingent on motor stereotypy for approximately 30 s. In the toy play condition, caregivers provided unrestricted access to preferred activities and attention throughout the 5-min session. Finally, the alone condition involved having caregivers step out of the treatment space for the duration of the 5-min condition.

Response Reduction Assessment

On the basis of caregiver input on socially acceptable procedures, the following procedures were included on the basis of the hypothesis that they may be associated with reductions in motor stereotypy: *Contingent demands* included contingent presentation of one-step gross motor demands

consistent with RIRD procedures (Ahearn et al., 2007). *Time out* included contingent removal of caregivers from the treatment room for 30-s following instances of motor stereotypy. *Hands down* included caregivers placing their hands on Toby's arms and physically prompting them down for 30-s contingent on motor stereotypy (e.g., Verriden & Roscoe, 2019). The term "response reduction assessment" was used in place of "punisher assessment" as we thought this might improve acceptability when it was later taught by caregivers to other staff (e.g., teachers, therapists). Similar to the attention condition of the FA, caregivers provided Toby with 2-min of high-quality attention prior to the start of each assessment session. Once sessions began, caregivers informed Toby that they would be unavailable and described the contingency in place for his stereotypy (e.g., "if you don't have quiet hands, we will help you"). Contingent on target behavior, caregivers were trained to implement the relevant consequence. Conditions were randomized in each series (i.e., a series included the presentation of each of the three test conditions) to avoid potential carry-over effects; all sessions were 5 min in duration and were evaluated using a multielement design.

Baseline

Percentages of motor stereotypy from the divided attention condition of the FA were used as the initial baseline for the treatment evaluation. Subsequent baseline phases involved replicating procedures in a similar manner. The final baseline phase occurred in the natural environment but was otherwise consistent procedurally.

Treatment Evaluation

An ABACADAD withdrawal design was used to evaluate a progressive sequence of phases to reduce the duration of Toby's motor stereotypy. All sessions were 5 min, and approximately 6–9 sessions were conducted per day over three consecutive days of treatment. Data were collected on the duration of motor stereotypy across sessions and converted to percentages. The first treatment phase (B) involved the use of *noncontingent reinforcement (NCR)*; during this condition, Toby was given approximately 10-s of high-quality attention from caregivers on a fixed-time 60-s schedule, regardless of whether stereotypy was occurring. The NCR schedule was calculated by dividing the total duration of pauses between bouts of stereotypy (i.e., 200 s) by the total number of pauses between bouts (i.e., 3 pauses), in the condition with the longest duration of stereotypy, yielding an inter-response time of 66.7 s. The reinforcement delivery interval was set at 60 s to pre-empt motivation to engage in stereotypy. Motor stereotypy did not receive programmed consequences during the NCR condition. The second treatment phase (C) included both NCR and *differential reinforcement of incompatible behavior (DRI).* During this phase, in addition to NCR, caregivers were trained to implement the procedure contingent on Toby's appropriate mands for preferred items or attention (i.e., "snack please"). That is, caregivers would present a visual stimulus (a rules card) to depict behavioral expectations that involved responses incompatible with motor stereotypy (i.e., calm body, quiet hands). Contingent on demonstration of the incompatible response for 3-s, caregivers provided access to requested items or activities. Contingent on mands that coincided with stereotypy, caregivers would gesture to the visual stimulus. Following evidence that NCR + DRI were not sufficient to reduce stereotypy to socially significant levels, the third treatment phase (D) incorporated the results of the response reduction assessment. During this phase, all reinforcement procedures from the previous phase remained in place; however, contingent on motor stereotypy, caregivers used a hands-down procedure involving 30-s of physically prompting Toby's hands down by his sides. At the beginning of each session, caregivers signaled the condition by placing a rubber bracelet on Toby's arm and provided instructions (e.g., "remember to have quiet hands when you wear the bracelet"). After the session ended, caregivers removed the bracelet, signaling that motor stereotypy would not contact the hands-down procedure. Percent reductions were calculated using the last three sessions of each treatment phase.

Social Validity

In each phase of the treatment evaluation, caregivers were asked to rate their impressions of the feasibility and efficacy of the procedures on a scale of 0 to 10, where 0 = minimally acceptable, 5 = neither acceptable nor unacceptable, and 10 = maximally acceptable (see Supporting Information). These ratings were incorporated into the clinical decision-making process along with visual inspection of behavior data to determine the components of the final treatment package.

Response Measurement, Interobserver Agreement, and Procedural Fidelity

Data were collected by doctoral-level behavior analysts with over 10 years of experience collecting continuous behavior data via pencil and paper, as well as a digital timer to track duration data. *Motor stereotypy* was defined as any period of gross-motor stereotypic movement involving two or more parts of Toby's body (e.g., arm waving and body rocking). Examples included arm waving and/or leg thrusting, where extremities were fully extended in a flailing-type motion. Non-examples included movements involving smaller muscle groups (i.e., finger flicking, wrist twitching) as they did

not meet the standard for interfering behaviors warranting intervention. Duration onset occurred following the first instance where the operational criteria were met and offset occurred following 3 consecutive seconds where these responses had stopped occurring. The percentage of session time with motor stereotypy was calculated by dividing the total duration of stereotypy by the total session duration and multiplying the quotient by 100. To ensure the data accurately reflected changes in Toby's behavior, the percentage of motor stereotypy was corrected by subtracting the time his hands were blocked from the total session duration.

All functional analysis and treatment evaluation sessions conducted in the treatment space were recorded using a HIPAA-compliant cloud-based system to allow for independent secondary coding of interobserver agreement (IOA). In the functional analysis, 50% of all sessions were coded for IOA, and 33% of sessions were coded for IOA in the treatment evaluation. Baseline and treatment sessions that occurred in the natural environment could not be coded for IOA due to a lack of recording capabilities or a second clinician. Agreement for the duration of motor stereotypy and time spent in the hands-down procedure was calculated using a partial-interval agreement method. Each session was divided into 10-s intervals. Data collectors continuously timed the duration of motor stereotypy and hands down; the duration within each interval was then divided by the duration of the interval and converted to a percentage. Agreement on the nonoccurrence of behavior was scored as 100% for that interval. The percentage of intervals was then averaged across the session to produce a session-level agreement value. Agreement for motor stereotypy during the functional analysis was 92% (range 86–100%), and during the treatment evaluation was 90% (range 80–100%). Agreement for the hands-down procedure was 100%.

Caregivers' procedural fidelity was evaluated both in vivo and upon review of video footage between appointments. Skills were taught by the first author using behavioral skills training (e.g., Sarokoff & Sturmey, 2004), with modeling occurring via role play during the functional analysis and during the first phase of the treatment evaluation (sessions 5–7). Procedural fidelity was calculated during video review by dividing correct responses (e.g., delivery of programmed contingencies within 3 s of motor stereotypy) by total opportunities and multiplying by 100. Fidelity was coded for 33% of both functional analysis and treatment evaluation sessions, including representative samples of each condition in the functional analysis and each phase of the treatment evaluation. The only exception was generalization sessions, where video review was not available. Procedural fidelity averaged 94% (range 92–98%) during the functional analysis and 96% (range 94–100%) during the treatment evaluation.

Results

Functional Analysis

Results from Toby's FA are shown in Fig. 1. Toby engaged in the highest level of stereotypy in the attention condition ($M = 24.3\%$), and moderate levels of stereotypy in

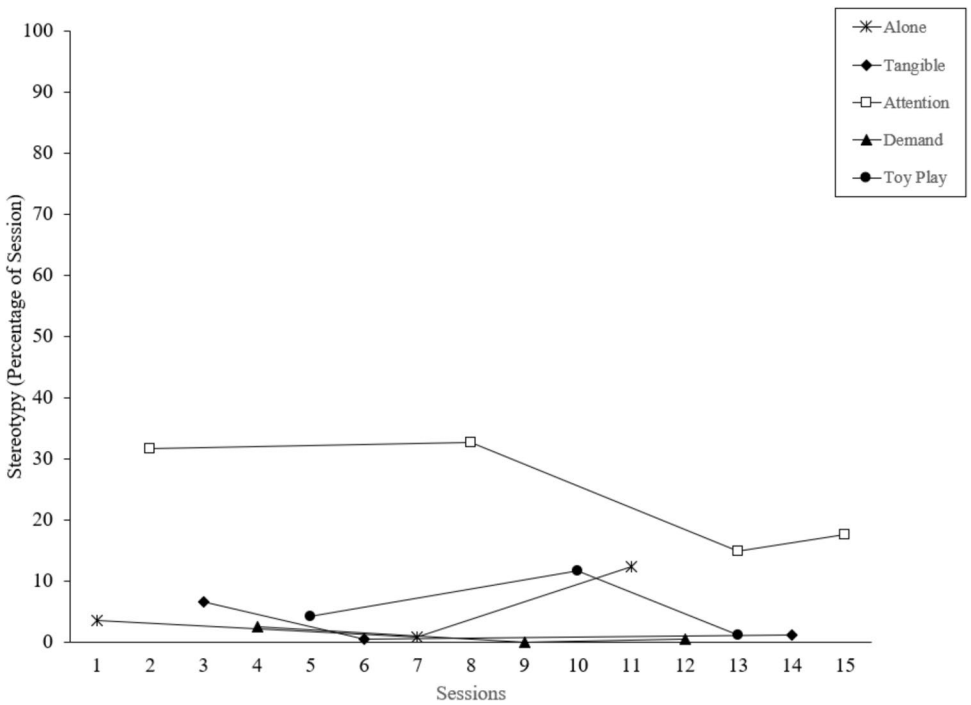

Fig. 1 Results of functional analysis

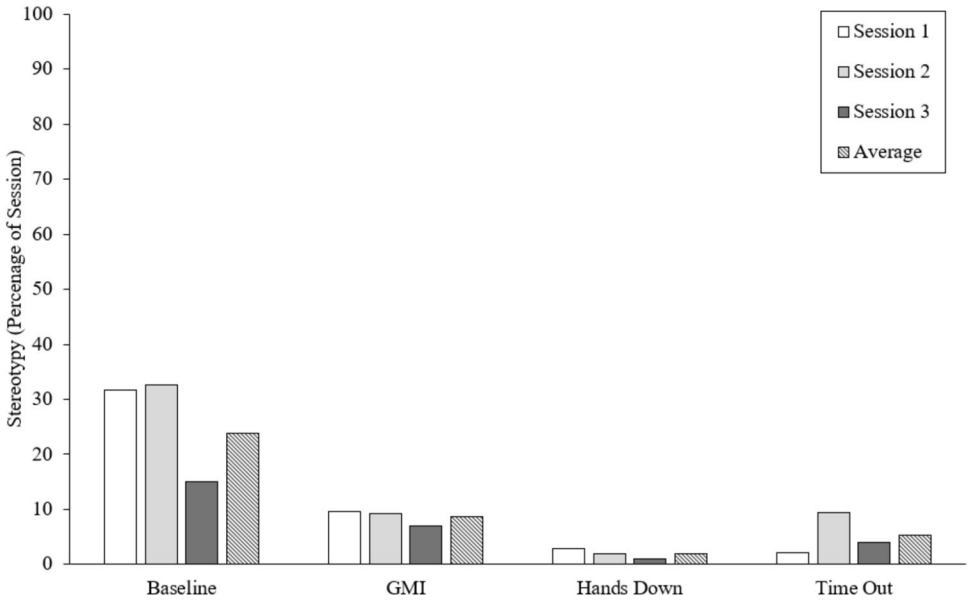

Fig. 2 Results of response reduction assessment. *Note*. *GMI* Gross motor imitation

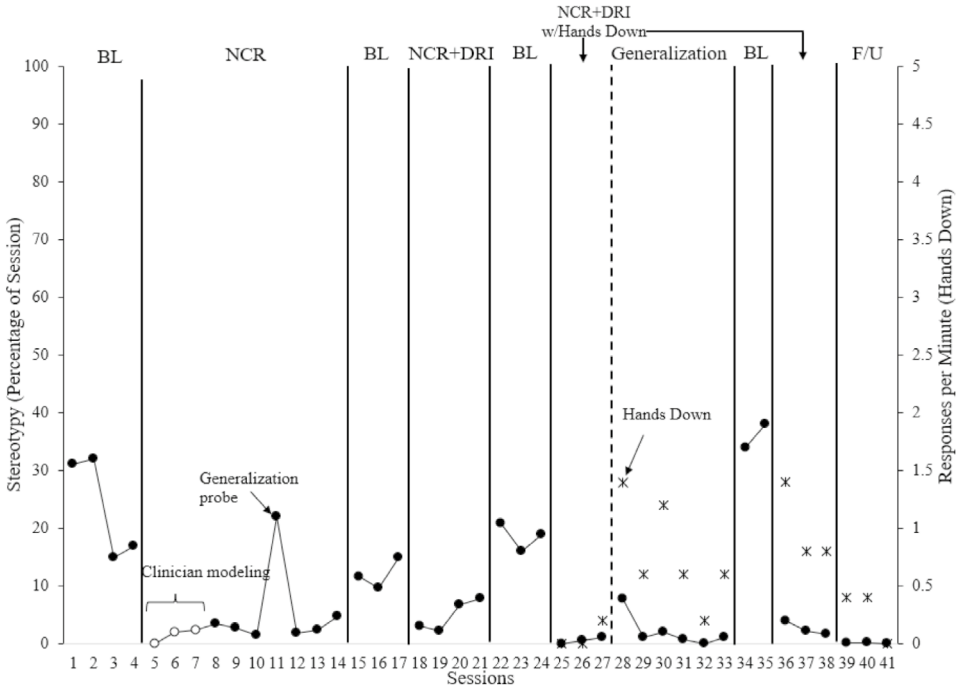

Fig. 3 Results of treatment evaluation. *Note. BL (A phase)* baseline; *NCR (B phase)* noncontingent reinforcement; *NCR + DRI (C phase)* noncontingent reinforcement and differential reinforcement of incompatible behavior; *NCR + DRI w/ Hands Down (D phase)* reinforcement and response reduction procedures; *F/U* follow-up sessions conducted via telehealth

the alone ($M = 5.7\%$) and toy play ($M = 5.7\%$) conditions. Lowest levels of stereotypy were observed in the escape ($M = 1.1\%$) and tangible ($M = 2.8\%$) conditions. Session 9 (escape) was the only condition where Toby was observed to engage in zero instances of motor stereotypy. These results suggest that Toby's motor stereotypy was primarily maintained by social positive reinforcement (i.e., attention). However, as motor stereotypy duration occurred at moderate rates in both the alone and toy play conditions, a level of differentiation (LOD; Hagopian et al., 2023) metric was used to quantify the degree to which Toby's motor stereotypy in the alone condition was differentiated from the toy play condition. Specifically, the level-of-differentiation formula reflects the proportional rate of responding across conditions; it is calculated via the following formula: LOD = [1 − (mean rate in the play condition ÷ mean rate in the no-interaction condition)] × 100. Toby's LOD score is 0%, suggesting that there was likely a secondary automatic function (e.g., Hagopian et al., 2023).

Response Reduction Assessment

Results of the response reduction assessment are shown in Fig. 2. All three procedures were observed to reduce the level of stereotypy from baseline. Gross motor demands were associated with the highest levels of stereotypy ($M = 8.6\%$) while reported by caregivers as being moderately socially valid (caregiver efficacy score = 5); alternatively, the lowest levels were observed in the hands-down condition ($M = 1.93\%$), with caregivers reporting high social validity (caregiver efficacy score = 8).

Noncontingent Reinforcement (NCR)

Results from Toby's treatment evaluation are depicted in Fig. 3. The first phase of the treatment evaluation (NCR; sessions 5–14, excepting session 11) produced reductions in motor stereotypy from baseline ($M = 91\%$). However, during a generalization probe (session 11), target behavior was observed to return to baseline levels, raising concerns about generality and prompting low social validity scores from caregivers (e.g., caregiver efficacy score = 2).

Noncontingent Reinforcement and Differential Reinforcement of Incompatible Behavior (NCR + DRI)

Following the reversal to baseline, the NCR + DRI phase resulted in an initial reduction in motor stereotypy ($M = 82.7\%$), followed by an increasing trend of stereotypy in sessions 20 and 21. Caregivers scored the anticipated efficacy as 4, noting their need for an empirically derived strategy to curb and redirect stereotypy in emergent situations (i.e., in a parking lot) and necessitating the response reduction assessment. Reductions of motor stereotypy were observed with the inclusion of the hands-down procedure following return to baseline ($M = 98\%$; sessions 25–27), and caregivers' instructions about the programmed contingency and wearable stimuli alone were sufficient to reduce the behavior to 0% in the first two sessions. Reapplication of the stimuli following a brief reversal resulted in quick reductions of motor stereotypy in the natural environment ($M = 93.2\%$). Caregivers reported that the final treatment package was acceptable toward addressing their concerns (i.e., social validity score = 8 out of 10).

Follow-Up

Follow-up data were collected over 3 sessions (i.e., sessions 39–41) approximately 2 weeks following discharge from clinical services via a HIPAA-compliant audio-visual conferencing application (i.e., Zoom) in the natural environment. Results suggested near-zero percentages of motor stereotypy ($M = 0.12\%$), with $M = 0.2$ instances of the hands-down procedure per minute. Caregivers continued to report high social validity scores outside of the training context (i.e., 8 out of 10).

Discussion

The current study sought to extend findings from Verriden and Roscoe (2019) on the use of a contingency-based punishment procedure to address socially significant motor stereotypy. Toby's complex motor movements produced significant safety concerns due to poor environmental awareness when he was engaging in the response (e.g., wandering from caregivers into areas that could jeopardize his safety). The use of a systematic evaluation aided in avoiding contraindications for programming punishment procedures. The primary goal of the response reduction assessment was to evaluate the potential for inadvertent reinforcement (i.e., attention implementing consequent-based procedures contingent on target behavior). Results of the treatment evaluation suggested an 89% reduction from the initial baseline, and a 93% reduction in responding in the community when the hands-down procedure was withdrawn.

The results of the current study should be tempered by its limitations. Despite data corrections for the hands-down procedure, the highest rates of blocking occurred during community generalization sessions, peaking at 1.1 instances per minute (sessions 36–38) when the punishment stimulus was present. These limitations notwithstanding, the current study extends research on empirically derived consequent-based procedures for the reduction of multiply maintained complex motor stereotypy in clinical service delivery settings. Future research should extend the use of these procedures across additional participants in such settings and consider including wearable stimuli to signal the presence of programmed contingencies.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s40617-026-01182-0.

Acknowledgements The authors would like to thank Kim Sloman and Steven Lindauer for input on previous versions of this manuscript. There are no conflicts of interest or funding sources associated with this study. All study materials can be made available by the corresponding author upon request.

Author Contributions The first author was responsible for the study conception and design. Material preparation, data review, and manuscript editing were performed by Thomas "Kris" Franzmann and Joshua Mellott. All authors read and approved the final manuscript.

Funding The authors did not receive support from any organization for the submitted work.

Data Availability De-identified data will be made available by the corresponding author upon reasonable request.

Declarations

Ethics Approval The study was performed in accordance with the ethical standards as laid down in the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards. The study was deemed exempt by the Johns Hopkins University Institutional Review Board (IRB). All participants provided consent for the study procedures.

Conflicts of interest No conflicts of interest have been identified.

References

Ahearn, W. H., Clark, K. M., MacDonald, R. P., & Chung, B. I. (2007). Assessing and treating vocal stereotypy in children with autism. *Journal of Applied Behavior Analysis, 40*(2), 263–275. https://doi.org/10.1901/jaba.2007.30-06

Akers, J. S., Davis, T. N., Gerow, S., & Avery, S. (2020). Decreasing motor stereotypy in individuals with autism spectrum disorder: A systematic review. *Research in Autism Spectrum Disorders, 77*, Article 101611. https://doi.org/10.1016/j.rasd.2020.101611

Behavior Analyst Certification Board. (2020). Ethics code for behavior analysts. https://bacb.com/wp-content/ethics-code-for-behavior-analysts/

Cook, J. L., & Rapp, J. T. (2018). To what extent do practitioners need to treat stereotypy during academic tasks? *Behavior Modification, 44*(2), 228–264. https://doi.org/10.1177/0145445518808226

Cunningham, A. B., & Schreibman, L. (2008). Stereotypy in autism: The importance of function. *Research in Autism Spectrum Disorders, 2*(3), 469–479. https://doi.org/10.1016/j.rasd.2007.09.006

Hagopian, L. P., Falligant, J. M., Frank-Crawford, M. A., Yenokyan, G., Piersma, D. E., & Kaur, J. (2023). Simplified methods for identifying subtypes of automatically maintained self-injury. *Journal of Applied Behavior Analysis, 56*(3), 575–592. https://doi.org/10.1002/jaba.1005

Heyvaert, M., Saenen, L., Campbell, J. M., Maes, B., & Onghena, P. (2014). Efficacy of behavioral interventions for reducing problem behavior in persons with autism: An updated quantitative synthesis of single-subject research. *Research on Developmental Disabilities, 35*(10), 2463–2476. https://doi.org/10.1016/j.ridd.2014.06.017

Iwata, B. A., Dorsey, M. F., Slifer, K. J., Bauman, K. E., & Richman, G. S. (1994). Toward a functional analysis of self-injury. *Journal of Applied Behavior Analysis, 27*(2), 197–209. https://doi.org/10.1901/jaba.1994.27-197

Kennedy, C. H., Meyer, K. A., Knowles, T., & Shukla, S. (2000). Analyzing the multiple functions of stereotypical behavior for students with autism: Implications for assessment and treatment. *Journal of Applied Behavior Analysis, 33*(4), 559–571.

Leaf, J. B., Cihon, J. H., Javed, A., Klick, S., Ferguson, J. L., Milne, C., Creem, A., Arthur, S., Saunders, M. S., Olive, M. L., Ross, R. K., Leaf, R., & McEachin, J. (2022). A call for discussion on stereotypic behavior. *European Journal of Behavior Analysis, 23*(2), 156–180. https://doi.org/10.1080/15021149.2022.2112810

Roantree, C. F., & Kennedy, C. H. (2006). A paradoxical effect of presession attention on stereotypy: Antecedent attention as an establishing, not an abolishing, operation. *Journal of Applied Behavior Analysis, 39*(3), 381–384. https://doi.org/10.1901/jaba.2006

Sarokoff, R. A., & Sturmey, P. (2004). The effects of behavioral skills training on staff implementation of discrete trial teaching. *Journal of Applied Behavior Analysis, 37*(4), 535–538. https://doi.org/10.1901/jaba.2004.37-535

Verriden, A. L., & Roscoe, E. M. (2019). An evaluation of a punisher assessment for decreasing automatically reinforced problem behavior. *Journal of Applied Behavior Analysis, 52*(1), 205–226. https://doi.org/10.1002/jaba.509