

Université de Montréal

Monitorage des mesures physiologiques et des comportements répétitifs associés au stress chez les enfants ayant un trouble du spectre de l'autisme

par Marie-Michèle Dufour

École de psychoéducation
Faculté des arts et des sciences

Thèse présentée

en vue de l'obtention du grade de Philosophiæ Doctor (Ph.D.)

en psychoéducation

Avril 2020

© Marie-Michèle Dufour, 2020

Résumé

Le trouble du spectre de l'autisme se caractérise par la présence de difficultés au plan de la communication sociale et par la présence de comportements répétitifs et d'intérêts restreints (American Psychiatric Association, 2013). Les enfants ayant un TSA présentent plusieurs difficultés concomitantes qui les rendent plus susceptibles de vivre des niveaux de stress élevés, comme des déficits dans la sphère de la communication, de la socialisation et des fonctions exécutives, ainsi que la présence de particularités sensorielles (Grodén et al., 1994, 2005). Malgré que ces enfants soient plus à risque de vivre du stress, plusieurs enjeux méthodologiques rendent difficile sa mesure et plus particulièrement chez ceux qui sont non verbaux. Pour ces raisons, le recours aux mesures physiologiques pour évaluer le stress auprès de cette clientèle est d'une grande pertinence. Par contre, les sensibilités sensorielles de ces enfants pourraient les rendre plus susceptibles de ne pas tolérer ces mesures. Le premier article de cette thèse vise donc à évaluer l'efficacité de l'intervention comportementale renforcement différentiel d'autres comportements (*differential reinforcement of other behavior*; DRO) pour augmenter la tolérance au port d'une ceinture cardiaque chez deux enfants non verbaux ayant un TSA. Les résultats obtenus démontrent que cette intervention a été efficace pour amener ces enfants à augmenter leur tolérance au port du dispositif cardiaque. Un autre aspect qui a reçu beaucoup d'attention dans les dernières années est l'implication du stress dans l'explication des comportements répétitifs et stéréotypés chez les personnes ayant un TSA. Les résultats des études antérieures sont contradictoires (de Vaan et al., 2018; Gabriels et al., 2013; Hutt et al., 1975; Lydon et al., 2015; Yang et al., 2015) et ont principalement utilisé des mesures indirectes des comportements répétitifs. Pour cette

raison, le deuxième article de cet ouvrage vise à évaluer le lien entre le cortisol salivaire, le rythme cardiaque et des mesures d'observations directes de la stéréotypie chez quatre enfants minimalement verbaux ayant un TSA. Les résultats montrent que le cortisol et le rythme cardiaque sont significativement liés à la stéréotypie globale et motrice, sans que ces liens soient observés avec la stéréotypie vocale. Finalement, mesurer directement les comportements stéréotypés demande beaucoup de ressources, ce qui pourrait expliquer la prépondérance des mesures indirectes dans les études. Comme pour l'évaluation du stress, il importe de réfléchir aux méthodes alternatives abordables et accessibles qui pourraient améliorer la mesure de ces comportements. Le troisième article de cette thèse a évalué l'efficacité d'un algorithme d'intelligence artificielle (IA) dans la reconnaissance de la stéréotypie vocale chez des enfants ayant un TSA. Les résultats démontrent que la performance de l'algorithme est supérieure à la reconnaissance due au hasard. Bien que des recherches futures soient nécessaires pour augmenter l'efficacité de cette méthode, l'IA représente une technologie novatrice ayant le potentiel d'améliorer significativement les méthodes actuellement utilisées pour mesurer la stéréotypie. En conclusion, le présent ouvrage explore différentes avenues novatrices pour mieux comprendre et de monitorer la stéréotypie chez les enfants ayant un TSA.

Mots-clés : autisme, stéréotypie, renforcement différentiel d'autres comportements (DRO), cortisol, rythme cardiaque, intelligence artificielle.

Abstract

Autism spectrum disorder (ASD) is characterized by the presence of difficulties in social communication and the presence of repetitive behaviors and restricted interests (American Psychiatric Association, 2013). Children with ASD have several concurrent difficulties, such as deficits in communication, socialization, and executive function, as well as the presence of sensory peculiarities that make them more likely to experience high levels of stress (Groden et al., 2005). Although these children are at increased risk for stress, a number of methodological issues make it difficult to measure, particularly in non-verbal children. For these reasons, the use of physiological measures to assess stress among this group is highly relevant. On the other hand, the sensory sensitivities of these children could potentially make them more likely to be intolerant to these measures. Therefore, the first study in this thesis aims to evaluate the effectiveness of differential reinforcement of other behavior (DRO) to increase compliance with wearing a heart rate monitor in two non-verbal children with ASD. The results obtained portray that this intervention was effective in getting these children to increase their compliance to wearing a cardiac device. Another aspect that has received much attention in recent years is the involvement of stress in explaining repetitive behaviors in individuals with ASD. However, the results of previous studies have been producing contradictory results (de Vaan et al., 2018; Gabriels et al., 2013; Hutt et al., 1975; Lydon et al., 2015; Yang et al., 2015), and have mainly been using indirect measures of stereotypy. For this reason, the second study in this thesis aims to evaluate the relationship between salivary cortisol, heart rate, and direct observational measures of stereotypy in four minimally verbal children with ASD. The results show that cortisol and heart rate are significantly related

to global and motor stereotypy, but not to vocal stereotypy. Finally, measuring stereotypy requires a lot of resources, which could explain the preponderance of indirect measuring in studies on stress. As with the measurement of stress, it is important to consider affordable and alternative methods that could improve the measurement of these behaviors, and therefore the third study evaluated the effectiveness of an artificial intelligence (AI) algorithm in the recognition of vocal stereotypy in children with ASD. The results show that the performance of the algorithm is superior to recognition due to chance. Although future research is needed to increase the effectiveness of this method, AI represents an innovative technology with the potential to significantly improve the methods currently used to measure vocal stereotypy. In conclusion, this thesis explores different innovative methods to better understand and monitor stereotypy in children with ASD.

Keywords: autism, stereotypy, differential reinforcement of other behavior (DRO), cortisol, heart rate, artificial intelligence.

Table des matières

Résumé.....	i
Abstract.....	iii
Table des matières.....	v
Liste des tableaux.....	vii
Liste des figures	viii
Liste des sigles et des abréviations	x
Remerciements.....	xiii
Contribution des auteurs aux articles.....	xvi
Présentation de la thèse.....	1
Chapitre I : Introduction générale	6
Contexte théorique	7
Stress et TSA.....	8
La mesure du stress.....	10
Comportements répétitifs et stéréotypés.....	12
La mesure de la stéréotypie.....	16
Objectifs de la thèse	18
Chapitre II : Article 1	29
Increasing compliance with wearing a medical device in children with autism.....	30
Chapitre III : Article 2	48
Salivary cortisol, heart rate, and direct observation measures of stereotypy	49
in minimally verbal children with autism	49
Chapitre IV : Article 3.....	79
Artificial intelligence for the measurement of vocal stereotypy.....	80
Chapitre V : Discussion générale et conclusion	115
Résumé des principaux résultats empiriques	116

Implications pour la recherche.....	118
Implications pour la pratique	120
Implications pour la psychoéducation	122
Forces et limites de l'étude doctorale	126
Futures études	127
Conclusion	129

Liste des tableaux

Chapitre III

Article 2

Table 1

Participant Characteristics72

Table 2

Operational Definitions for Each Participant's Behaviors73

Table 3

Spearman's Rho Correlations74

Table 4

Summary of Multiple Regression Analysis for Stereotypy Regressed on Mean Cortisol Values and Heart Rate75

Chapitre IV

Article 3

Table 1

Participant Characteristics104

Table 2

Number of Sessions and Duration of Dataset Per Participant105

Table 3

Between-Participant Analyses: Accuracy, Kappa, and Correlation for All Sessions and Those Without Music Sessions.106

Table 4

Within-Participant Analyses: Accuracy, Kappa, and Correlation for Each Participant.
.....107

Table 5

Hybrid Approach: Accuracy, Kappa, and Correlation for Each Participant.....108

Liste des figures

Chapitre II

Article 1

Figure 1. Percentage of time complying with wearing the heart rate monitor (primary y-axis) and frequency of device removal (secondary y-axis) for Leo (upper panel) and Adam (lower panel) during baseline and differential reinforcement of other behavior (DRO) sessions. The values above or below the data points identify the duration of the DRO schedules at the start of sessions. The absence of value indicates that the starting schedule was the same as in the previous session.....47

Chapitre III

Article 2

Figure 1. Level of mean value of cortisol (left panels) and percentage of time engaged in global stereotypy (right panels) across three conditions: Control, novelty and unpredictability.....76

Figure 2. Percentage of time engaged in motor stereotypy (left panels) and vocal stereotypy (right panels) across three conditions: Control, novelty and unpredictability.....77

Figure 3. Mean value of cortisol (ug/dl) and heart rate (HR) compared to percentage of time engaged in global stereotypy, motor stereotypy and vocal stereotypy.....78

Chapitre IV

Article 3

Figure 1. Visual representation of the design of the artificial neural network109

Figure 2. Between-participant analyses: Correlation between the percentages of vocal stereotypy measured by the machine learning algorithm and those measured by the

human observer across all sessions for each participant.....	110
<i>Figure 3. Between-participant analyses: Correlation between the percentages of vocal stereotypy measured by the machine learning algorithm and those measured by the human observer across sessions without music only for each participant.....</i>	<i>111</i>
<i>Figure 4. Within-participant analyses: Correlation between the percentages of vocal stereotypy measured by the machine learning algorithm and those measured by the human observer across sessions for each participant</i>	<i>112</i>
<i>Figure 5. Hybrid analyses: Correlation between the percentages of vocal stereotypy measured by the machine learning algorithm and those measured by the human observer across sessions for each participant.....</i>	<i>113</i>
<i>Figure 6. Correlation between the kappa scores and the percentage of engagement in each session for Billy-Peter.....</i>	<i>114</i>

Liste des sigles et des abréviations

AAC : Analyse appliquée du comportement

ABC : Aberrant Behavior Checklist

AI : Artificial intelligence

APA : American Psychiatric Association

ASD : Autism spectrum disorder

AUCg : Area under the curve with respect to ground

BPI-01 : Behavior problems inventory

CARS-2 : Childhood Autism Rating Scale – Second Edition

CFI : Comparative fit index

DNRO : Differential negative reinforcement of other behavior

DRO : Differential reinforcement of other behavior

DSM : Diagnostic and Statistical Manual

FESP : Faculté des études supérieures et postdoctorales

HHS : Hypothalamo-hypophysio-surrénalien

HR : Heart rate

INSPQ : Institut national de santé publique du Québec

IA : Intelligence artificielle

IOA : Interobserver agreement

IQ : Intelligence quotient

MFCC : Mel filterbank cepstrum coefficient

MO : Motivating operation

NCR : Noncontingent reinforcement

PAD : Potentiel d'adaptation de la personne

PEX : Potentiel expérientiel de l'environnement

QI : Quotient intellectuel

RBS-R : Repetitive Behavior Scale-Revised

RMSEA : Root mean square error of approximation

SNS : Système nerveux sympathique

TLI : Tucker-Lewis index

TSA : Trouble du spectre de l'autisme

À toutes les familles côtoyant de près ou de loin le trouble du spectre de l'autisme. C'est vous qui motivez mon désir de m'investir dans le développement des connaissances dans ce domaine.

À Nan et Gp Guy. Je sais que vous auriez tout fait pour assister à ce moment et qu'en quelque sorte, vous y êtes.

Remerciements

Les remerciements portent non seulement sur l'aboutissement de cette thèse, mais sur l'ensemble des six dernières années qui ont été remplies de défis, d'accomplissements, de grandes joies et de grandes peines. La présence et le soutien de mes proches ont été des éléments cruciaux, qui m'ont portée tout au long de ce long parcours.

Jim.

Ma thèse s'est terminée dans un contexte de distanciation physique. Nous sommes confinés à la maison depuis plus de 50 jours au moment de rédiger ces lignes. Cette situation hors du commun aura mis nos capacités d'adaptation à rude épreuve, mais aura surtout eu comme effet de remettre en lumière tout l'amour, l'unicité et la complicité de notre équipe. Merci d'avoir l'œil sur l'horizon et de voir au loin ce qui n'est pas perceptible pour moi. Merci pour ton soutien constant, ta présence, ton calme. Cette thèse, c'est aussi un peu la tienne, parce qu'elle n'aurait jamais abouti sans toi. Trois, quatre p'tites fins de semaine de job plus tard...infini merci.

Clémence.

Ma fille. Ton arrivée et la création de notre famille est sans contredit la plus belle chose qui me soit donnée de vivre. L'accomplissement de cette thèse confirme qu'il est possible d'aller au bout de plusieurs rêves simultanément. Mon plus grand souhait pour maintenant, et pour les années à venir, est de te démontrer qu'avec de la volonté et de la persévérance tout est possible.

Mes parents, Jérôme et Jean Ann.

C'est en évoluant dans l'environnement sain et aimant que vous avez bâti pour nous que j'ai pu construire des bases solides qui me permettent aujourd'hui de me

propulser et de constamment repousser mes limites. Merci pour votre amour inconditionnel et pour votre soutien constant. Ces deux éléments valent tout l'or du monde.

Mathieu.

Mon frère, nos grandes discussions *deep* sur la vie, autant que nos fous rires interminables ont été essentiels au maintien de mon équilibre. C'est un privilège pour moi d'être aux premières loges et de te voir évoluer et réaliser tes rêves. Merci pour tout, on t'aime Payain.

Mes stores.

À l'aube de la trentaine, rares sont les gens qui peuvent compter sur des amitiés vieilles d'un quart de siècle et je suis privilégiée d'avoir cette chance. La vie a mis sur notre chemin des épreuves de taille dans les dernières années qui ont mis en lumière la puissance de notre soutien mutuel. Merci d'être qui vous êtes et de faire partie de ma vie. Audri et Joe, votre présence et votre soutien ont fait une différence significative, mille mercis.

Marie-Philippe.

Partager les bons et particulièrement les moins bons moments avec toi ont rendu ce parcours tellement plus doux et agréable. Merci pour tout.

Sur une note plus académique et formelle :

Marc.

Si on additionne les années de maîtrise et de doctorat, ça fait près d'une décennie que tu diriges mes travaux (!!). Merci de m'avoir poussé à développer mes capacités à

leur plein potentiel. Merci pour ta rigueur et pour tes encouragements. Merci surtout de n'avoir jamais douté de l'aboutissement de cette thèse.

Mes collègues du laboratoire de recherche et plus particulièrement Valérie, Malena et Isabelle.

Ce fût extrêmement stimulant d'évoluer à vos côtés dans les dernières années. Valérie, merci pour les multiples conversations. Ta vision claire, ton objectivité et ta grande intelligence m'ont éclairée face à plusieurs situations autant personnelles qu'académiques. Malena, tu as fait une différence significative dans mes premières années de doctorat, merci pour tout! Isabelle, merci pour ta présence, pour l'espace offert lorsque j'avais besoin de ventiler... et pour le thé!

Lyanne Levasseur, Emily Heng, Virginie Chevrère et Lydia Trudel.

Merci pour l'aide à la collecte et à la cotation des données, votre implication a fait une énorme différence.

Sur une note moins formelle, merci au Café Léopard de m'avoir offert un refuge confortable pour rédiger. Merci à Alexandra Stréliski d'avoir bercé mes angoisses et d'avoir adouci mes maintes et maintes heures de rédaction.

Finalement, merci aux familles qui ont participé à cette thèse. Je sais que ce n'est pas facile de trouver du temps pour participer à des projets de recherche, particulièrement de le faire de manière bénévole. Merci de tout cœur, c'est grâce à vous si on peut continuer de faire avancer les connaissances.

Contribution des auteurs aux articles

Conformément aux règlements de la Faculté des études supérieures et postdoctorales (FESP), l'auteur de la thèse a contribué de manière « essentielle, majeure et déterminante » aux trois articles que constitue cet ouvrage. Plus particulièrement, j'ai planifié et conçu les protocoles des trois articles. J'ai collecté toutes les données du premier article et plus de la moitié des données du second. Le troisième article s'est basé sur des données secondaires que j'ai recodées entièrement. J'ai procédé à l'analyse des données des trois études et j'ai rédigé les trois articles. Marc Lanovaz a supervisé l'ensemble de mes travaux, tout en m'offrant un soutien essentiel à l'accomplissement de l'ensemble des étapes ayant mené au présent dépôt, en plus d'écrire le code Python qui a permis de mener les analyses du troisième article. Pierrich Plusquellec m'a soutenue avec son expertise du cortisol et Patrick Cardinal m'a aidée dans la sélection des méthodes d'analyse en ce qui a trait à l'extraction et l'analyse des sons du dernier article.

Présentation de la thèse

Le présent ouvrage a été rédigé par articles et contient cinq principaux chapitres : l'introduction, les trois articles empiriques et la discussion générale. Le premier chapitre présente le contexte théorique dans lequel s'insère la présente thèse. Les principaux sujets traités sont : les caractéristiques de l'autisme, le stress, les mesures physiologiques, les comportements répétitifs et stéréotypés et l'intelligence artificielle. Finalement, les trois objectifs de la thèse sont présentés : (a) déterminer l'efficacité d'une intervention comportementale pour augmenter la tolérance au port d'une ceinture cardiaque chez des enfants non verbaux ayant un TSA, (b) évaluer la relation entre le cortisol salivaire, le rythme cardiaque et des mesures d'observations directes de la stéréotypie chez des enfants minimalement verbaux ayant un TSA et (c) déterminer l'efficacité d'un algorithme d'intelligence artificielle dans l'identification de la stéréotypie vocale chez des enfants ayant un TSA.

Le deuxième chapitre présente l'article intitulé « Increasing compliance with wearing a medical device in children with autism », qui a été publié dans le *Journal of Applied Behavior Analysis* (voir Dufour & Lanovaz, 2020). L'article visait à évaluer l'efficacité d'une intervention comportementale pour augmenter la tolérance au port d'un dispositif cardiaque chez deux enfants ayant un TSA. L'idée de cet article est venue à la suite de difficultés lors de la collecte de données du deuxième article. L'objectif ultime de cette étude était d'améliorer et de faire avancer les possibilités de prises de mesures physiologiques chez les populations présentant des sensibilités sensorielles. De ce fait, cet article est une reproduction de trois études précédentes (Cook et al., 2015; DeLeon et al., 2008; Richling et al., 2011) qui avaient

démontré l'efficacité d'interventions comportementales pour augmenter la tolérance à différents dispositifs médicaux (bracelets médicaux, lunettes et orthèses) et d'étendre ces résultats au port d'un dispositif cardiaque.

Dans le chapitre trois, nous avons évalué le lien entre les comportements stéréotypés, le cortisol salivaire et le rythme cardiaque chez des enfants minimalement verbaux ayant un TSA. Le titre de l'article est : « Salivary cortisol, heart rate measures and behavioral measures of stereotypy in children with autism spectrum disorder » et sera soumis sous peu à une revue spécialisée en autisme. Cette étude a été motivée par l'inconsistance des résultats des études précédentes qui ont évalué le lien entre le stress et les comportements répétitifs (Gabriels et al., 2013; García-Villamizar & Rojahn, 2015; Hutt et al., 1975; Lydon et al., 2013; Lydon, Healy, Mulhern et Hughes, 2015; Lydon, Healy, Roche et al., 2015; Willemsen-Swinkels et al., 1998; Yang et al., 2015).

La mesure directe de ces comportements dans le deuxième article a mis en lumière les limites inhérentes à cette méthode de prise de mesure (ressources humaines et monétaires) qui restreignent considérablement ses possibilités d'utilisation. C'est ce constat qui a motivé la création du dernier article qui visait à explorer des méthodes alternatives abordables et accessibles pour mesurer ces comportements. Plus particulièrement le troisième article de cet ouvrage «Artificial intelligence for the measurement of stereotypy» avait comme objectif d'évaluer l'efficacité d'un algorithme d'intelligence artificielle dans la détection de la stéréotypie vocale chez des enfants ayant un TSA. Cet article a été accepté pour publication dans la revue *Journal of the Experimental Analysis of Behavior*. La reconnaissance automatique de ces comportements rendrait possible la prise de données rigoureuses en contexte clinique et permettrait de réduire considérablement les ressources nécessaires à la

mise en place de projets de recherche sur le stress qui devraient inclure des mesures de stéréotypie vocale.

En conclusion, le dernier chapitre résume les principaux résultats des trois études, présente les forces et les faiblesses méthodologiques de la thèse et propose des idées et des pistes de recherches futures. L'accent est mis sur les implications au plan de la recherche et au plan clinique, tout en contextualisant l'ensemble de cette thèse dans la réalité psychoéducative.

Références

- Dufour, M.-M., & Lanovaz, M. J. (2019). Increasing compliance with wearing a medical device in children with autism. *Journal of Applied Behavior Analysis, 53*(2), 1089-1096. <https://doi.org/10.1002/jaba.628>
- Cook, J. L., Rapp, J. T., & Schulze, K. A. (2015). Differential negative reinforcement of other behavior to increase wearing of a medical bracelet. *Journal of Applied Behavior Analysis, 48*(4), 901-906. <https://doi.org/10.1002/jaba.228>
- DeLeon, I. G., Hagopian, L. P., Rodriguez-Catter, V., Bowman, L. G., Long, E. S., & Boelter, E. W. (2008). Increasing wearing of prescription glasses in individuals with mental retardation. *Journal of Applied Behavior Analysis, 41*(1), 137-142. <https://doi.org/10.1901/jaba.2008.41-137>
- Gabriels, R. L., Agnew, J. A., Pan, Z., Holt, K. D., Reynolds, A., & Laudenslager, M. L. (2013). Elevated repetitive behaviors are associated with lower diurnal salivary cortisol levels in autism spectrum disorder. *Biological Psychology, 93*(2), 262-268. <https://doi.org/10.1016/j.biopsycho.2013.02.017>
- García-Villamizar, D., & Rojahn, J. (2015). Comorbid psychopathology and stress mediate the relationship between autistic traits and repetitive behaviours in adults with autism: Mediators of repetitive behaviour in autism. *Journal of Intellectual Disability Research, 59*(2), 116-124. <https://doi.org/10.1111/jir.12083>
- Hutt, C., Forrest, S. J., & Richer, J. (1975). Cardiac arrhythmia and behaviour in autistic children. *Acta Psychiatrica Scandinavica, 51*(5), 361-372. <https://doi.org/10.1111/j.1600-0447.1975.tb00014.x>

- Lydon, S., Healy, O., & Dwyer, M. (2013). An examination of heart rate during challenging behavior in autism spectrum disorder. *Journal of Developmental and Physical Disabilities, 25*(1), 149-170. <https://doi.org/10.1007/s10882-012-9324-y>
- Lydon, S., Healy, O., Mulhern, T., & Hughes, B. M. (2015). Heart rate measurement during stereotyped motor behavior in autism spectrum disorder. *Journal of Developmental and Physical Disabilities, 27*(5), 677-699. <https://doi.org/10.1007/s10882-015-9445-1>
- Lydon, S., Healy, O., Roche, M., Henry, R., Mulhern, T., & Hughes, B. M. (2015). Salivary cortisol levels and challenging behavior in children with autism spectrum disorder. *Research in Autism Spectrum Disorders, 10*, 78-92. <https://doi.org/10.1016/j.rasd.2014.10.020>
- Richling, S. M., Rapp, J. T., Carroll, R. A., Smith, J. N., Nystedt, A., & Siewert, B. (2011). Using noncontingent reinforcement to increase compliance with wearing prescription prostheses. *Journal of Applied Behavior Analysis, 44*(2), 375-379. <https://doi.org/10.1901/jaba.2011.44-375>
- Willemsen-Swinkels, S. H. N., Buitelaar, J. K., Dekker, M., & van Engeland, H. (1998). Subtyping stereotypic behavior in children: The association between stereotypic behavior, mood, and heart rate. *Journal of Autism and Developmental Disorder, 28*, 547-557. <https://doi.org/10.1023/A:1026008313284>
- Yang, C.-J., Tan, H.-P., Yang, F.-Y., Wang, H.-P., Liu, C.-L., He, H.-Z., Sang, B., Zhu, X.-M., & Du, Y.-J. (2015). The cortisol, serotonin and oxytocin are associated with repetitive behavior in autism spectrum disorder. *Research in Autism Spectrum Disorders, 18*, 12-20. <https://doi.org/10.1016/j.rasd.2015.07.002>

Chapitre I : Introduction générale

Contexte théorique

Le trouble du spectre de l'autisme (TSA) est un trouble neurodéveloppemental qui se caractérise par la présence de difficultés au plan de la communication sociale et par la présence de comportements répétitifs et d'intérêts restreints (American Psychiatric Association, 2013). La communication sociale réfère à la réciprocité socioémotionnelle, aux comportements de communication non verbale ainsi qu'au développement, au maintien et à la compréhension des relations interpersonnelles (APA, 2013). Pour sa part, la deuxième catégorie de difficultés comprend des mouvements moteurs ou vocaux répétitifs et stéréotypés, une adhérence aux routines et aux rituels, des intérêts restreints et souvent intenses et la présence de sensibilités sensorielles (hypo ou hyper réactivité; APA, 2013).

La publication de la cinquième version du *Diagnostic and Statistical Manual* en 2013 (*DSM-5*) a entraîné son lot de nouveautés dans la conception et la définition du TSA. Un des changements majeurs est le fait d'aborder le trouble selon une conception dimensionnelle, ce qui permet une meilleure représentation du caractère hétérogène de la symptomatologie autistique. Les personnes qui se trouvent sur le spectre de l'autisme présentent des niveaux de fonctionnement qui peuvent différer significativement. Certaines personnes qui ont un TSA ont un niveau de fonctionnement qui leur permet d'acquérir l'autonomie nécessaire à une intégration sociale complète (cursus scolaire régulier, avoir et maintenir un emploi). À l'opposé, certaines personnes qui ont un TSA auront besoin d'un soutien substantiel tout au long de leur vie. Notamment, des études montrent qu'entre 30% et 70% des enfants ayant un TSA présentent aussi une déficience intellectuelle, fréquemment associée à une absence ou à un retard significatif du langage (quotient intellectuel (QI) < 70; Bertrand et al., 2001; Chakrabarti & Fombonne, 2001, 2005; Fombonne, 2009).

La prévalence du TSA a connu une progression significative dans les dernières années. Le taux annuel d'augmentation de la prévalence chez les enfants âgés entre 2 et 14 ans au Canada a été évalué comme variant entre 9,7 % à 14,6 % (Ouellette-Kuntz et al., 2014). En 2015, la prévalence au Québec était évaluée à 1,6 % chez la population âgée entre 5 et 17 ans, ce qui représente près de 17 000 enfants (Agence de Santé Publique du Canada, 2018). La différence entre les sexes est une particularité qui mérite d'être soulignée : il y a quatre à cinq fois plus de garçons que de filles qui reçoivent un diagnostic de TSA (INSPQ, 2017). La progression de la prévalence des dernières années s'expliquerait entre autres par l'évolution des critères diagnostiques, une meilleure connaissance du trouble, une meilleure accessibilité aux services, la substitution de diagnostic et la méthodologie des études (Campbell et al., 2011). Sans que ce soit possible d'éliminer l'hypothèse d'une augmentation réelle, il est certain que l'ensemble de ces facteurs contribue à l'augmentation observée et qu'il n'y a probablement pas d'*épidémie* d'autisme à proprement dit (Campbell et al., 2011).

Stress et TSA

Plusieurs déficits associés au TSA rendent ces individus plus susceptibles de vivre du stress (communication, socialisation, fonctions exécutives, sensibilités sensorielles; Groden et al., 1994, 2005). Les difficultés à établir et à maintenir des relations sociales, à comprendre les règles sociales non explicites, à décoder les expressions faciales et à saisir adéquatement les informations à double sens, l'ironie et le sarcasme contribuent à rendre ces personnes à risque de percevoir diverses situations sociales comme étant stressantes (Volkmar, 2011). Dans les années 1920, Hans Selye a été le premier chercheur à définir le stress comme étant une réponse non-spécifique du corps causée par des irrégularités dans son fonctionnement (Tan & Yip, 2018). Selye nomme cette réaction le Syndrome général de l'adaptation dont il décrit

trois phases principales : la réaction générale d'alarme, la phase de résistance et la phase d'épuisement (Tan & Yip, 2018).

Suite aux travaux de Selye (1974) portant sur le stress biologique, plusieurs auteurs ont tenté de définir le stress psychologique. Une multitude de définitions ont été prêtées à ce concept dont nombre d'entre elles ont été critiquées comme étant trop larges et difficiles à opérationnaliser (Grant et al., 2003). L'influence des circonstances environnementales représente un point commun à l'ensemble de ces définitions. La plus couramment utilisée est celle de Lazarus et Folkman (1984) introduisant la notion d'interactionnisme entre l'individu et son environnement : « Le stress psychologique implique une relation particulière entre la personne et son environnement qui est évaluée par la personne comme excédant ses ressources en mettant son bien-être en danger ». Plus récemment, des chercheurs ont mis en lumière quatre éléments liés à une réponse de stress : le manque de contrôle, l'imprévisibilité, la nouveauté et l'atteinte à l'égo (Dickerson & Kemeny, 2004). Lorsque l'un ou plusieurs de ces éléments sont présents dans l'évaluation de l'environnement faite par la personne, une réaction physiologique de stress est déclenchée.

Les enfants ayant un TSA ont moins de facilité à décoder le monde social qui les entoure et surtout à le prévoir; ils sont donc plus à risque d'être confrontés à des situations imprévisibles. De plus, la nouveauté peut être difficile à gérer pour ceux d'entre eux qui présentent une adhérence aux routines et aux rituels. Selon certains chercheurs, les enfants ayant un TSA présenteraient une suractivation du cortex pariétal lorsque soumis à la nouveauté en comparaison avec les enfants neurotypiques, les rendant ainsi plus sensibles à cet ingrédient du stress (Sokhadze et al., 2009). Finalement, ces individus sont plus vulnérables à la stigmatisation et à la victimisation, qui impliquent directement une menace à

l'égo (Sentenac et al., 2011). À la lumière de ces informations, il est possible d'affirmer que les personnes ayant un TSA sont plus à risque de présenter des niveaux de stress élevés.

Le vécu d'un stress élevé en fréquence et en intensité peut avoir de nombreux impacts sur le développement et le fonctionnement des enfants ayant un TSA. Le stress représente un prédicteur au développement des troubles anxieux, qui serait la comorbidité la plus commune chez cette population (INSPQ, 2017; Vasa et al., 2014). Le stress vécu de manière répétitive peut aussi avoir des effets négatifs sur le cerveau en développement qui peuvent entraîner des conséquences à long terme, telles que la difficulté à acquérir de nouvelles connaissances et à s'adapter aux situations adverses (Shonkoff et al., 2012).

La mesure du stress

Malgré que cette population soit plus à risque de vivre du stress, plusieurs enjeux méthodologiques rendent difficile sa mesure chez les enfants ayant un TSA et plus particulièrement chez ceux qui sont non verbaux. Ces enfants n'ont pas les capacités requises pour répondre aux types de mesure du stress couramment utilisés comme les questionnaires autorapportés ou les entrevues. Ces considérations méthodologiques sont à prendre très au sérieux considérant qu'entre 30% et 70% des enfants ayant un TSA présentent une déficience intellectuelle, fréquemment associée à une absence ou à un retard significatif du langage (QI < 70; Bertrand et al., 2001; Chakrabarti & Fombonne, 2001, 2005; Fombonne, 2009).

Pour ces raisons, le recours aux mesures physiologiques pour évaluer le stress auprès de cette clientèle est d'une grande pertinence. La réponse physiologique de stress est bien étudiée chez l'humain. Brièvement, elle consiste en une cascade d'hormones, dont la principale est le cortisol (Taylor & Corbett, 2014). Souvent appelé l'hormone du stress, le cortisol est une mesure de l'activation de l'axe hypothalamo-hypophysio-surrénalien (HHS)

qui s'active dans les minutes (voire les heures) suivant le stress perçu. En s'activant, cet axe déclenche une séquence de sécrétion d'hormones, impliquant ainsi le système limbique, l'hypothalamus, l'hypophyse et les glandes surrénales (Gunnar & Adam, 2012). Le cortisol sécrété dans le sang est à la base de la réaction de *fuite ou combat*, qui a pour but de préparer le corps à fuir un danger perçu ou à le combattre. Le corps humain sécrète cette hormone quotidiennement de manière normative, mais la concentration du cortisol augmente lors de la perception d'une menace. La sécrétion normative du cortisol suit un rythme circadien. Sa plus grande production se produit dans la deuxième moitié de la nuit pour atteindre son sommet de 30 à 40 minutes après le réveil. Par la suite, le cortisol diminue tout au long de la journée pour atteindre son niveau le plus bas durant la première partie de la nuit. (Fries et al., 2009).

Par ailleurs, lors de la réaction de *fuite ou combat*, le système nerveux sympathique (SNS) est activé, ce qui entraîne la stimulation du système cardiovasculaire, du système immunitaire, des glandes endorphines et des régions cérébrales associées à la mémoire et aux émotions (Jansen et al., 1995). Donc, la fréquence cardiaque a elle aussi été identifiée comme étant une mesure valide pour évaluer les différences individuelles de réactivité à des stressors environnementaux (Grodén et al., 2005).

Bien que les mesures physiologiques soient d'une grande pertinence pour mesurer le stress chez la population ayant un TSA, les particularités sensorielles de ces enfants pourraient les rendre plus susceptibles de ne pas tolérer les mesures (p. ex. contact du coton dans la bouche pour le cortisol salivaire, port d'une ceinture cardiaque). Même si plusieurs études incluant des mesures physiologiques ont été menées auprès de cette population, à notre connaissance, aucune étude ne s'est intéressée à cet enjeu précis (Heathers et al., 2019; Magnuson et al., 2019; Muscatello & Corbett, 2017) Or, des chercheurs ont démontré que

certaines interventions comportementales sont efficaces pour aider les enfants à accepter de porter différents dispositifs comme des lunettes, des orthèses et des dispositifs auditifs (Cook, et al., 2015; DeLeon et al., 2008; Richling et al., 2011). Ainsi, nous avons tenté d'étendre ces résultats au port d'une ceinture cardiaque pour mesurer le stress des enfants non verbaux ayant un TSA, ce qui fait l'objet du premier article de cette thèse. La validation d'une intervention pour faciliter le port de la ceinture cardiaque pourrait avoir une portée significative dans la mesure du stress chez les enfants qui n'ont pas les capacités pour répondre aux mesures autorapportées couramment utilisées.

Comportements répétitifs et stéréotypés

Les manifestations, le fonctionnement physiologique et les moyens pour réduire le stress sont des sujets largement étudiés chez la population ayant un TSA (Bitsika et al., 2018; Hollock et al., 2014; Viau et al., 2010). Un autre aspect qui a reçu beaucoup d'attention dans les dernières années est l'implication du stress dans l'explication des comportements répétitifs et stéréotypés chez les personnes ayant un TSA. De par la définition des critères diagnostiques, toutes les personnes qui ont un TSA présentent aussi une forme de comportements répétitifs et stéréotypés (Lewis & Bodfish, 1998). Des auteurs ont montré que ces individus présentent des comportements répétitifs de nature variée qui sont compris dans les catégories suivantes : la stéréotypie, l'automutilation, les compulsions, les tics, les rituels, les comportements de rigidité ainsi que les comportements restreints (Bodfish et al., 2000). Bien que les comportements répétitifs et stéréotypés aient fait l'objet de plusieurs recherches dans les dernières années, plusieurs zones grises demeurent. En effet, plusieurs hypothèses sont émises quant à la fonction de ces comportements sans qu'aucune ne fasse consensus. Les trois principales hypothèses quant à la fonction de ces comportements sont (1)

l'autostimulation ou le renforcement automatique, (2) la manifestation du stress et (3) l'autorégulation.

La première hypothèse découle de l'approche comportementale et soutient que les comportements répétitifs sont maintenus par un renforcement automatique positif puisque l'apparition du comportement en lui seul contribue à son maintien (Lanovaz & Sladeczek, 2012; Lovaas et al., 1987; Rapp & Vollmer, 2005). Selon cette hypothèse, l'occurrence du comportement à elle seule produit un renforcement interne (automatique), qui est indépendant des conditions externes à l'individu (Vollmer, 1994). Lanovaz (2011) propose un modèle explicatif de la stéréotypie qui inclut des variables neurologiques et comportementales. Selon ce modèle, la stéréotypie serait le produit d'une anomalie dans les processus neurologiques, qui ferait en sorte que les comportements provoqueraient un plaisir (renforcement automatique). À titre d'exemple, l'autostimulation vocale ne procure pas de plaisir à une personne neurotypique, mais une anomalie dans le système nerveux central des personnes ayant un TSA mènerait à une plus grande sécrétion de dopamine lors de la production de ce comportement, ce qui entraînerait la personne à le produire à une fréquence excessive. Selon cette perspective, les comportements répétitifs seraient indépendants de l'état psychologique de la personne.

La deuxième hypothèse soutient que les comportements répétitifs sont une manifestation du stress, de l'excitation ou de la fatigue vécus par ces enfants (Muthugovindan & Singer, 2009). En effet, des chercheurs ont publié des résultats démontrant que la présence de comportements répétitifs serait associée à un niveau de cortisol plus élevé (de Vaan et al., 2018; Lydon, Healy, Roche et al., 2015; Yang et al., 2015). Selon ces auteurs, les comportements répétitifs seraient davantage une manifestation physique de l'état

psychologique des enfants. Une autre étude a trouvé une corrélation positive entre la stéréotypie et le niveau de cortisol pour les enfants âgés entre six et 12 ans, sans toutefois observer ce lien pour ceux âgés entre 13 à 18 ans (Bitsika et al., 2015). Ces résultats suggèrent que le lien entre le stress et la stéréotypie pourrait varier selon les périodes développementales. Récemment, García-Villamizar et Rojahn (2015) ont démontré que le stress pourrait potentiellement modérer la relation entre les symptômes autistiques et la présence de comportements répétitifs.

La troisième hypothèse soutient que les comportements répétitifs répondent à un besoin sensoriel et que le fait de manifester ces comportements pourrait contribuer à l'autorégulation de ces individus. Des études portant sur des animaux ont démontré que les comportements répétitifs pourraient représenter une stratégie de coping en réponse à une exposition à un environnement stressant (Langen et al., 2011; Lewis et al., 2007). Groden et al. (1994) avancent que, bien qu'inadaptés, ces comportements représentent une stratégie fonctionnelle pour répondre à un stresser et ainsi, faire diminuer le niveau de stress ressenti. Selon cette perspective, les comportements répétitifs auraient une fonction régulatrice. Soutenant cette hypothèse, les résultats d'une étude ont illustré que la fréquence cardiaque de neuf enfants ayant un TSA diminuait suite à l'occurrence de comportements répétitifs (Hutt et al., 1975). Dans le même ordre d'idées, une étude a montré que les enfants ayant un TSA qui présentent un niveau élevé de stéréotypie auraient des niveaux de cortisol plus faibles (Gabriels et al., 2013). Les auteurs de cette recherche avancent deux hypothèses explicatives : soit la stéréotypie a une fonction d'autorégulation, soit le niveau du cortisol est faible, dû à une exposition prolongée au stress. Ces derniers résultats sont en contradiction

avec ceux qui soutiennent que la stéréotypie est associée à un niveau élevé de stress (de Vaan et al., 2018; Lydon, Healy, Roche et al., 2015; Yang et al., 2015).

Une explication pour réconcilier ces résultats contradictoires est celle selon laquelle la fonction de ces comportements serait multiple, incluant l'élévation du niveau physiologique durant des périodes de faible stimulation environnementale, le relâchement d'une excitation trop grande ou la prévention de stimulation additionnelle lors de périodes où l'excitation physiologique est élevée (Lydon, Healy, Mulhern, & Hughes, 2015). Une autre étude a démontré que les comportements répétitifs pourraient répondre à des fonctions différentes selon l'état émotionnel de la personne au moment d'émettre les comportements (Willemsen-Swinkels et al., 1998). Devant cette inconsistance des connaissances relatives au lien entre les comportements répétitifs et le stress, plusieurs auteurs font appel à une investigation plus approfondie de leur fonction (Gabriels et al., 2013; Lydon, Healy, Roche et al., 2015).

Comme mentionné précédemment, les comportements répétitifs et stéréotypés comprennent plusieurs catégories (stéréotypies, automutilations, compulsions, tics, rituels, comportements de rigidité, comportements restreints; Bodfish et al., 2000). Le fait qu'une multitude de comportements tombe sous la même catégorie peut, selon nous, permettre d'expliquer en partie la divergence des résultats passés. Pour cette raison, le deuxième article de cette thèse s'est centré uniquement sur la sous-catégorie de la stéréotypie et plus particulièrement sur les formes vocales et motrices. La stéréotypie vocale se définit par des vocalisations répétitives qui persistent en l'absence de conséquences sociales (Ahearn et al., 2007; Lanovaz & Sladeczek, 2012), alors que la stéréotypie motrice est définie comme étant des comportements structurés, répétitifs, coordonnés et rythmiques (Mahone et al., 2004).

Le type de mesure utilisé peut aussi avoir une influence sur les différents résultats. En effet, la vaste majorité des études passées ont mesuré les comportements répétitifs à l'aide de questionnaires. Puisque les informations amassées par ces outils proviennent d'observateurs externes (parents, éducateurs, donneurs de soin), un biais inhérent sous-tend ce type de mesure (van de Mortel, 2008). À notre connaissance, aucune étude n'a mesuré la relation entre le cortisol salivaire et des mesures d'observations directes de la stéréotypie. De plus, aucune étude à notre connaissance n'a combiné plusieurs mesures physiologiques (cortisol et fréquence cardiaque) à des mesures d'observations directes des comportements stéréotypés. Pour ces raisons, l'objectif du deuxième article est d'évaluer la relation entre le cortisol salivaire, le rythme cardiaque et des mesures d'observations directes de la stéréotypie chez des enfants minimalement verbaux ayant un TSA.

La mesure de la stéréotypie

Comme mentionné précédemment, plusieurs défis entourent la mesure du stress chez les personnes ayant un TSA et il est parfois nécessaire d'avoir recours à des méthodes alternatives, telles que les mesures physiologiques. Au-delà des défis spécifiques à la mesure du stress, il y a plusieurs défis inhérents à la mesure des comportements de manière plus large chez les personnes qui ont un TSA et les comportements stéréotypés n'en font pas exception. Ces difficultés liées à la mesure de la stéréotypie pourraient permettre d'expliquer l'absence d'étude sur le stress qui inclut des mesures d'observations directes.

Dans plusieurs situations, la mesure comportementale de la stéréotypie par un observateur humain soulève une problématique importante liée aux ressources nécessaires et aux coûts qui y sont associés. Si le comportement ciblé est présent à une fréquence très élevée, l'observateur doit généralement noter le comportement à partir d'un enregistrement vidéo

puisque ce n'est pas possible de le faire lors de la séance d'observation. Pour assurer la fidélité de la mesure, un deuxième observateur indépendant est aussi requis pour assurer un accord interjuges (Mudford et al., 2009). Considérant la réalité des milieux de pratique, il est irréaliste de s'attendre à ce que des données soient collectées de manière aussi rigoureuse, ce qui représente une barrière au suivi objectif et rigoureux des effets des interventions psychoéducatives. Les défis liés à la collecte ont aussi un impact sur le développement des connaissances. Les ressources et les coûts importants qui y sont associés limitent le nombre de recherches qui peuvent être menées.

Comme pour la mesure du stress, il importe de réfléchir aux méthodes alternatives qui pourraient améliorer la mesure de ces comportements. Une solution prometteuse à cette problématique est l'intelligence artificielle (IA). La portée et les possibilités des algorithmes d'IA sont bien connues à ce jour et ils arrivent aujourd'hui à accomplir des tâches dont on ne les aurait pas cru capables il y a de cela quelques années (p. ex. la reconnaissance des émotions; Ko., 2018). Comme l'IA l'a fait dans maints autres domaines, il est logique de penser que la puissance de ces algorithmes puisse aussi venir améliorer, voire révolutionner le domaine de la mesure des comportements. Un des champs d'études de l'IA est l'apprentissage machine. Basés sur des modèles statistiques, ces algorithmes arrivent à faire ressortir des patrons dans des ensembles de données complexes.

Quelques études ont évalué l'efficacité d'algorithmes d'apprentissage machine dans la reconnaissance de comportements moteurs stéréotypés (Goodwin et al., 2011; Min & Tewfik, 2010; Rad & Furlanello, 2016; Westeyn et al., 2005). Par contre, aucune étude à notre connaissance n'avait évalué l'efficacité de ces algorithmes dans la reconnaissance de la stéréotypie vocale. Puisque 48% des personnes ayant un retard du développement présentent

une forme vocale de comportements stéréotypés (Chebli et al., 2016), l'automatisation de la prise de données des comportements stéréotypés vocaux permettrait de diminuer significativement les ressources et les coûts qui sont nécessaires actuellement pour collecter ces données. De ce fait, plus de recherches pourraient être menées et il deviendrait réaliste que ce type de comportements soit mesuré dans des contextes cliniques.

Objectifs de la thèse

La présente thèse doctorale vise : (a) à déterminer l'efficacité d'une intervention comportementale pour augmenter la tolérance au port d'une ceinture cardiaque chez deux enfants non verbaux ayant un TSA, (b) à évaluer la relation entre le cortisol salivaire, le rythme cardiaque et des mesures d'observations directes de la stéréotypie chez des enfants minimalement verbaux ayant un TSA et (c) à déterminer l'efficacité d'un algorithme d'intelligence artificielle dans l'identification de la stéréotypie vocale chez des enfants ayant un TSA.

Références

- Agence de Santé Publique (2018). *Trouble du spectre de l'autisme chez les enfants et les adolescents au Canada 2018, Un rapport du système national de surveillance du trouble du spectre de l'autisme*. Repéré à : <https://www.canada.ca/fr/sante-publique/services/publications/maladies-et-affections/trouble-spectre-autisme-enfants-adolescents-canada-2018.html#a3-5>
- 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>
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (5th ed.)*. American Psychiatric Publishing.
- Bertrand, J., Mars, A., Boyle, C., Bove, F., Yeargin-Allsopp, M., & Decoufle, P. (2001). Prevalence of autism in a United States population: The brick township, New Jersey, investigation. *Pediatrics, 108*(5), 1155-1161. <https://doi.org/10.1542/peds.108.5.1155>
- Bitsika, V., Sharpley, C. F., Agnew, L. L., & Andronicos, N. M. (2015). Age-related differences in the association between stereotypic behaviour and salivary cortisol in young males with an autism spectrum disorder. *Physiology & Behavior, 152*, 238-243. <https://doi.org/10.1016/j.physbeh.2015.10.010>
- Bitsika, V., Sharpley, C. F., McMillan, M. E., & Agnew, L. L. (2018). Background cortisol versus social anxiety as correlates of HPA-axis recovery from stress in boys with autism spectrum disorder. *International Journal of Developmental Neuroscience, 71*(1), 52-60. <https://doi.org/10.1016/j.ijdevneu.2018.08.004>

- Bodfish, J. W., Symons, F. J., Parker, D. E., & Lewis, M. H. (2000). Varieties of repetitive behavior in autism: Comparisons to mental retardation. *Journal of Autism and Developmental Disorders, 30*(3), 237-243. <https://doi.org/10.1023/a:1005596502855>
- Campbell, C. A., Davarya, S., Elsabbagh, M., Madden, L., & Fombonne, E. (2011). Prevalence and the controversy. In J.L. Matson & P. Sturmey (Eds.), *Autism and child psychopathology series, international handbook of autism and pervasive developmental disorders* (pp. 25-35). https://doi.org/10.1007/978-1-4419-8065-6_3
- Chakrabarti, S., & Fombonne, E. (2001). Pervasive developmental disorders in preschool children. *Journal of the American Medical Association, 285*, 3093-3099. <https://doi.org/10.1001/jama.285.24.3093>
- Chakrabarti, S., & Fombonne, E. (2005). Pervasive developmental disorders in preschool children: Confirmation of high prevalence. *American Journal of Psychiatry, 162*, 1133-1141. <https://doi.org/10.1176/appi.ajp.162.6.1133>
- Chebli, S. S., Martin, V., & Lanovaz, M. J. (2016). Prevalence of stereotypy in individuals with developmental disabilities: A systematic review. *Journal of Autism and Developmental Disorders, 3*(2), 107-118. <https://doi.org/10.1007/s40489-016-0069-x>
- Cook, J. L., Rapp, J. T., & Schulze, K. A. (2015). Differential negative reinforcement of other behavior to increase wearing of a medical bracelet. *Journal of Applied Behavior Analysis, 48*(4), 901-906. <https://doi.org/10.1002/jaba.228>
- DeLeon, I. G., Hagopian, L. P., Rodriguez-Catter, V., Bowman, L. G., Long, E. S., & Boelter, E. W. (2008). Increasing wearing of prescription glasses in individuals with mental retardation. *Journal of Applied Behavior Analysis, 41*(1), 137-142. <https://doi.org/10.1901/jaba.2008.41-137>

- de Vaan, G., Beijers, B., Vervloed, M. P. J., Knoors, H., Bloeming-Wolbrink, K. A., De Weerth, C., & Verhoeven, L. (2018). *The relationship between stress and autism spectrum disorder in people with sensory and intellectual disabilities: Evidence from cortisol levels*. [thèse de doctorat, Radboud University]. Radboud Repository. <https://repository.ubn.ru.nl/bitstream/handle/2066/191699/191699.pdf?sequence=1#page=166>
- Dickerson, S. S., & Kemeny, M. E. (2004). Acute stressors and cortisol responses: A theoretical integration and synthesis of laboratory research. *Psychological Bulletin*, *130*(3), 355-391. <https://doi.org/10.1037/0033-2909.130.3.355>
- Fombonne, E. (2009). Epidemiology of pervasive developmental disorders. *Pediatric Research*, *65*(6), 591-598. <https://doi.org/10.1203/PDR.0b013e31819e7203>
- Fries, E., Dettenborn, L., & Kirschbaum, C. (2009). The cortisol awakening response (CAR): Facts and future directions. *International Journal of Psychophysiology*, *72*(1), 67-73. <https://doi.org/10.1016/j.ijpsycho.2008.03.014>
- Gabriels, R. L., Agnew, J. A., Pan, Z., Holt, K. D., Reynolds, A., & Laudenslager, M. L. (2013). Elevated repetitive behaviors are associated with lower diurnal salivary cortisol levels in autism spectrum disorder. *Biological Psychology*, *93*(2), 262-268. <https://doi.org/10.1016/j.biopsycho.2013.02.017>
- García-Villamizar, D., & Rojahn, J. (2015). Comorbid psychopathology and stress mediate the relationship between autistic traits and repetitive behaviours in adults with autism: Mediators of repetitive behaviour in autism. *Journal of Intellectual Disability Research*, *59*(2), 116-124. <https://doi.org/10.1111/jir.12083>

- Goodwin, M. S., Intille, S. S., Albinali, F., & Velicer, W. F. (2011). Automated detection of stereotypical motor movements. *Journal of Autism and Developmental Disorders*, *41*(6), 770-782. <https://doi.org/10.1007/s10803-010-1102-z>
- Grant, K. E., Compas, B. E., Stuhlmacher, A. F., Thurm, A. E., McMahon, S. D., & Halpert, J. A. (2003). Stressors and child and adolescent psychopathology: Moving from markers to mechanisms of risk. *Psychological Bulletin*, *129*(3), 447-466. doi:10.1037/0033-2909.129.3.447
- Groden, J., Cautela, J., Prince, S., & Berryman, J. (1994). The impact of stress and anxiety on individuals with autism and developmental disabilities. In E. Schopler & G. B. Mesibov (Eds.), *Behavioral issues in autism* (pp. 177-194). Plenum Press.
- Groden, J., Goodwin, M. S., Baron, M., Groden, G., Velicer, W. F., Lipsitt, L. P., Hofmann, S. G., & Plummer, B. (2005). Assessing cardiovascular responses to stressors in individuals with autism spectrum disorders. *Focus on Autism and Other Developmental Disabilities*, *20*(4), 244-252. <https://doi.org/10.1177/10883576050200040601>
- Gunnar, M. R., & Adam, E. K. (2012). The hypothalamic-pituitary-adrenocortical system and emotion: Current wisdom and future directions. *Monographs of the Society for Research in Child Development*, *77*(2), 109-119. <https://doi.org/10.1111/j.1540-5834.2011.00669.x>
- Heathers, J. A. J., Gilchrist, K. H., Hegarty-Craver, M., Grego, S., & Goodwin, M. S. (2019). An analysis of stereotypical motor movements and cardiovascular coupling in individuals on the autism spectrum. *Biological Psychology*, *142*, 90-99. <https://doi.org/10.1016/j.biopsycho.2019.01.004>

- Hollocks, M. J., Howlin, P., Papadopoulos, A. S., Khondoker, M., & Simonoff, E. (2014). Differences in HPA-axis and heart rate responsiveness to psychosocial stress in children with autism spectrum disorder with and without co-morbid anxiety. *Psychoneuroendocrinology*, *46*, 32-45. <https://doi.org/10.1016/j.psyneuen.2014.04.004>
- Hutt, C., Forrest, S. J., & Richer, J. (1975). Cardiac arrhythmia and behaviour in autistic children. *Acta Psychiatria Scandinavica*, *51*, 361-372. <https://doi.org/10.1111/j.1600-0447.1975.tb00014.x>
- Institut National de Santé Publique (INSPQ; 2017). *Surveillance du trouble du spectre de l'autisme au Québec*. Repéré à : https://www.inspq.qc.ca/sites/default/files/publications/2310_surveillance_trouble_spectre_autisme.pdf
- Jansen, A. S. P., Nguyen, X. V., Karpitskiy, V., Mettenleiter, T. C., & Loewy, A. D. (1995). Central command neurons of the sympathetic nervous system: Basis of the fight-or-flight response. *Science*, *270*, 644-646.
- Ko, B. C. (2018). A brief review of facial emotion recognition based on visual information. *Sensors*, *18*(2), 401. <https://doi.org/10.3390/s18020401>
- Langen, M., Kas, M. J., Staal, W. G., van Engeland, H., & Durston, S. (2011). The neurobiology of repetitive behavior of mice. *Neuroscience & Biobehavioral Reviews*, *35*(3), 345-355. <https://doi.org/10.1016/j.neubiorev.2010.02.004>
- Lanovaz, M. J. (2011). Towards a comprehensive model of stereotypy: Integrating operant and neurobiological interpretations. *Research in Developmental Disabilities*, *32*(2), 447-455. <https://doi.org/10.1016/j.ridd.2010.12.026>

- Lanovaz, M. J., & Sladeczek, I. E. (2012). Vocal stereotypy in individuals with autism spectrum disorders: A review of behavioral interventions. *Behavior Modification*, 36(2), 146-164. <https://doi.org/10.1177/0145445511427192>
- Lazarus, R. S., & Folkman, S. (1984). *Stress, appraisal, and coping*. Springer.
- Lewis, M. H., Tanimura, Y., Lee, L. W., & Bodfish, J. W. (2007). Animal models of restricted repetitive behavior in autism. *Behavioural Brain Research*, 176, 66–74. <https://doi.org/10.1016/j.bbr.2006.08.023>
- Lovaas, I., Newsom, C., & Hickman, C. (1987). Self-stimulatory behavior and perceptual reinforcement. *Journal of Applied Behavior Analysis*, 20(1), 45-68. <https://doi.org/10.1901/jaba.1987.20-45>
- Lydon, S., Healy, O., Mulhern, T., & Hughes, B. M. (2015). Heart rate measurement during stereotyped motor behavior in autism spectrum disorder. *Journal of Developmental and Physical Disabilities*, 27(5), 677-699. <https://doi.org/10.1007/s10882-015-9445-1>
- Lydon, S., Healy, O., Roche, M., Henry, R., Mulhern, T., & Hughes, B. M. (2015). Salivary cortisol levels and challenging behavior in children with autism spectrum disorder. *Research in Autism Spectrum Disorders*, 10, 78-92. <https://doi.org/10.1016/j.rasd.2014.10.020>
- Magnuson, J. R., Peatfield, N. A., Fickling, S. D., Nunes, A. S., Christie, G., Vakorin, V., D'Arcy, R. C. N., Ribary, U., Larocci, G., Moreno, S., & Doesburg, S. M. (2019). Electrophysiology of inhibitory control in the context of emotion processing in children with autism spectrum disorder. *Frontier in Human Neuroscience*, 13, 1-12. <https://doi.org/10.3389/fnhum.2019.00078>

- Mahone, E. M., Bridges, D., Prahme, C., & Singer, H. S. (2004). Repetitive arm and hand movements (complex motor stereotypies) in children. *The Journal of Pediatrics*, *145*(3), 391-395. <https://doi.org/10.1016/j.jpeds.2004.06.014>
- Min, C.-H., & Tewfik, A. H. (2010). Novel pattern detection in children with autism spectrum disorder using iterative subspace identification. *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2266-2269. <https://doi.org/10.1109/ICASSP.2010.5495885>
- Muscatello, R. A., & Corbett, B. A. (2017). Comparing the effects of age, pubertal development, and symptom profile on cortisol rhythm in children and adolescents with autism spectrum disorder. *Autism Research*, *11*(1), 110-120. <https://doi.org/10.1002/aur.1879>
- Muthugovindan, D., & Singer, H. (2009). Motor stereotypy disorders. *Current Opinion in Neurology*, *22*, 131-136. <https://doi.org/10.1097/WCO.0b013e328326f6c8>
- Mudford, O. C., Taylor, S. A., & Martin, N. T. (2009). Continuous recording and interobserver agreement algorithms reported in the Journal of Applied Behavior Analysis (1995-2005). *Journal of Applied Behavior Analysis*, *42*(1), 165-169. <https://doi.org/10.1901/jaba.2009.42-165>
- Ouellette-Kuntz, H., Coo, H., Lam, M., Breitenbach, M. M., Hennessey, P. E., Jackman, P. D., Lewis, M. E. S., Dewey, D., Bernier, F. P., & Chung, A. M. (2014). The changing prevalence of autism in three regions of Canada. *Journal of Autism and Developmental Disorders*, *44*, 120-136. <https://doi.org/10.1007/s10803-013-1856-1>
- Rad, N. M., & Furlanello, C. (2016). Applying deep learning to stereotypical motor movement detection in autism spectrum disorders. *2016 IEEE 16th International Conference on*

Data Mining Workshops (ICDMW), 1235-1242.

<https://doi.org/10.1109/ICDMW.2016.0178>

Rapp, J. T., & Vollmer, T. R. (2005). Stereotypy I: A review of behavioral assessment and treatment. *Research in Developmental Disabilities*, 26(6), 527-547.

<https://doi.org/10.1016/j.ridd.2004.11.005>

Richling, S. M., Rapp, J. T., Carroll, R. A., Smith, J. N., Nystedt, A., & Siewert, B. (2011).

Using noncontingent reinforcement to increase compliance with wearing prescription prostheses. *Journal of Applied Behavior Analysis*, 44(2), 375-379.

<https://doi.org/10.1901/jaba.2011.44-375>

Selye, H. (1974). *Stress without distress*. McClelland and Stewart Limited.

Sentenac, M., Arnaud, C., Gavin, A., Molcho, M., Gabhainn, S. N., & Godeau, E. (2011). Peer victimization among school-aged children with chronic conditions. *Epidemiologic reviews*, 38(1), 1-9.

<https://doi.org/10.1093/epirev/mxr024>

Shonkoff, J. P., Garner, A. S., Siegel, B. S., Dobbins, M. I., Earls, M. F., McGuinn, L.,

Pascoe, J., & Wood, D. L. (2012). The lifelong effects of early childhood adversity and toxic stress. *Pediatrics*, 129(1), 232-246. <https://doi.org/10.1542/peds.2011-2663>

Sokhadze, E., Baruth, J., Tasman, A., Sears, L., Mathai, G., El-Baz, A., & Casanova, M. F.

(2009). Event-related potential study of novelty processing abnormalities in autism.

Applied Psychophysiol Biofeedback, 34(1), 37-51. <https://doi.org/10.1007/s10484-009-9074-5>

Tan, S. Y., & Yip, A. (2018). Hans Selye (1907-1982): Founder of the stress theory. *Medicine in Stamps*, 59(4), 170-171.

<https://doi.org/10.11622/smedj.2018043>

- Taylor, J. L., & Corbett, B. A. (2014). A review of rhythm and responsiveness of cortisol in individuals with autism spectrum disorders. *Psychoneuroendocrinology*, *49*, 207-228. <https://doi.org/10.1016/j.psyneuen.2014.07.015>
- van de Mortel, T. F. (2008). Faking it: Social desirability response bias in self-report research. *Australian Journal of Advanced Nursing*, *25*(4), 40-48.
- Vasa, R. A., Carroll, L. M., Nozzolillo, A. A., Mahajan, R., Mazurek, M. O., Bennett, A. E., Wink, L. K., & Bernal, M. P. (2014). A systematic review of treatments for anxiety in youth with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *44*(12), 3215-3229. <https://doi.org/10.1007/s10803-014-2184-9>
- Viau, R., Arsenault-Lapierre, G., Fecteau, S., Champagne, N., Walker, C.-D., & Lupien, S. (2010). Effect of service dog on salivary cortisol secretion in autistic children. *Psychoneuroendocrinology*, *35*(8), 1187-1193. <https://doi.org/10.1016/j.psyneuen.2010.02.004>
- Volkmar, F. R. (2011). Understanding the social brain in autism. *Developmental Psychobiology*, *53*(5), 428-434. <https://doi.org/10.1002/dev.20556>
- Vollmer, T. R. (1994). The concept of automatic reinforcement: Implications for behavioral research in developmental disabilities. *Research in Developmental Disabilities*, *15*(3), 187-207. [https://doi.org/10.1016/0891-4222\(94\)90011-6](https://doi.org/10.1016/0891-4222(94)90011-6)
- Westeyn, T., Vadas, K., Xuehai B., Starner, T., & Abowd, G. D. (2005). Recognizing mimicked autistic self-stimulatory behaviors using HMMs. *Ninth IEEE International Symposium on Wearable Computers (ISWC'05)*, 164-169. <https://doi.org/10.1109/ISWC.2005.45>

- Willemsen-Swinkels, S. H. N., Buitelaar, J. K., Dekker, M., & van Engeland, H. (1998). Subtyping stereotypic behavior in children: The association between stereotypic behavior, mood, and heart rate. *Journal of Autism and Developmental Disorder*, *28*, 547-557. <https://doi.org/10.1023/A:1026008313284>
- Yang, C.-J., Tan, H.-P., Yang, F.-Y., Wang, H.-P., Liu, C.-L., He, H.-Z., Sang, B., Zhu, X.-M., & Du, Y.-J. (2015). The cortisol, serotonin and oxytocin are associated with repetitive behavior in autism spectrum disorder. *Research in Autism Spectrum Disorders*, *18*, 12-20. <https://doi.org/10.1016/j.rasd.2015.07.002>

Chapitre II : Article 1

Increasing compliance with wearing a medical device in children with autism

Marie-Michèle Dufour^{1,2} et Marc J. Lanovaz^{1,2}

¹École de psychoéducation, Université de Montréal

²Centre de recherche de l'Institut universitaire en santé mentale de Montréal

Cet article a été publié dans *Journal of Applied Behavior Analysis*
Dufour, M.-M., & Lanovaz, M. J. (2020). Increasing compliance with wearing a medical device in children with autism. *Journal of Applied Behavior Analysis*, 53(2), 1089-1096.
<https://doi.org/10.1002/jaba.628>

Résumé

Les professionnels de la santé ont fréquemment recours à l'utilisation de dispositifs médicaux pour évaluer la santé et le bien-être de leurs patients, ainsi que pour améliorer leur qualité de vie. Les sensibilités sensorielles souvent associées au trouble du spectre de l'autisme (TSA) rendent ces individus plus susceptibles de refuser de porter ces différents dispositifs. Pour résoudre ce problème, nous avons systématiquement reproduit des recherches antérieures en examinant les effets du renforcement différentiel d'autres comportements (*differential reinforcement of other behavior*; DRO) pour augmenter la tolérance au port d'une ceinture cardiaque chez deux enfants ayant un TSA. L'intervention a permis d'augmenter la tolérance à 100% chez les deux participants lorsqu'un renforçateur était administré toutes les 90 secondes. Les résultats indiquent que le DRO ne nécessite pas la mise en œuvre de l'extinction pour augmenter la tolérance au port de la ceinture. D'autres recherches sont nécessaires pour évaluer si la densité des renforçateurs peut être diminuée.

Mots-clés: autisme, tolérance, renforcement différentiel d'autres comportements, moniteur cardiaque, intervention

Abstract

Health professionals often recommend the use of medical devices to assess the health, monitor the well-being, or improve the quality of life of their patients. Children with autism may present challenges in these situations as their sensory peculiarities may increase refusals to wear such devices. To address this issue, we systematically replicated prior research by examining the effects of differential reinforcement of other behavior (DRO) to increase compliance with wearing a heart rate monitor in 2 children with autism. The intervention increased compliance to 100% for both participants when an edible reinforcer was delivered every 90 s. The results indicate that DRO does not require the implementation of extinction to increase compliance with wearing a medical device. More research is needed to examine whether the reinforcement schedule can be further thinned.

Keywords: autism, compliance, differential reinforcement, heart rate monitor, intervention

Increasing Compliance with Wearing a Medical Device in Children with Autism

Health professionals may require that their patients wear a medical device (e.g., heart rate monitor, prosthesis, prescription glasses, medical bracelet, electroencephalograph) to assess their health, to monitor their well-being, or to improve their quality of life (Cuvo, 2011). Children with autism spectrum disorders (ASD) who have limited, or no, means of communication may benefit from such technology, as they may be unable to report important medical information (e.g., abnormal internal sensations, feelings of anxiety, allergic reactions). One potentially relevant area of application is for the measurement and monitoring of stress. Researchers have suggested that health professionals may assess the stress levels of children with ASD who have no functional form of communication, by monitoring heart rate in combination with other measures (Gabriels et al., 2013; Hollocks et al., 2014). Notably, these devices (e.g., heart rate monitors, electrodermal sensors) can detect and report increases in levels of stress and anxiety (Fletcher et al., 2010; Poe et al., 2010), and may be used to suggest interventions to prevent these manifestations (Liu et al., 2008).

When an individual has limited means of communication, using heart rate monitors allows for the collection of data on levels of stress that could not have been recorded with other traditional methods (questionnaires, interviews). In addition to monitoring physical health, cardiac measurements have the potential for increasing the understanding of challenging behavior, and eventually guiding the selection of interventions. As stress and anxiety may function as motivating operations (MOs) for challenging behavior, behavior analysts should consider their behavioral and physiological manifestations as problems of social significance (Friman et al., 1998). If additional research is conducted to support this approach, practitioners and researchers may eventually use heart rate to monitor the presence

or absence of physiological MOs. That said, children with ASD may have increased rates of refusal to wear such devices (Johnson & Rodriguez, 2013).

Researchers have shown that behavioral interventions are effective at increasing compliance with wearing medical devices in children with developmental disabilities (Cook et al., 2015; DeLeon et al., 2008; Richling et al., 2011). For example, DeLeon et al. (2008) implemented noncontingent reinforcement (NCR) to increase the wearing of prescription glasses in four participants with intellectual disabilities. Their results indicated that NCR alone increased compliance with one participant, whereas the addition of response blocking and response cost were necessary for the remaining participants. In a replication and extension, Richling et al. (2011) found that NCR alone increased compliance with wearing foot orthopedics and hearing aids in two children with developmental disabilities. More recently, Cook et al. (2015) reported that differential negative reinforcement of other behavior (DNRO) increased compliance with wearing a medical bracelet in one child with autism. Their DNRO procedure consisted of removing the device and praising the child when he was compliant during a predetermined interval. At any other times, the trainer blocked attempts at removing the bracelet.

Both DeLeon et al. (2008) and Cook et al. (2015) implemented response blocking as part of their intervention, which can be a limitation for some individuals. The use of response blocking to implement escape extinction may not be possible or desirable in certain settings or with specific individuals (Athens & Vollmer, 2010). For example, response blocking may lead to aggression (e.g., Hagopian & Toole, 2009), which may prevent its implementation. When NCR alone is ineffective (e.g., DeLeon et al., 2008) and response blocking is not an option,

practitioners and researchers may require alternatives to increase compliance with wearing medical devices.

Differential reinforcement of other behavior (DRO) is an alternative intervention that does not require response blocking. That is, the DRO procedure may prevent the occurrence of challenging behavior that are evoked by the response blocking procedure implemented in DNRO. Researchers have used DRO alone and as part of treatment packages, to increase compliance with medical procedures (Carton & Schweitzer, 1996; Cuvo, Godard et al., 2010; Cuvo, Reagan et al., 2010; Goetz et al., 1975; Shabani & Fisher, 2006), but we found no study that focused solely on wearing a medical device. As children with ASD may need to wear medical devices for assessment and treatment of some co-occurring conditions, examining the effects on an intervention that does not require response blocking appears important. The current study is a systematic replication of prior research that implemented interventions to increase compliance with wearing different medical devices (Cook et al., 2015; Richling et al., 2011). Even though several studies demonstrated the effectiveness of behavioral procedures in increasing compliance, we aimed to extend this knowledge by assessing whether a DRO schedule was effective in the absence of an extinction component. Therefore, the purpose of the study was to assess the effects of DRO on compliance with wearing a heart rate monitor.

Method

Participants and Settings

Two children, diagnosed with ASD by an independent multidisciplinary team, participated in the study. The two participants were part of a larger study that required them to wear a heart rate monitor for 30-min periods. More specifically, the university's research ethics board approved the larger study that aimed to identify behavioral correlates of

physiological stress. To participate in the larger study, the children had to: (a) already have a diagnosis of ASD, and (b) have a score of 3.5 or more on the verbal communication subscale of the Childhood Autism Rating Scale (CARS-2; Schopler et al., 2010). This score indicates that speech may be absent or, if present, verbal communication takes the form of peculiar language (jargon or echolalia; Schopler et al., 2010). We included the latter criterion as the larger study focused on children with limited or no functional means of communication. As the larger study involved the measurement of heart rate as a main dependent variable, all children who refused to wear the device (i.e., pull the heart rate monitor off their chest when clipped it on) participated in the present study.

Leo (fictional name) was a 5-year-old boy with severe symptoms of ASD (according to the CARS-2) who did not use functional mands or tacts to communicate. Adam (fictional name) was a 9-year-old boy who also presented severe symptoms of ASD and had no functional means of communication. To our knowledge, neither of the participants used an augmentative and alternative communication system. All sessions took place in each participant's home. Leo's sessions occurred in his basement playroom and Adam's sessions occurred in his family's living room. During the sessions, participants had access to toys they typically interacted with at home. Both participants had access to blocks, three to four books, and a tablet. These items remained consistent across sessions and no other toys were available.

Data Collection and Interobserver Agreement (IOA)

The trainer (i.e., the first author) recorded each session on video and subsequently measured the duration that each participant complied with wearing the heart rate monitor and the frequency of device removal. *Compliance* with wearing the device was defined as the heart rate monitor being in direct contact with the participant's chest. The participant could touch

the device and still be in compliance, but a *noncompliance* occurred as soon as he pulled on the heart rate monitor so that it was no longer in contact with his chest. *Device removal* was defined as the trainer unclipping the heart rate monitor and taking it off the participant contingent on noncompliance. The trainer used continuous recording to measure the duration of compliance with wearing the device as well as the frequency of device removal. For the analyses, she converted the duration measure into a percentage of time by dividing the duration of compliance by the duration of the session (300 s) and multiplying the quotient by 100. A second observer scored 33% of sessions for each participant. The experimenters calculated IOA by using the block-by-block method with 10-s intervals (Mudford et al., 2009). The mean IOA scores for compliance and device removal, respectively, were 94% (range, 89% to 97%) and 99% (range, 98% to 100%) for Leo, and 94% (range, 83% to 98%) and 100% for Adam.

Experimental Design and Procedures

The experimenters used a nonconcurrent multiple baseline across participants to evaluate the effects of DRO intervention on compliance with wearing the heart rate monitor and on device removals. Each child first participated in a preference assessment to identify the reinforcer delivered during DRO. Then, the experimenters monitored compliance and device removals across baseline and DRO sessions. The trainer typically conducted 1 to 4 sessions per day, once or twice per week with each participant.

Preference assessment

Prior to baseline, the trainer evaluated each participant's preference for five edible items (M&Ms™, Skittles™, Gummy bears™, nachos, potato chips). using a paired-choice assessment procedure (Fisher et al., 1992). The parents proposed some of the five edible items,

but the trainer made the final selection so that it remained consistent across participants. All possible pairs were presented in random order with position of placement counterbalanced. The experimenters chose arbitrary rather than functional reinforcers (i.e., escape) because edibles could be consumed in a shorter amount of time and it prevented the use of response blocking. Leo and Adam preferred the potato chips and the nachos, respectively.

Baseline

Each baseline session lasted 5 min and began when the trainer fitted the participant with the heart rate monitor (a Polar H7 Bluetooth Heart Rate Sensor & Fitness Tracker). The therapist started the 5-min timer simultaneously. Although the participant ultimately had to wear the device for 30 min for the larger study, the experimenters targeted 5 min for the current experiment as it was easier to conduct multiple short sessions with the participants to start. Fitting the device involved clipping the strap below the chest muscle while ensuring that the sensor was in contact with the skin. Like the escape condition described in the recent study by Cook et al. (2015), the trainer removed the heart rate monitor for 30 s each time the participant pulled on it so that it was no longer in contact with the skin of his chest. At the end of the 30 s, the trainer refitted the participant with the monitor. The timed 5-min session continued to run whether the participant pulled on the sensor or not. The session ended when 5 min had elapsed

Differential reinforcement of other behavior

The DRO condition was similar to the baseline condition: the session lasted 5 min, the 30-s escape contingency remained in effect and the participant had access to the same toys. However, the trainer simultaneously implemented the DRO intervention. During DRO, the participant received his preferred edible reinforcer if the heart rate monitor stayed in contact

with his chest skin for the entire duration of the interval. If the participant met the criterion, the trainer said, “Congratulations”, provided a small piece of edible reinforcer, left the sensor on the participant’s chest, and restarted the timer. As soon as the participant pulled on the device so that it was no longer in contact with the chest, the trainer removed the sensor for 30 s without saying anything. Following 30 s, the trainer fitted the sensor to the participant and reset the DRO interval. Unlike Cook et al. (2015), the trainer did not provide access to escape based on compliance, nor did the trainer block any attempt to remove the device. Escape remained contingent on noncompliance (i.e., pulling the monitor off his chest).

Whenever the participant achieved the criterion to receive his reinforcer for three consecutive intervals, the trainer increased the duration of the schedule. Contrarily, the participant returned to the previous denser schedule when he failed to achieve the criterion for three consecutive intervals. The duration of the DRO schedule could increase or decrease within a session. The trainer implemented within-session changes in the DRO schedule as it allowed more rapid increases in the interval schedule duration, which could in turn reduce the time required to meet the terminal criterion. The DRO schedules were 5, 10, 15, 20, 30, 40, 45, 60 and 90 s. The experimenters chose 90 s as the terminal DRO schedule because it was deemed realistic to provide the reinforcer on a 90-s schedule within their larger study (which required that the child wear the heart rate monitor for 30 min).

When 5 min elapsed, the session ended regardless of whether the participant was in the middle of an interval. The starting interval criterion of each session was determined based on the last interval applied in the previous session. For example, if the participant met the criterion at 15 s for three consecutive intervals towards the end of the second session, the starting interval criterion of the third session would be 20 s. The participants reached the

termination criterion when they wore the heart rate monitor 100% of the time for three consecutive sessions at the 90-s schedule.

Results

Figure 1 presents the results for Leo and Adam. During the baseline phase, Leo (upper panel) wore the heart rate device between 0% and 20% of the sessions ($M = 6\%$) and the trainer removed the device a mean of 10 times per session. The duration of compliance increased rapidly after the implementation of the DRO sessions, leading Leo to comply with wearing the device between 43% and 100% of the sessions ($M = 79\%$). The frequency of device removal decreased consequently to a mean of 1.8 per DRO session. It took 13 sessions (i.e., 65 min of training) for Leo to comply with wearing the device 100% of the time with the 90-s intervals. Adam wore the heart rate monitor between 3% and 29% ($M = 11\%$) of the baseline sessions and the trainer removed the device a mean of 7 times per session during this phase. During the DRO sessions, he wore the device between 28% and 100% ($M = 79\%$) of the sessions and we removed the device approximately 1.5 times per session. Adam required 27 sessions to reach the termination criterion (i.e., 125 min of training).

Discussion

The results indicate that DRO, without an extinction component, increased compliance with wearing the heart rate monitor for both participants. The participants only required 65 to 125 min of training to reach the termination criterion (i.e., wearing the device for 100% of the session while on a 90-s DRO schedule). Most importantly, the intervention did not require the implementation of response blocking or escape extinction. Conceptually, wearing a medical device does not involve engagement in a specific behavior (i.e., no movement or control of muscles), which is why DRO is well suited to explain the changes that we observed in the

current study. The delivery of the preferred edible reinforced engagement in behavior other than attempting to remove the heart rate monitor. The increase in compliance in the absence of an extinction component also suggests that the preferred edibles were reinforcers that could effectively compete with the escape contingency.

This study extends previous research regarding the relevance of behavioral procedures in training children with developmental disabilities to comply with wearing medical devices (e.g., Cook et al., 2015; DeLeon et al., 2008; Richling et al., 2011). These findings also remain consistent with prior research on the effectiveness of positive reinforcement without an extinction component in increasing compliance (Lalli et al., 1999; Piazza et al., 1997). From a practical standpoint, the results of this study contribute to the research literature by showing that DRO may be an alternative to DNRO and NCR for increasing compliance with wearing a medical device. Practitioners may consider DRO as an intervention when implementing response blocking and escape extinction is unfeasible or counterproductive, or when NCR alone is ineffective. Considering that no response blocking was needed, DRO may be a useful procedure for the practitioner to use in a variety of contexts, particularly when blocking evokes challenging behavior.

In the current study, the trainer delivered an arbitrary reinforcer, as providing escape in DNRO would have required the implementation of response blocking. Response blocking may evoke other challenging behaviors (Hagopian & Toole, 2009), which the experimenters wanted to avoid. That being said, prior research has shown that the quality of the reinforcers may influence the response to an intervention (Lalli et al., 1999). Thereby, the choice of arbitrary reinforcers may have influenced the effectiveness of the DRO. To address this issue,

researchers should compare a function-based intervention (DNRO) with a non-function-based intervention (DRO) in the future.

The main limitation of the study is that the intervention ended when the participant met the termination criterion with a 90-s DRO schedule. However, health professionals may require that some devices (e.g., hearing aids, prescription glasses) be worn all day in which case the schedule would need to be considerably thinner. Future research should address this issue as a priority by further increasing the duration of the DRO schedule. Moreover, the experimenters did not conduct a functional analysis prior to implementing DRO as this type of intervention should be effective regardless of function. Performance during baseline indicated that compliance was low during the escape contingency, but the analysis does not rule out the possibility that the behavior was multiply controlled. Future research should address this limitation by conducting a functional analysis beforehand. Another limitation is that the small sample size and nonconcurrent nature of the design could have affected the internal validity. In sum, researchers should replicate these procedures with a larger sample, with other types of devices and with a more rigorous design, while further thinning the reinforcement schedule to extend the generality of the findings.

References

- Athens, E. S., & Vollmer, T. R. (2010). An investigation of differential reinforcement of alternative behavior without extinction. *Journal of Applied Behavior Analysis, 43*(4), 569-589. <https://doi.org/10.1901/jaba.2010.43-569>
- Carton, J. S., & Schweitzer, J. B. (1996). Use of a token economy to increase compliance during hemodialysis. *Journal of Applied Behavior Analysis, 29*(1), 111-113. <https://doi.org/10.1901/jaba.1996.29-111>
- Cook, J. L., Rapp, J. T., & Schulze, K. A. (2015). Differential negative reinforcement of other behavior to increase wearing of a medical bracelet. *Journal of Applied Behavior Analysis, 48*(4), 901-906. <https://doi.org/10.1002/jaba.228>
- Cuvo, A. J. (2011). Training children with autism and pervasive developmental disorders to comply with healthcare procedures: Theory and research. In J. L. Matson (Ed.), *International handbook of autism and pervasive disorders* (pp. 381-395). Springer.
- Cuvo, A. J., Godard, A., Huckfeldt, R., & DeMattei, R. (2010). Training children with autism spectrum disorders to be compliant with an oral assessment. *Research in Autism Spectrum Disorders, 4*(4), 681-696. <https://doi.org/10.1016/j.rasd.2010.01.007>
- Cuvo, A. J., Reagan, A. L., Ackerlund, J., Huckfeldt, R., & Kelly, C. (2010). Training children with autism spectrum disorders to be compliant with a physical exam. *Research in Autism Spectrum Disorders, 4*(2), 168-185. <https://doi.org/10.1016/j.rasd.2009.09.001>
- DeLeon, I. G., Hagopian, L. P., Rodriguez-Catter, V., Bowman, L. G., Long, E. S., & Boelter, E. W. (2008). Increasing wearing of prescription glasses in individuals with mental retardation. *Journal of Applied Behavior Analysis, 41*(1), 137-142. <https://doi.org/10.1901/jaba.2008.41-137>

- Fisher, W. W., Piazza, C. C., Bowman, L. G., Hagopian, L. P., Owens, J. C., & Slevin, I. (1992). A comparison of two approaches for identifying reinforcers for persons with severe and profound disabilities. *Journal of Applied Behavior Analysis, 25*(2), 491-498. <https://doi.org/10.1901/jaba.1992.25-491>
- Fletcher, R. R., Dobson, K., Goodwin, M. S., Eydgahi, H., Wilder-Smith, O., Fernholz, D., Kuboyama, Y., Hedman, E. B., Poh, M.-Z., & Picard, R. W. (2010). iCalm: Wearable sensor and network architecture for wirelessly communicating and logging autonomic activity. *IEEE Transactions on Information Technology in Biomedicine, 14*, 215-223. <https://doi.org/10.1109/TITB.2009.2038692>
- Friman, P. C., Hayes, S. C., & Wilson, K. G. (1998). Why behavior analysts should study emotion: The example of anxiety. *Journal of Applied Behavior Analysis, 31*(1), 137-156. <https://doi.org/10.1901/jaba.1998.31-137>
- Gabriels, R. L., Agnew, J. A., Pan, Z., Holt, K. D., Reynolds, A., & Laudenslager, M. L. (2013). Elevated repetitive behaviors are associated with lower diurnal salivary cortisol levels in autism spectrum disorder. *Biological Psychology, 93*(2), 262-268. <https://doi.org/10.1016/j.biopsycho.2013.02.017>
- Goetz, E. M., Holmberg, M. C., & LeBlanc, J. M. (1975). Differential reinforcement of other behavior and noncontingent reinforcer as control procedures during the modification of a preschooler's compliance. *Journal of Applied Behavior Analysis, 8*(1), 77-82. <https://doi.org/10.1901/jaba.1975.8-77>
- Hagopian, L. P., & Toole, L. M. (2009). Effects of response blocking and competing stimuli on stereotypic behavior. *Behavior Modification, 24*(2), 117-125. <https://doi.org/10.1002/bin.278>

- Hollocks, M. J., Howlin, P., Papadopoulos, A. S., Khondoker, M., & Simonoff, E. (2014). Differences in HPA-axis and heart rate responsiveness to psychosocial stress in children with autism spectrum disorders with and without co-morbid anxiety. *Psychoneuroendocrinology*, *46*, 32-45. <https://doi.org/10.1016/j.psyneuen.2014.04.004>
- Johnson, N. L., & Rodriguez, D. (2013). Children with autism spectrum disorder at a pediatric hospital: A systematic review of the literature. *Pediatric Nursing*, *39*, 131-141.
- Lalli, J. S., Vollmer, T. R., Progar, P. R., Wright, C., Borrero, J., Daniel, D., Barthold, C. F., Tocco, K., & May, W. (1999). Competition between positive and negative reinforcement in the treatment of escape behavior. *Journal of Applied Behavior Analysis*, *32*(3), 285-296. <https://doi.org/10.1901/jaba.1999.32-285>
- Liu, C., Conn, K., Sarkar, N., & Stone, W. (2008). Physiology-based affect recognition for computer-assisted intervention of children with Autism Spectrum Disorder. *International Journal of Human-Computer Studies*, *66*(9), 662-677. <https://doi.org/10.1016/j.ijhsc.2008.04.003>
- Mudford, O. C., Taylor, S. A., & Martin, N. T. (2009). Continuous recording and interobserver agreement algorithms reported in the Journal of Applied Behavior Analysis (1995-2005). *Journal of Applied Behavior Analysis*, *42*(1), 165-169. <https://doi.org/10.1901/jaba.2009.42-165>
- Piazza, C. C., Fisher, W. W., Hanley, G. P., Remick, M. L., Contrucci, S. A., & Aitken, T. L. (1997). The use of positive and negative reinforcement in treatment of escape-maintained destructive behavior. *Journal of Applied Behavior Analysis*, *30*(2), 279-298. <https://doi.org/10.1901/jaba.1997.30-279>
- Poe, M. Z., Swenson, N. C., & Picard, R. W. (2010). A wearable sensor for unobstrusive,

long-term assessment of electrodermal activity. *IEEE Transactions on Information Technology in Biomedicine*, 57, 1243-1252.

<https://doi.org/10.1109/TBME.2009.2038487>

Richling, S. M., Rapp, J. T., Carroll, R. A., Smith, J. N., Nystedt, A., & Siewert, B. (2011).

Using noncontingent reinforcement to increase compliance with wearing prescription prostheses. *Journal of Applied Behavior Analysis*, 44(2), 375-379.

<https://doi.org/10.1901/jaba.2011.44-375>

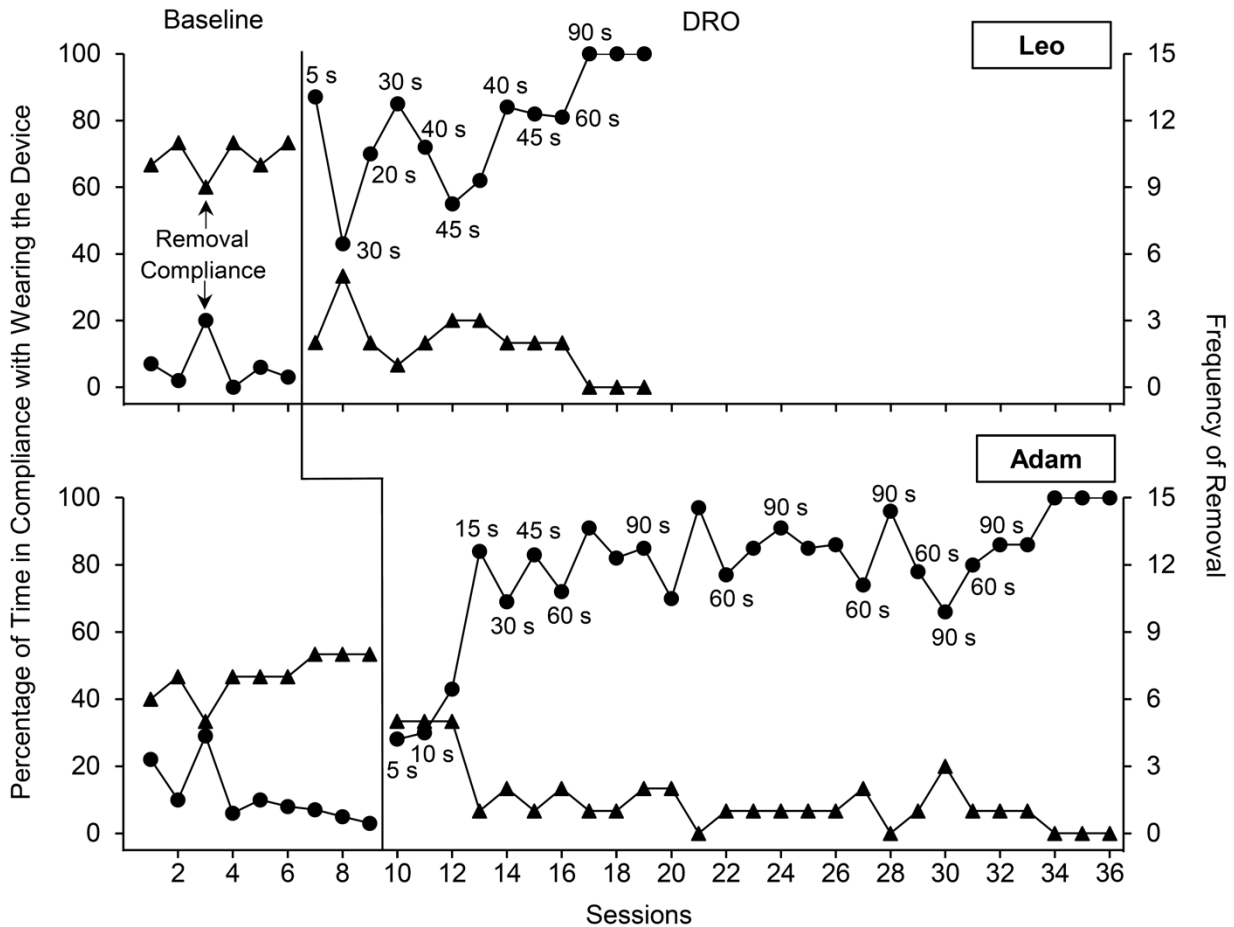
Schopler, E., Van Bourgondien, M. E., Wellman, G. J., & Love, S. R. (2010). *The Childhood Autism Rating Scale*. (2nd ed.). Western Psychological Services.

Shabani, D. B. & Fisher, W. W. (2006). Stimulus fading and differential reinforcement for the treatment of needle phobia in a youth with autism. *Journal of Applied Behavior*

Analysis, 39(4), 449-451. <https://doi.org/10.1901/jaba.2006.30-05>

Figure 1

Percentage of Time Complying With Wearing the Heart Rate Monitor (Primary y-Axis) and Frequency of Device Removal (Secondary y-Axis) for Leo (Upper Panel) and Adam (Lower Panel) During Baseline and Differential Reinforcement of Other Behavior (DRO) Sessions.



The Values Above or Below the Data Points Identify the Duration of the DRO Schedules at the Start of Sessions. The Absence of Value Indicates That the Starting Schedule was the Same as in the Previous Session.

Chapitre III : Article 2

**Salivary cortisol, heart rate, and direct observation measures of stereotypy
in minimally verbal children with autism**

Marie-Michèle Dufour^{1,2}, Marc J. Lanovaz^{1,2} et Pierrich Plusquellec^{1,2}

¹École de psychoéducation, Université de Montréal

²Centre de recherche de l'Institut universitaire en santé mentale de Montréal

Une adaptation de cet article sera soumise prochainement dans une revue spécialisée en autisme

Résumé

Plusieurs études ont fait état de résultats contradictoires lors de l'évaluation de l'association entre le stress et les comportements stéréotypés. L'objectif de notre étude était d'examiner la relation entre le cortisol salivaire, les mesures de fréquence cardiaque et des mesures d'observations directes des comportements stéréotypés chez des enfants minimalement verbaux présentant un trouble du spectre de l'autisme (TSA). À cette fin, nous avons combiné un devis par alternance de traitements à des analyses de régression multiple pour évaluer l'association entre ces variables chez quatre enfants ayant un TSA. Nos résultats ont montré que la valeur moyenne du cortisol était négativement associée à la stéréotypie globale et motrice, tandis que la fréquence cardiaque était positivement liée à la stéréotypie globale et motrice. Aucune relation significative n'a été trouvée entre la valeur moyenne du cortisol, la fréquence cardiaque et la stéréotypie vocale. Nos résultats mettent en évidence la relation complexe entre le stress et la stéréotypie et mettent en lumière la nécessité de poursuivre les recherches sur ce sujet.

Mots-clés : autisme, fréquence cardiaque, cortisol salivaire, stéréotypie, stress

Abstract

Several studies have reported conflicting results when assessing associations between stress and repetitive behaviors. The purpose of our study was to examine the relationship between salivary cortisol, heart rate measures, and direct observation measure of stereotypy in minimally verbal children with autism spectrum disorder (ASD). To this end, we combined alternating-treatment designs and multiple regression analyses to evaluate the association between these variables in four children with ASD. Our results showed that the mean value of cortisol was negatively associated with global and motor stereotypy, whereas heart rate (HR) was positively linked to global and motor stereotypy. No significant relation was found between mean value of cortisol, HR and vocal stereotypy. Our results highlight the complex relationship between stress and stereotypy and the need for further research.

Keywords: autism, heart rate, salivary cortisol, stereotypy, stress

Salivary Cortisol, Heart Rate, and Direct Observation Measures of Stereotypy in Minimally Verbal Children with Autism

Repetitive behaviors represent a core features of autism spectrum disorder (ASD; American Psychiatric Association, 2013). Stereotypy is one form of repetitive behaviors that is often classified in two broad categories: vocal and motor stereotypy. Vocal stereotypy is defined as repetitive vocalizations which persists in the absence of social consequences (Ahearn et al., 2007; Lanovaz & Sladeczek, 2011) whereas motor stereotypy is defined by patterned, repetitive, coordinated, rhythmic and non-reflexive behaviors (Mahone et al., 2004). According to reviews of the literature, at least 50% of individuals with ASD show at least one form of motor stereotypy (Melo et al., 2020). This proportion approaches 90% when vocal forms of stereotypy are included (Chebli et al., 2016).

Researchers have proposed three main hypotheses to explain the maintenance of stereotypy in ASD: (1) the self-stimulatory hypothesis, (2) the manifestation of stress hypothesis and (3) the self-regulation hypothesis. The self-stimulatory hypothesis considers that stereotypy is maintained by automatic positive reinforcement (Lovaas et al., 1987; Rapp & Vollmer, 2005). That is, stereotypy generates forms of stimulation that function as positive reinforcement for the behavior. Hence, individuals with ASD who engage in stereotypy access this reinforcement. Supporting this hypothesis, Lydon et al. (2013) examined the heart rate (HR) of three children with ASD around stereotypy and found that stereotypy is related to an increase in HR for all participants. They measured behavior using direct observation and they rated the severity of the behavior using a three-point scale ranging from (1) mild to (3) severe. The authors reported that stereotypy could be positively reinforced in allowing the individual to access a state of high physiological arousal (Lydon et al., 2013). The large body of research

on functional analysis showed that stereotypy tends to occur most when individuals with ASD have the least environmental stimulation and demands (Querim et al., 2013).

The proponents of stress hypothesis argue that stereotypy is a manifestation of a physical state including stress and excitement (Muthugovindan & Singer, 2009). Several studies found positive correlation between cortisol basal levels and stereotypy, as measured by questionnaires, in individuals with ASD (Bitsika et al., 2015; García-Villamizar & Rojahn, 2015; Lydon, Healy, Roche et al., 2015; Yang et al., 2015). Specifically, individuals with higher levels of cortisol tend to show higher levels of stereotypy. These authors hypothesized that repetitive behavior could be a manifestation of the internal state of the individual such as stress. In contrast, Bitsika et al., (2015) found a positive association between cortisol basal levels and stereotypy in children, but not in adolescents, with ASD. Although these studies associate cortisol with stereotypy, the authors mentioned that the underlying exact mechanism remains unknown.

Groden et al. (1994) created an adaptive and maladaptive theoretical model that contrasts people who adopts appropriate coping strategies with people who do not. In their model, the authors interpreted stereotypy as a coping strategy within a population that does not have appropriate coping mechanisms. At least three studies have supported this hypothesis (Heathers et al., 2019; Hutt et al., 1975; Gabriels et al., 2013). First, Hutt et al. (1975) noted that HR of nine ASD children decreased following the occurrence of stereotypy. Second, Gabriels et al. (2013) found that children who present a higher level of stereotypy have lower levels of cortisol. After controlling for age, intelligence quotient (IQ), and the type of repetitive behaviors, the authors found that the children who were in the high repetitive behaviors group showed 36% lower diurnal salivary cortisol in comparison with those who

were in the low repetitive behavior group. The authors of both studies suggested that repetitive behavior could play a self-regulation function in this population. A more recent study evaluated the synchronicity between cardiovascular measures and direct measures of motor stereotypy (hand flapping and body rocking). The authors found a similar cardiovascular pattern of acceleration and deceleration unrelated to physical demands across the 10 participants that supported the regulatory function hypotheses (Heathers et al., 2019).

Lydon, Healy, Mulhern, & Hughes (2015) found no significant link between HR and stereotypy in five children with ASD. Faced with these inconsistent results and in agreement with previous authors (Hutt et al., 1975; Sroufe et al., 1973), the authors stipulated that stereotypy function could be multiple, including elevation of physiological arousal during a period of low environmental stimulation (i.e., self-stimulatory function), and releasing excitement or preventing overload of stimulation in periods of high physiological arousal (i.e., self-regulatory function). Another hypothesis that help support these contradictory results across studies is that stereotypy may serve different functions depending on the internal state of an individual and the surrounding environmental conditions (Willemsen-Swinkels et al., 1998).

As indicated in the previous paragraph, the vast majority of past studies measured repetitive behaviors with questionnaires (Bitsika et al., 2015; Gabriels et al., 2013; Lydon, Healy, Roche et al., 2015; Yang et al., 2015). The most frequently used questionnaire for measuring repetitive behaviors is the Repetitive Behavior Scale – Revised (RBS-R; Bodfish et al., 1999, 2000), but the Aberrant Behavior Checklist (ABC; Aman et al., 1985) and the Behavior Problems Inventory (BPI-01; Rojahn et al., 2001) were also used in previous research. Being reported by caregivers, these measures are somewhat biased or at least not as

precise as direct observation measures (van de Mortel, 2008). To our knowledge, no study has evaluated the relation between cortisol and stereotypy using observational measures.

Moreover, no study combined multiple biological measures (HR and cortisol) and direct observation measures. The inconsistency in the results of past studies highlight the need for further research and identify the need to include direct observation measures of stereotypy. Finally, no study evaluated stress (cortisol and HR) and stereotypy strictly in a minimally verbal subpopulation of ASD. Our hypothesis is that heterogeneity in samples of the past studies combined with the low proportion of studies using direct measurements of stereotypy could potentially explain the divergent results. Therefore, the purpose of our study is to evaluate the relation between salivary cortisol, heart rate measures, and direct observation measures of stereotypy in minimally verbal children with ASD.

Method

Participant and Settings

Four children between the ages of 5 and 9 years old participated in this study. To be included in the study, the children had to: (a) already have a diagnosis of ASD provided by an independent multidisciplinary team, (b) present at least one form of stereotypy, and (c) have a score of 3.5 or more on the verbal communication subscale of the Childhood Autism Rating Scale (CARS-2; Schopler et al., 2010). This score indicates that speech was absent or, if present, took the form of peculiar language (jargon or echolalia; Schopler et al., 2010). All participants presented severe symptoms of ASD according to the CARS-2 (T-score > 49). Ana and Oli (fictitious names) did not have means of communication other than informal sounds. Matt and Jim (fictitious names) pronounced nonfunctional words (i.e. echolalia). No participant was taking any psychotropic medication that could have impacted neuroendocrine

functioning. Table 1 presents participant characteristics. Our university research ethics board approved the present study and we obtained consent from the legal guardian of each participant. All sessions took place in each participant's home.

Measures

The Childhood Autism Rating Scale (CARS-II)

The CARS-II (Schopler et al., 2010) is a 15-item observation-based rating system that provided the severity of the autistic symptomology. The CARS-II is also used in the diagnostic process. For children, a score of 30 represents the cut-off for an ASD diagnosis. Psychometric properties of this tool have been evaluated and found to be adequate (Parkhurst & Kawa, 2018). This measure was not used for diagnostic purposes (which was conducted by an independent multidisciplinary team), but only to characterize our participants and determine eligibility (i.e., minimally verbal status).

Saliva Collection

The first author and research assistants collected sixty samples of saliva from each participant for a total of 240 saliva samples. Four saliva samples were taken on each day that an observation took place (before, during and after a session of observation). To prevent teaching undesirable behaviors to the participants (e.g., spitting) and to facilitate data collection, we used the same procedure as described by Dozier et al. (2006). The saliva samples were taken with Salivabio children's swab®. The data collector briefly introduced the swab into the child's mouth in order to slightly humidify. She then dipped the swab in a little bag containing 0.8g of flavored beverage crystal (Tang orange flavor drink mix®). The data collector put back the swab in the child's mouth until it was wet with saliva. She puts the wet swab in the designated storage tube and put in the freezer for preservation prior to all samples

being sent out to the lab. A study reported no significant differences between samples collected with and without low levels (0.8g) of drink mix (Gordon et al., 2005). All the saliva collection took place in the afternoon, which represents a reliable procedure that reduces inter-participant variability (Sharpley et al., 2016).

The lab used a high sensitivity enzyme immune assay kit from Salimetrics State College, PA to analyze the salivary cortisol concentrations. The samples were brought to room temperature to be centrifuged at 15000xg (3000rpm) for 15 min. The sensitivity of this assay ranged from 0.012 - 3 $\mu\text{g}/\text{dl}$. The lowest limit of detection is 0.007 $\mu\text{g}/\text{dl}$. All samples with a CV% >15 were rerun, unless the difference between the two values was below .03 $\mu\text{g}/\text{dl}$. To examine the cortisol level, we used the mean value level of cortisol for each session. This decision was based on two arguments: (1) in using the mean value, we eliminated the impact of the missing value of our data and (2) according to the 5% trimmed mean test ran with SPSS, outliers had no effect on this indicator whereas they do on areas under the curve with respect to ground (AUC_g).

Heart rate Variability

During each session, the participants were wearing a Polar® heart rate thoracic belt monitor (PolarH7®). The mobile app EliteHRV® recorded the signals and created a text file that contained a series of cardiac intervals from which a custom Java application extracted a mean heart rate per min.

Direct Observation Measures

The first author, a doctoral student, videotaped each session, and manually coded each video recording second-by-second with the support of two research assistants. Table 2 presents the behavioral definitions for each form of stereotypy emitted by the participants. If

stereotypy was present at any time during a second, the observer coded the behavior as occurring during the second. When coding, the observer also made the distinction between vocal and motor stereotypy. We also created a global stereotypy variable which was the total percentage of time that the participant emitted stereotypy, vocal and motor combined. For the analysis, a percentage was calculated by dividing duration in which the participant was engaging in stereotypy by the total duration of the session and multiplying the quotient by 100. A second observer measured stereotypy on 33% of the recordings. The mean second-by-second interobserver agreement was .96 (range: .93-.98) and the mean kappa interobserver agreement was .86 (range: .68-.96).

Procedures

Each child participated in 15, 30-min observation periods. For each session, four cortisol measures were taken: 10 min prior to the beginning of the session, 15 min after it started, at the end of the session and 20 min after it ended. The participants wore the heart rate belt for the entire duration of the session. To make the level of stress vary during our observations, we designed three conditions inspired from the stress ingredients of Dickerson and Kemeny (2004): control, unpredictability, and novelty. Each child participated in five sessions of each condition for a total of 15 sessions per participant. The conditions varied randomly to avoid order effects. We used an alternating-treatment design to compare levels of stereotypy across three conditions. In the week preceding the first session, the trainer visited each participant's home twice for a 30-min play period. The purpose of these visits was for the child to habituate to the data collector's presence and to the heart rate belt. During these visits, the data collector also observed the child and filled the CARS-II.

Control

In this condition, the participants had access to preferred items, received non-contingent attention from the data collector, and no demands were made. The trainer stayed with the child, participated in the play, and ignored all forms of stereotypy.

Unpredictability

In the first 5 min of this condition, the participants had access to their preferred item. Then, the trainer removed that item and offered a less appreciated toy for the next 10 min. After that, the participants were asked to perform two different tasks for the next 10 min (2 x 5 min). For the last 5 min, the data collector offered another non-preferred item to the child. The order of the task and the less appreciated item presentation varied during each session in order to keep the unpredictability variable.

Novelty

In this condition, participants were allowed to interact with their preferred item for the first 5 min. Then, the trainer removed that item and offered a new toy for the next 10 min. After that, the participants were asked to perform a new task (10 min). Finally, another new toy was presented for the last 5 min of the session. All toys or tasks presented in this condition were items with which the child had never interacted before.

Analysis

Cortisol Analysis

To examine patterns across conditions, the first author graphed stereotypy and cortisol levels on alternating-treatment graphs. Due to missing data, we didn't generate any graphs with cardiac data. We applied a rule developed by Lanovaz et al. (2019) to identify any significant differences between conditions. Prior to the main analysis, we performed an overall

inspection of our data as well as an assessment of the normality of our variables. Inspection of the data revealed some missing data (0% stereotypy; 10% HR; 1,2% cortisol). Since we decided to use the mean indicator of cortisol, these missing data had no or few effects on our analyses. The three cortisol missing samples were due to the collection of an insufficient volume of saliva for the analysis. HR missing values were due to an app technical problem in saving the data or a Bluetooth connection problem between the belt and the app. Our inspection revealed that cortisol data was positively skewed, as it is frequently the case with cortisol distributions (Adam & Kumari, 2009). We tried to apply a natural logarithmic transformation in order to normalize the distribution. Since the transformation did not produce the expected effect (i.e., normalized the distribution, produced different statistical outcome), we decided to keep the original data.

We ran Spearman rho correlation to provide an overview of the association between our different variables. Furthermore, we used a multiple regression test to assess the relation between our variables; mean cortisol value and HR as the independent variables and the stereotypy as the dependent variable. We decided to use the number of sessions ($n = 60$) as the number of observations, and to control for the non-independence of our data (considering that these 60 sessions belong to 4 participants). The analysis was performed using Mplus. The `type=COMPLEX` function with clustering was applied in order to take into account the interdependence of the data. This function allows to obtain standard errors that reflect the nature of the clustering and that is robust to non-independence and non-normality (McNeish et al., 2017). As our study was preliminary, we considered $p < 0.05$ as significant and $p < 0.1$ as marginally significant. Because the chi-square (χ^2) is sensitive to sample size, model fit was assessed using the Root Mean Square Error of Approximation (RMSEA), the Comparative Fit

Index (CFI), and the Tucker-Lewis Index (TLI; Hu & Bentler, 1999; Marsh et al., 2005; Yu, 2002). RMSEA values smaller than 0.08 and 0.06 respectively suggest acceptable and excellent model fit. Values above .90 and .95 for the CFI and TLI respectively indicate adequate and excellent model fit. To explain whether our results was explained by between or within-participant patterns, we generated correlation graphs that illustrated cortisol and stereotypy data.

Results

Figure 1 shows the variation of mean values of cortisol (left panels) and the percentage of global stereotypy (right panels) across the three conditions. No condition generated a distinct level of cortisol for the four participants. Nevertheless, the patterns showed enough variability to conduct further analyses. Figure 2 presents variations in percentage of motor and vocal stereotypy across conditions. The three conditions were undifferentiated in terms of cortisol levels and all stereotypy levels (global, motor and vocal). We applied the rule developed by Lanovaz et al. (2019) and the results showed that there was no significant difference between the three conditions. Hence, we treated all the sessions as being undifferentiated in terms of condition for the upcoming analyses.

Table 3 presents the Spearman rho correlations between all the variables and Figure 3 offers a more detailed perspective of these correlations. The upper panel of the Figure 3 shows the relationship between the percentage of time engaged in global stereotypy and the mean value of cortisol and HR for each participant. The middle panel presents the relationship with motor stereotypy, and the lower panel demonstrates the relationship between cortisol and HR with the percentage of time engaged in vocal stereotypy. Each form of data point refers to a different participant, and each point refers to a different session. The mean value of cortisol

was negatively associated with the percentage of global stereotypy ($r_s = -.392, p < 0.01$), whereas the HR was positively associated with the percentage of global stereotypy ($r_s = .533, p < 0.01$). The mean value of cortisol was negatively associated with the percentage of motor stereotypy ($r_s = -.508, p < 0.01$) and the HR and motor stereotypy were positively associated ($r_s = .640, p < 0.01$). No significant correlation was found between mean value of cortisol, HR and vocal stereotypy.

The multiple regression using Mplus revealed that the model fitted the data well (RMSEA = .00, CFI = 1.00, TLI = 1.00). We tested three models, with different dependent variables: percentage of global stereotypy, percentage of motor stereotypy only, and percentage of vocal stereotypy only. Independent variables remained the same across the models (mean value of cortisol and HR). Table 4 displays the standardized regression coefficient (β) and the R-square (R^2). In the first model, both independent variables significantly contributed to the outcome. The mean cortisol value found itself to be negatively related to the percentage of global stereotypy ($\beta = -.291, p < 0.05$), whereas HR contributed positively to the outcome ($\beta = .433, p < 0.05$). The mean value of cortisol and the HR allowed to explain 35% of the percentage of global stereotypy's variance ($R^2 = .353, p < 0.05$). Mean cortisol values being negatively related to the percentage of motor stereotypy ($\beta = -.254, p < 0.05$), and HR contributing positively to the outcome ($\beta = .557, p < 0.001$) signifies that both independent variables contributed significantly to the second model. The mean value of cortisol and the HR explained 47% of the percentage of global stereotypy's variance ($R^2 = .466, p < 0.001$). Unlike the first two models, the mean value of cortisol and HR did not contribute significantly to the third model (percentage of vocal stereotypy). Figure 3 presents the correlations between the variables with each participant represented by a different type of

point. A visual analysis of Figure 3 clearly shows that these relationships were an artefact of between-participant differences rather than within-participant patterns.

Discussion

Altogether, our results indicate that the mean level of cortisol was negatively associated with global and motor stereotypy. These results conflict with three previous studies which identified positive correlation between cortisol levels and stereotypy (Bitsika et al., 2015; Lydon, Healy, Roche et al., 2015; Yang et al., 2015). Our sample's characteristic and the direct observation measures of stereotypy may explain part of these conflicting results. Although our design does not allow us to establish a directional link, our results are consistent with those of Gabriels et al. (2013) who showed that children with ASD presenting higher level of stereotypy had lower cortisol levels. The negative correlation between cortisol, motor and global stereotypy tends to support the hypothesis according to which stereotypy could serve a self-regulation function. Further studies are needed to clarify the causal relationship between these variables.

Our analyses also found a positive relationship between HR and global and motor stereotypy. These results are consistent with those of Lydon et al. (2013) who found that stereotypy led to an increase in HR for three participants with ASD. HR data must be interpreted with caution considering that a stereotypy often simultaneously increase physical agitation and hence, can have a direct impact on the HR. That said, these results tend to support the self-stimulation or automatic reinforcement function hypothesis. The current study did not identify an association between cortisol, HR and vocal stereotypy. This result highlights the possibility that the relationship between stress and stereotypy can vary depending on the form of stereotypy. Previous studies had highlighted the possibility that

stereotypy has different functions depending on the internal state of the individual and/or environmental characteristics. The results of our research add another layer of complexity to this hypothesis in showing that different forms of stereotypy could serve different functions.

To our knowledge, our study is the first to combine multiple biological measures (HR and cortisol) and direct observation measures of stereotypy. Our results highlight the complexity of the relationship between biological measures and stereotypy. We believe that direct observation measures are central to elucidating this complex relationship and our study is a first step in this direction. The inconsistency in the literature highlights the need to develop effective measures that will identify the multiple functions of stereotypy. These findings would have a direct clinical impact, considering that function is central to all behavioral interventions. To address this issue, a large amount of data is needed. An inherent limit of direct observation measures are the resources needed to collect the data (human and financial). Artificial intelligence (AI) algorithms would be a well-suited tool in order to automate the measurement of those behaviors.

This study has some limitations that should be noted. First, our conditions did not yield differential stress levels, which limited the range of cortisol values studied. Second, we did not conduct an experimental functional assessment to assess the environmental variables associated with stereotypy due to the limited amount of data available. Examining environmental variables in the future may uncover novel association between physiological measures and specific behavioral functions. The participants in this study are more likely to experience hypersensitivity. Thereby the physiological equipment could have contributed to participant discomfort, movement artifact and Bluetooth disconnection. Finally, the small size of our sample and the nature of our design could have an impact on the external validity of our

results. Results from our statistical analysis results demonstrated a relationship between cortisol, HR and stereotypy, but the visual analysis of Figure 3 do not clearly show a relation on a within-participant basis. It would be interesting to investigate ways to identify patterns between physiological measures and participant characteristics. Machine learning algorithms would also be well suited to investigate these new paths. For example, researchers could use Machine learning algorithms to determine which combination of physiological measures allows a better prediction of the stereotypy level. The current study did not have sufficient data to do so.

The association between stress and stereotypy remains unclear and future studies are needed to clarify these relationships. In the future, studies should use direct observation measures with a larger sample and with a design that will permit the assessment of the relationship directionality. Considering that the relationship between stress and stereotypy seems to vary between developmental period (Bitsika et al., 2015), emotional state of the individual (Lydon et al., 2013; Willemsen-Swinkels et al., 1998) and form of the stereotypy, we highly recommend that future evaluation be conducted with samples that are as heterogenous as possible. To conclude, we suggest considering less intrusive devices when collecting biological variables such as wearable stress monitoring devices (Gul Airij et al., 2016).

References

- Adam, E. K., & Kumari, M. (2009). Assessing salivary cortisol in large-scale, epidemiological research. *Psychoneuroendocrinology*, *34*(10), 1423-1436.
<https://doi.org/10.1016/j.psyneuen.2009.06.011>
- 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>
- Aman, M., Singh, N., Stewart, A., & Field, C. (1985). The Aberrant Behavior Checklist: A behavior ratings scale for the assessment of treatment effects. *American Journal of Mental Deficiency*, *89*, 485-491.
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (5th ed.)*. American Psychiatric Publishing.
- Bitsika, V., Sharpley, C. F., Agnew, L. L., & Andronicos, N. M. (2015). Age-related differences in the association between stereotypic behaviour and salivary cortisol in young males with an Autism Spectrum Disorder. *Physiology & Behavior*, *152*, 238-243. <https://doi.org/10.1016/j.physbeh.2015.10.010>
- Bodfish, J. W., Symons, F. J., Parker, D. E., & Lewis, M. H. (2000). Varieties of repetitive behavior in autism: Comparisons to mental retardation. *Journal of Autism and Developmental Disorders*, *30*, 237–243. <https://doi.org/10.1023/a:1005596502855>
- Bodfish, J. W., Symons, F. W., & Lewis, M. H. (1999). *The repetitive behavior scale*. In Western Carolina Center research reports.

- Chebli, S. S., Martin, V., & Lanovaz, M. J. (2016). Prevalence of stereotypy in individuals with developmental disabilities: A systematic review. *Review Journal of Autism and Developmental Disorders*, 3(2), 107-118. <https://doi.org/10.1007/s40489-016-0069-x>
- Dickerson, S. S., & Kemeny, M. E. (2004). Acute stressors and cortisol responses: A theoretical integration and synthesis of laboratory research. *Psychological Bulletin*, 130(3), 355-391. <https://doi.org/10.1037/0033-2909.130.3.355>
- Dozier, M., Manni, M., Gordon, M. K., Peloso, E., Gunnar, M. R., Stovall-McClough, K. C., Eldreth, D., & Levine, S. (2006). Foster children's diurnal production of cortisol: An exploratory study. *Child Maltreatment*, 11(2), 189-197. <https://doi.org/10.1177/1077559505285779>
- Gabriels, R. L., Agnew, J. A., Pan, Z., Holt, K. D., Reynolds, A., & Laudenslager, M. L. (2013). Elevated repetitive behaviors are associated with lower diurnal salivary cortisol levels in autism spectrum disorder. *Biological Psychology*, 93(2), 262-268. <https://doi.org/10.1016/j.biopsycho.2013.02.017>
- García-Villamizar, D., & Rojahn, J. (2015). Comorbid psychopathology and stress mediate the relationship between autistic traits and repetitive behaviours in adults with autism: Mediators of repetitive behaviour in autism. *Journal of Intellectual Disability Research*, 59(2), 116-124. <https://doi.org/10.1111/jir.12083>
- Gordon, M. K., Peloso, E., Auker, A., & Dozier, M. (2005). Effect of flavored beverage crystals on salivary cortisol enzyme-immunoreactive assay measurements. *Developmental Psychobiology*, 47(2), 189-195. <https://doi.org/10.1002/dev.20081>
- Groden, J., Cautela, J., Prince, S., & Berryman, J. (1994). The Impact of stress and anxiety on individuals with autism and developmental disabilities. In E. Schopler & G. B.

- Mesibov (Éds.), *Behavioral issues in autism* (p. 177-194). Springer US.
https://doi.org/10.1007/978-1-4757-9400-7_9
- Gul Airij, A., Bakhteri, R., & Khali-Hani, M. (2016). Smart wearable stress monitoring device for autistic children. *Jurnal Teknologi*, 78, 75-81. <https://doi.org/10.11113/jt.v78.9453>
- Heathers, J. A. J., Gilchrist, K. H., Hegarty-Craver, M., Grego, S., & Goodwin, M. S. (2019). An analysis of stereotypical motor movements and cardiovascular coupling in individuals on the autism spectrum. *Biological Psychology*, 142, 90-99.
<https://doi.org/10.1016/j.biopsycho.2019.01.004>
- Hu, L., & Bentler, P. M. (1999). Cut-off criteria for fit indices in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1-55. <https://doi.org/10.1080/10705519909540118>
- Hutt, C., Forrest, S. J., & Richer, J. (1975). Cardiac arrhythmia and behaviour in autistic children. *Acta Psychiatrica Scandinavica*, 51(5), 361-372.
<https://doi.org/10.1111/j.1600-0447.1975.tb00014.x>
- Lanovaz, M. J., Cardinal, P., & Francis, M. (2019). Using a visual structured criterion for the analysis of alternating-treatment designs. *Behavior Modification*, 43(1), 115-131.
<https://doi.org/10.1177/0145445517739278>
- Lanovaz, M. J., & Sladeczek, I. E. (2011). Vocal stereotypy in children with autism: Structural characteristics, variability, and effects of auditory stimulation. *Research in Autism Spectrum Disorders*, 5(3), 1159-1168. <https://doi.org/10.1016/j.rasd.2011.01.001>
- Lovaas, I., Newsom, C., & Hickman, C. (1987). Self-stimulatory behavior and perceptual reinforcement. *Journal of Applied Behavior Analysis*, 20(1), 45-68.
<https://doi.org/10.1901/jaba.1987.20-45>

- Lydon, S., Healy, O., & Dwyer, M. (2013). An examination of heart rate during challenging behavior in autism spectrum disorder. *Journal of Developmental and Physical Disabilities, 25*(1), 149-170. <https://doi.org/10.1007/s10882-012-9324-y>
- Lydon, S., Healy, O., Mulhern, T., & Hughes, B. M. (2015). Heart rate measurement during stereotyped motor behavior in autism spectrum disorder. *Journal of Developmental and Physical Disabilities, 27*(5), 677-699. <https://doi.org/10.1007/s10882-015-9445-1>
- Lydon, S., Healy, O., Roche, M., Henry, R., Mulhern, T., & Hughes, B. M. (2015). Salivary cortisol levels and challenging behavior in children with autism spectrum disorder. *Research in Autism Spectrum Disorders, 10*, 78-92. <https://doi.org/10.1016/j.rasd.2014.10.020>
- Mahone, E. M., Bridges, D., Prahme, C., & Singer, H. S. (2004). Repetitive arm and hand movements (complex motor stereotypies) in children. *The Journal of Pediatrics, 145*(3), 391-395. <https://doi.org/10.1016/j.jpeds.2004.06.014>
- Marsh, H. W., Hau, K.-T., & Kong, C.-K. (2002). Multilevel causal ordering of academic self-concept and achievement: Influences of language of instruction (English compared with Chinese) for Hong Kong Students. *American Educational Research Journal, 39*(3), 727-763. <https://doi.org/10.3102/00028312039003727>
- McNeish, D., Stapleton, L. M., & Silverman, R. D. (2017). On the unnecessary ubiquity of hierarchical linear modeling. *Psychological Methods, 22*(1), 114-140. <https://doi.org/10.1037/met0000078>
- Melo, C., Ruano, L., Jorge, J., Pinto Ribeiro, T., Oliveira, G., Azevedo, L., & Temudo, T. (2020). Prevalence and determinants of motor stereotypies in autism spectrum

- disorder: A systematic review and meta-analysis. *Autism*, 24(3), 1-22.
<https://doi.org/10.1177/1362361319869118>
- Muthugovindan, D., & Singer, H. (2009). Motor stereotypy disorders. *Current Opinion in Neurology*, 22(2), 131-136. <https://doi.org/10.1097/WCO.0b013e328326f6c8>
- Parkhurst, J., & Kawa, J. M. (2018). Childhood Autism Rating Scales. In J. Kreutzer, J. DeLuca, & B. Caplan (Eds.), *Encyclopedia of Clinical Neuropsychology* (pp. 1-3). Springer International Publishing. https://doi.org/10.1007/978-3-319-56782-2_1530-2
- Querim, A. C., Iwata, B. A., Roscoe, E. M., Schlichenmeyer, K. J., Viruès Ortega, J., & Hurl, K. E. (2013). Functional analysis screening for problem behavior maintained by automatic reinforcement. *Journal of Applied Behavior Analysis*, 46(1), 47-60.
<https://doi.org/10.1002/jaba.26>
- Rapp, J. T., & Vollmer, T. R. (2005). Stereotypy I: A review of behavioral assessment and treatment. *Research in Developmental Disabilities*, 26(6), 527-547.
<https://doi.org/10.1016/j.ridd.2004.11.005>
- Rojahn, J., Matson, J. L., Lott, D., Esbensen, A. J., & Smalls, Y. (2001). The Behavior Problems Inventory: An instrument for the assessment of self-injury, stereotyped behavior and aggression/destruction in individuals with developmental disabilities. *Journal of Autism and Developmental Disorders*, 31(6), 577-588.
- Schopler, E., Van Bourgondien, M. E., Wellman, G. J., & Love, S. R. (2010). *The Childhood Autism Rating Scale* (2nd ed.). Western Psychological Services.
- Sharpley, C. F., Bitsika, V., Andronicos, N. M., & Agnew, L. L. (2016). Is afternoon cortisol more reliable than waking cortisol in association studies of children with an ASD? *Physiology & Behavior*, 155, 218-223. <https://doi.org/10.1016/j.physbeh.2015.12.020>

- Sroufe, L. A., Stuecher, H. U., & Stutzer, W. (1973). The functional significance of autistic behaviors for the psychotic child. *Journal of Abnormal Child Psychology*, *1*(3), 225-240. <https://doi.org/10.1007/BF00917733>
- van de Mortel, T. F. (2008). Faking it: Social desirability response bias in self-report research. *Australian Journal of Advanced Nursing*, *25*(4), 40-48.
- Willemsen-Swinkels, S. H. N., Buitelaar, J. K., Dekker, M., & van Engeland, H. (1998). Subtyping stereotypic behavior in children: The association between stereotypic behavior, mood, and heart rate. *Journal of Autism and Developmental Disorder*, *28*, 547-557. <https://doi.org/10.1023/A:1026008313284>
- Yang, C.-J., Tan, H.-P., Yang, F.-Y., Wang, H.-P., Liu, C.-L., He, H.-Z., Sang, B., Zhu, X.-M., & Du, Y.-J. (2015). The cortisol, serotonin and oxytocin are associated with repetitive behavior in autism spectrum disorder. *Research in Autism Spectrum Disorders*, *18*, 12-20. <https://doi.org/10.1016/j.rasd.2015.07.002>
- Yu, C.-Y. (2002). Evaluation of model fit indices for latent variable models with categorical and continuous outcomes. *Unpublished dissertation*. Retrieved from Mplus website <http://www.statmodel.com/download/Yudissertation.pdf>

Table 1*Participant Characteristics*

Participants	Age	Gender	CARS-II T-score	Verbal communication score (CARS-II)
Ana	6	F	61	4
Matt	9	M	55	3.5
Jim	5	M	58	3.5
Oli	5	M	49	4

Table 2*Operational Definitions for Each Participant's Behaviors*

	Target behavior	Behavioral definition
All	Repetitive movements	Two or more rapid back and forth movements of the same body part (without a clear functional purpose)
	Rocking	Two or more rapid back and forth movements of the torso
	Vocal stereotypy	Acontextual sounds or words produced by the vocal apparatus
	Toy stereotypy	Repetitively activating a toy that produces light or sounds
	Jumping	Repetitively having both feet leave the ground while standing
Ana	Object stereotypy	Two or more back and forth movements of an object
Matt	Rubbing	Rubbing any body part with hand for at least 2 consecutive seconds
	Scratching	Contact between the child's nails and any body part
	Visual stimulation	Putting his forearm or the crest of his elbow on his forehead or eyes
Jim	Ear touching	Two or more consecutive contacts between the child's ears and any part of his hand
Oli	Arm flapping	Two or more consecutive up and down arm movements
	Face rubbing	Touching his face with any part of an object

Table 3*Spearman's Rho Correlations*

	HR	Cortisol (mean)	Global stereotypy (%)	Motor stereotypy (%)	Vocal stereotypy (%)
HR	1.00	-.411**	.533**	.640**	-.156
Cortisol (mean)	-.411**	1.00	-.392**	-.508**	.129
Global stereotypy (%)	.533**	-.392**	1.00	.818**	.318*
Motor stereotypy (%)	.640**	-.508**	.818**	1.00	-.098
Vocal stereotypy (%)	-.156	.129	.318*	-.098	1.00

** $p < .01$ * $p < 0.05$

Table 4

Summary of Multiple Regression Analysis for Stereotypy Regressed on Mean Cortisol Values and Heart Rate

	β	S.E.	p
Model 1			
Global stereotypy (%)			
Cortisol (mean)	-.291	0.124	0.019
Heart rate	0.433	0.181	0.016
	R^2		
	0.353	0.155	0.023
Model 2			
Motor stereotypy (%)			
Cortisol (mean)	-0.254	0.117	0.030
Heart rate	0.557	0.137	0.000
	R^2		
	0.466	0.145	0.001
Model 3			
Vocal stereotypy (%)			
Cortisol (mean)	-0.077	0.078	0.324
Heart rate	-0.205	0.193	0.287
	R^2		
	0.038	0.069	0.583

Figure 1

Level of Mean Value of Cortisol (Left Panels) and Percentage of Time Engaged in Global Stereotypy (Right Panels) Across Three Conditions: Control, Novelty and Unpredictability

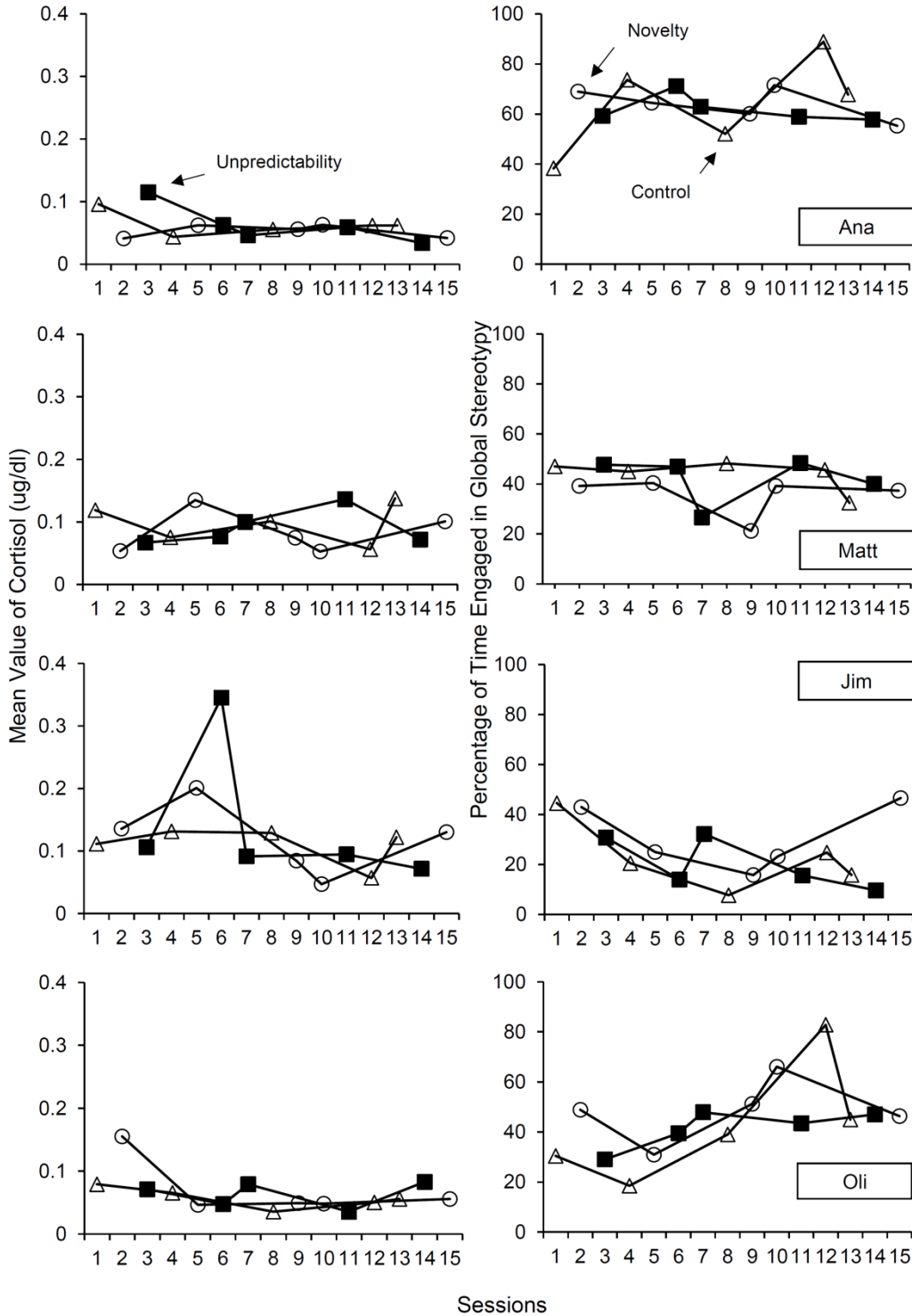


Figure 2

Percentage of Time Engaged in Motor Stereotypy (Left Panels) and Vocal Stereotypy (Right Panels) Across Three Conditions: Control, Novelty and Unpredictability

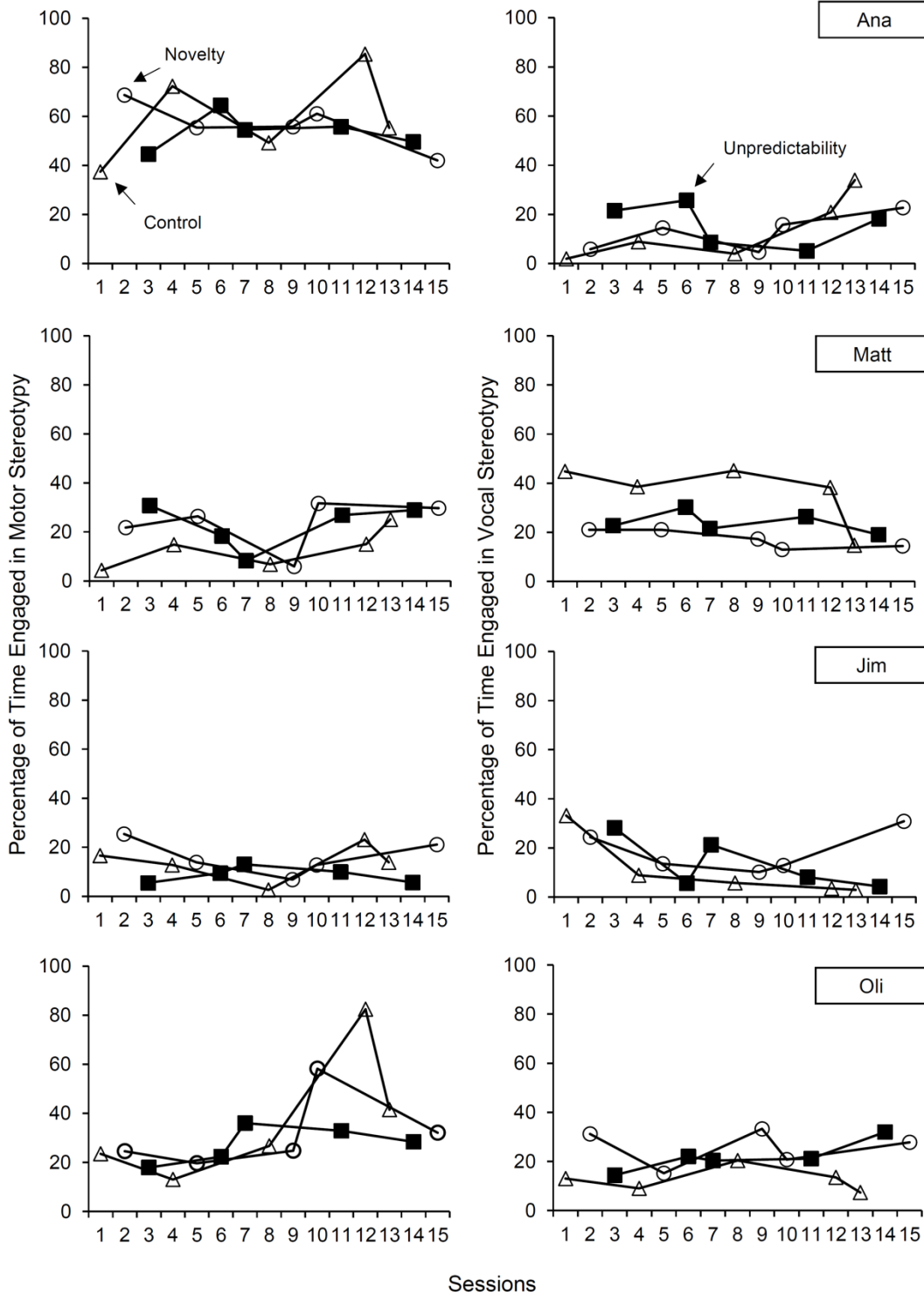
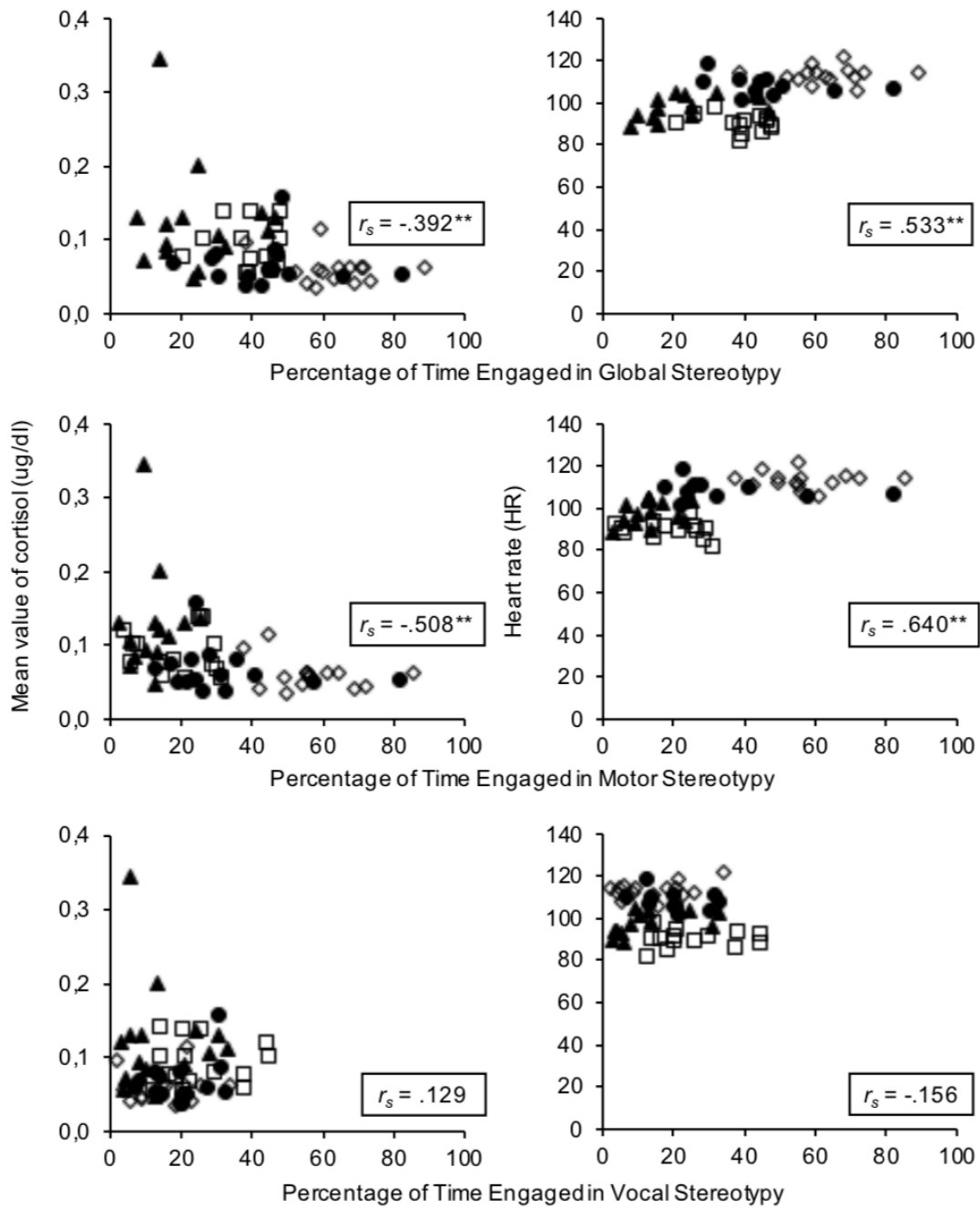


Figure 3

Mean Value of Cortisol (ug/dl) and Heart Rate (HR) Compared to Percentage of Time Engaged in Global Stereotypy, Motor Stereotypy and Vocal Stereotypy.



Chapitre IV : Article 3

Artificial intelligence for the measurement of vocal stereotypy

Marie-Michèle Dufour^{1,2}, Marc J. Lanovaz^{1,2}, and Patrick Cardinal³

¹École de psychoéducation, Université de Montréal

²Centre de recherche de l'Institut universitaire en santé mentale de Montréal

³Département de génie logiciel et des TI, École de technologie supérieure

Cet article a été accepté pour publication au *Journal of the Experimental Analysis of Behavior*

Résumé

Tant les chercheurs que les cliniciens ont fréquemment recours à l'observation directe pour mesurer divers comportements. Lorsque ces comportements sont trop complexes ou trop nombreux pour être mesurés in vivo, le recours à l'observation directe par des observateurs humains augmente considérablement les coûts de la recherche et du suivi des interventions dans la pratique. Pour résoudre ce problème, nous avons effectué une preuve de la validité du concept afin de déterminer si l'intelligence artificielle pouvait être plus performante que le hasard pour mesurer la stéréotypie vocale chez des enfants ayant un trouble du spectre de l'autisme (TSA). Nous avons utilisé un réseau de neurones artificiels et plus de 1 500 minutes d'enregistrements audio de huit enfants ayant un TSA pour entraîner nos modèles à mesurer la stéréotypie vocale. Nos résultats ont montré que le réseau de neurones artificiels a performé adéquatement dans la reconnaissance de la stéréotypie vocale (c.-à-d. corrélation session par session près ou au-dessus de 0.80 avec un observateur humain). Cela dit, des recherches supplémentaires avec plus de données devront être menées avant qu'il soit envisageable d'appliquer cette méthode dans un cadre de recherche ou de pratique.

Mots-clés: intelligence artificielle, autisme, apprentissage profond, mesure, stéréotypie vocale

Abstract

Both researchers and practitioners often rely on direct observation to measure and monitor behavior. When these behaviors are too complex or numerous to be measured in vivo, relying on direct observation using human observers increases the amount of resources required to conduct research and to monitor the effects of interventions in practice. To address this issue, we conducted a proof of concept examining whether artificial intelligence could measure vocal stereotypy in individuals with autism. More specifically, we used an artificial neural network with over 1,500 minutes of audio data from eight different individuals to train and test models to measure vocal stereotypy. Our results showed that the artificial neural network performed adequately (i.e., session-by-session correlation near or above .80 with a human observer) in measuring engagement in vocal stereotypy for six of eight participants. That said, researchers need to conduct additional research to further improve the generalizability of the approach.

Keywords: artificial intelligence, artificial neural network, autism, measurement, stereotypy

Artificial Intelligence for the Measurement Vocal Stereotypy

Whether experimental or applied, the science of behavior analysis targets a wide range of topics that aim to understand and to improve the functioning of human organisms (Skinner, 1951). A common thread central to this endeavor is the measurement of behavior. In most research involving human participants, researchers rely either on permanent products (e.g., response automatically recorded from pressing on a computer screen) or on direct observation to examine the effects of independent variables on behavior. Practicing behavior analysts must also use these measures to monitor behavior when directly intervening with individuals in a professional setting (Behavior Analyst Certification Board, 2017).

The reliance on direct observation using human observers in many contexts raises an important issue related to available resources. For example, using continuous recording to monitor multiple or high frequency behavior often requires an independent observer that scores the behavior from a video recording. Furthermore, researchers need to include a second observer to increase the believability of the results by monitoring interobserver agreement (Mudford et al., 2009). If a researcher has 20 hr of video recordings to score for one participant, the human resources can easily add up to 30 hr to 50 hr of work. This time does not include the resources and time involved in hiring and training additional staff to conduct these tasks. With many participants, these additional resources can add up rapidly and limit the amount of research that can be done or the number of intervention sessions that can be afforded.

One potential solution to significantly reduce response effort associated with direct observation is to use artificial intelligence. Broadly, artificial intelligence is “the study of how to make computers do things at which, at the moment, people do better” (Rich & Knight,

1991, p. 3). As such, the measurement of behavior is a topic well suited to artificial intelligence as human observers are currently better than computers at monitoring most types of behavior (Goodwin et al., 2011). One promising tool in artificial intelligence is machine learning, which involves training models to detect signals or patterns in data (see Lanovaz et al., 2020, for behavior analytic introduction to the topic). That is, machine learning takes data as input to develop mathematical models that allows them to predict the value or categorization of novel data.

One type of machine learning algorithm is the artificial neural network. Simply put, an artificial neural network takes input data from the experimenter, which are then transformed by mathematical functions to produce a prediction (Goodfellow et al., 2016). Typically, artificial neural networks contain three types of layer: (1) the input layer, (2) the hidden layer, and (3) the output layer (see Figure 1). The input layer receives features to train the model. The hidden neurons allow the model to learn more complex relationships between these features by transforming the data. Finally, the output layer provides the prediction of the model. Mathematically, the algorithm multiplies the input data by weights (which are initially set randomly) and transforms the product using an activation function to standardize the data. The result is then multiplied by a second series of weights and again transformed by an activation function. Next, the algorithm computes an error using the binary cross-entropy function, which compares the output values with the class labels (i.e., the true values). Finally, the algorithm retropropagates the gradient (derivative) of the error to update the weights. The retropropagation of the gradient of the error involves a variable called the learning rate, which determines how “fast” the model changes the weights. The retropropagation should produce updated weights that typically lead to more accurate predictions (less error) on the next pass.

Each pass across the steps is called an epoch. The whole process is akin to shaping in behavior analysis where the model updates itself to provide increasingly more accurate responses following feedback.

One challenge with artificial neural networks is that the training of models requires a large amount of data. Considering that individuals with developmental disability often engage in high rates of repetitive behavior, a starting point could be to apply these algorithms to this population. Individuals with developmental disability often engage in stereotypy, which are repetitive behaviors characterized by movement invariance that are not maintained by social contingencies (Rapp & Vollmer, 2005). Researchers and practitioners further divide stereotypy into two types: motor and vocal stereotypy. Some researchers have already evaluated the use of machine learning algorithms to identify motor stereotypy in this population (Goodwin et al., 2011; Min & Tewfik, 2010; Rad & Furlanello, 2016; Westeyn et al., 2005).

In the first study on the automated detection of stereotypy, Westeyn et al. (2005) used accelerometers to monitor behaviors and then applied a Hidden Markov Models to classify the data. Their model was capable of automatically and accurately detecting 69% of hand flapping in one typically developing adult who was mimicking stereotypy. Following this study, other researchers applied different algorithms (k-nearest neighbors and iterative subspace identification) to monitor body rocking and hand flapping in persons with autism with promising accuracy varying between 83% and 90% (Goodwin et al., 2011; Min & Tewfik, 2010). Finally, Rad and Furlanello (2016) applied artificial neural networks to detect motor stereotypy and found preliminary results that support the relevance of this approach.

The prior studies have all focused on motor stereotypy, but another form that should be targeted for reduction is vocal stereotypy (Rapp et al., 2013). A recent systematic review reported that 48% of individuals with developmental disability engage in at least one form of vocal stereotypy (Chebli et al., 2016). Examples of forms reported in the literature include monosyllable vocalizations, acontextual words or phrases, acontextual laughing and grunting (DiGennaro Reed et al., 2012; Lanovaz et al., 2011; Rapp et al., 2013; Spencer & Alkhanji, 2018; Weston et al., 2018). Researchers have also evaluated the potential of machine learning for measuring vocal stereotypy (Min & Fetzner, 2018, 2019). In their first article, Min and Fetzner (2018) used a traditional machine learning algorithm to detect vocal stereotypy in four children with autism spectrum disorder (ASD) with an accuracy between 73% and 93%. In a second study, Min and Fetzner (2019) applied an artificial neural network to detect vocal stereotypy with an accuracy of 85%.

A serious limitation of both prior studies was that the researchers only measured whether stereotypy was absent or present in brief videoclips. If the procedures are to be useful to researchers and practitioners, we must develop models that can measure the duration of stereotypy (not only its presence or absence) during longer sessions. As a matter of fact, finding ways to automate the measurement of the duration of vocal stereotypy could not only decrease the amount of resources required for conducting research, but also facilitate the monitoring of the behavior in applied settings. A second limitation is that the researchers did not extract videos that involved treatment. Researchers and practitioners are likely to measure vocal stereotypy in baseline and treatment sessions, which is why examining the models under both these conditions is important. One common treatment for vocal stereotypy is providing access to noncontingent music (i.e., continuous preferred music). Researchers have repeatedly

shown that this treatment is effective at reducing vocal stereotypy in children with ASD (Gibbs et al., 2018; Lanovaz et al., 2011; Saylor et al., 2012). Given the potential challenges of identifying vocal stereotypy during this treatment condition, the inclusion of such sessions when testing models appears essential.

Thus, the purpose of our study was to determine whether we could train a machine learning algorithm to measure the duration of vocal stereotypy using audio data extracted from video recordings of both baseline and treatment sessions. More specifically, our study examined whether models derived from an artificial neural network could produce session-by-session correlations at or above .80 when compared to the values measured by a human observer. We set the benchmark at .80 because (a) this correlation score is considered strong to very strong (Schober et al., 2018), and (b) prior research has found similar or higher correlations when examining the validity of discontinuous measurements methods (Leblanc et al., 2020). Given the purpose of our study, our research questions were:

- Can an artificial neural network produce a session-by-session correlation of .80 or more with human observers when measuring the duration of vocal stereotypy?
- What type of data analysis (i.e., within-participant, between-participant, or hybrid approach) produces the best measures of vocal stereotypy?

Method

Participants and Settings

To train and test our models, we measured vocal stereotypy in eight children with ASD who had previously participated in a study on the use of a mobile app to reduce engagement in stereotypy (Préfontaine et al., 2019). The mobile app involved the delivery of noncontingent

music for children who engaged mainly in vocal stereotypy. Two participants, Dave and Emile, also received differential reinforcement as an intervention during some of the sessions. An independent multidisciplinary team had provided a diagnosis of ASD to each child prior to their participation in the study. All participants received their schooling or their primary care in English or French. Table 1 presents the age, gender, and a description of the vocal stereotypy for each participant. Our university research ethics board approved both the original and the current study.

The sessions occurred in each child's home during regularly planned activities, which affected the quality of audio recordings. Therefore, we excluded sessions (a) where siblings could be heard in the background because the first author was unable to determine which sounds were emitted by whom when relying on the audio recordings, and (b) in which ambient noise or sounds impeded the measurement of vocal stereotypy by the first author. Each session lasted between 10 and 15 min. The total number of sessions was determined based on the efficacy of the intervention. Table 2 presents the number of sessions and the time in seconds of recordings for each participant that we used as part of the current study.

Extraction of Audio Recordings and Features

The first author extracted the audio from standard definition video recordings in .mpeg format using VLC®, an open source video software. The program extracted the audio to a .wav format with a sampling rate of 22,050 Hz, one audio channel and a bit rate of 16 Kbits/s. Both the human observers used this .wav file to measure vocal stereotypy. Artificial neural networks cannot analyze .wav files because the amount of information encoded is too large for processing with a typical computer. To address this issue, we used the package `python_speech_features` for Python to extract the Mel Filterbank Cepstrum Coefficient

(MFCC) from the audio files. The MFCC allows the extraction of a set of 26 audio features that have been widely used in machine learning, specifically in speech recognition (Chia et al., 2012; Kumar et al., 2011). These 26 features provide a description of the sound during short time windows. Our algorithm sampled the sound every 0.1 s for a time window of the same duration. Therefore, each second of recording was represented by 260 features (10 timesteps multiplied by 26 features).

Artificial Neural Networks

One type of machine learning algorithm is the artificial neural network. Simply put, an artificial neural network takes input data from the experimenter, which are then transformed by mathematical functions to produce a prediction (Goodfellow et al., 2016). Typically, artificial neural networks contain three types of layers: (1) the input layer, (2) the hidden layer, and (3) the output layer (see Figure 1). The input layer receives features to train the model. As indicated earlier, our input involved 260 features per second of recording. The hidden neurons, which process the inputs received from the input layer, allow the model to learn more complex relationships between the features of the sounds by transforming the data (see paragraph below).

Our raw data and Python code are freely available on the Open Science Framework (see <https://osf.io/e4vbs/>). Figure 1 depicts our artificial neural network. The input involved 260 features per second of recording (see extraction of audio recordings and features section). Based on broad recommendations by Heaton (2015) and on the computation power available to us, our neural network contained a single hidden layer with 128 neurons. Our model involved a single binary output value for each second: vocal stereotypy present (value = 1) or vocal stereotypy absent (value = 0). For our analysis, we used the Adam optimizer to set our

learning rate. The algorithm trained our models until the kappa metric (see below) had not improve for 10 consecutive epochs (i.e., loops) on the validation set. Given that our audio recordings contained more seconds containing silences than vocal stereotypy, we also applied a correction to the error to balance them out. To promote generalization to untrained exemplars, our algorithm also applied dropout regularization, which randomly left out 20% of data in each layer when making predictions.

Data Collection and Interobserver Agreement

The first author, a doctoral student, manually coded each audio recording on a second-by-second basis using Audacity®, an open source audio software. We defined vocal stereotypy as acontextual or unintelligible sounds or words produced by the vocal apparatus of the child. If vocal stereotypy was present even for a fraction of a second, the first author coded the behavior as occurring during the second (as in partial interval recording with 1-s intervals). Otherwise, she scored the behavior as not occurring during the second. A second observer measured vocal stereotypy on 42% of the recordings. The mean second-by-second interobserver agreement was 97% (range: 93%-99%) and the mean kappa interobserver agreement was .87 (range: .81-.94).

Procedures

Between-Participant Analysis. Our first analyses aimed to determine whether our models could predict the duration of vocal stereotypy for children whose data were not used during the training of the model. If the model produced adequate predictions, behavior analysts could develop models that could be applied to any child who engages in vocal stereotypy. To conduct the analysis, we used a variation of the leave one out cross-validation methodology (Wong, 2015). Our code divided our participants into three sets: the training set

(six participants), the validation set (one participant), and the test set (one participant). Our algorithm used the training set to train and update the model, and the validation set to determine when to stop the training and to select the model that produced the highest kappa value. The test set assessed generalization as it was not used during training, nor to select the most accurate model. The program repeated our analyses eight times so that each participant was in the test set once and in the validation set once.

For each model, we measured accuracy, the kappa statistic and a correlation on the test (generalization) set. To measure accuracy, the code divided the number of seconds on which the prediction of our model and the observation of the first coder agreed by the total number of seconds in the dataset. Accuracy can be easily skewed by unbalanced datasets by which a model is better at predicting the absence than the presence of vocal stereotypy. Therefore, we also measured agreement using the kappa statistic as it provides control over agreements that are the result of chance and balances the values of both possible outcomes. Kappa and accuracy only provide a within-session measure of agreement. Behavior analysts typically consider session-by-session patterns when analyzing single-case data in graphs. To address this issue, we added a measure comparing session-by-session values. Specifically, our program measured correlations on a session-by-session basis between the percentage of vocal stereotypy computed by the model and the percentage of stereotypy observed by the first coder. Since the original study evaluated the effectiveness of delivering noncontingent music, we had access to sessions with and without music.

To examine the potential effects of music on measurement, the previous analyses were repeated twice: once with a dataset including all sessions and a second time with a dataset

excluding sessions with music¹.

Within-Participant Analysis. Our second series of analyses involved examining whether we could produce better results using within-participant predictions. If the model produced adequate predictions, behavior analysts could score the first few sessions of a child for vocal stereotypy and then use the model to predict vocal stereotypy in subsequent sessions. In this case, we conducted the analysis for each participant individually. The test set contained a single session for the participant whereas the remaining sessions were divided between the training set (83% of the remaining sessions) and the validation set (16% of the remaining sessions). The code repeated the analysis once per session for each participant. Apart from the composition of the test, training, and validation sets, the procedures and analyses remained the same as for the between-participant analyses. Moreover, we did not test for the effects of music because our amount of data would have been insufficient for many participants. As our analysis produced multiple values for each participant, the article reports the means across sessions.

Hybrid Analysis. Our third series of analyses involved combining within- and between-participant data. As in the within-participant analysis, the test set contained a single session for the participant whereas the remaining sessions were divided between the training set (67% of the remaining sessions) and the validation set (33%² of the remaining sessions). However, we also added between-participant data in such a way that the training set contained 50% of within-participant data and 50% of data from other participants, which increased the

¹ Emile wore headphones during most of the noncontingent music sessions. For analyses purposes, these sessions were categorized as having no music as the audio recording contained no background music.

² We could add more sessions in the validation set because our training set was larger due to between-participant data.

number of samples. Because the between-participant data had a lot more samples, the algorithm picked the samples randomly to match the number from the within-participant component.

Results

As shown in Table 2, the data for our study involved between 6 and 35 sessions per participant for a total of 142 sessions from eight participants. The total duration of the 142 sessions was 99,564 s. The total duration of sessions with music was 25,887 s whereas the duration of sessions without music was 73,677 s (see Table 2). We used the previous data to develop our machine learning models using between-participant, within-participant, and hybrid analyses.

The left side of Table 3 and Figure 2 presents the results of the between-participant analyses. Five of the eight participants had kappa statistics above or close to 0.5, indicating moderate to substantial agreement between the human observer and the computer model. For these five participants, the session-by-session correlation between the human observer and the computer model remained above .80, which indicates a strong to very strong correlation (see Figure 2). Two participants (i.e., Alia and Nate) had negative correlations, which indicates that models were more likely to record an instance vocal stereotypy when the behavior was absent than when it was present. Therefore, we repeated the analysis on sessions without music only to determine whether the background music was misleading the algorithms (see right side of Table 3 and Figure 3). For Alia and Nate, all measures improved. However, the removal of music sessions also worsened the kappa and correlations for two participants (i.e., Owen and Dan).

Table 4 and Figure 4 present the results of the within-participant analyses. Rather than

using the data from the other participants to train the models (as in our between-participant analyses), the within-participant analyses consisted of training the models with the participant's own data. This manipulation involved a tradeoff: It reduced the amount of data available in the training set for each participant, but it also made the training set more similar to the vocal stereotypy that we were trying to measure. The results show that the kappa statistics were higher in the within-participant analysis than in the between-participant analysis for four participants. In contrast, the correlations improved for six of eight participants after removal of music sessions. A further examination of these data indicates that this result may be misleading. This improvement involved the three participants that had the lowest correlations in the between-participant analyses. As such, less participants achieved the .80 correlation criterion in the within-participant analyses (i.e., 4) than in the between-participant analyses (i.e., 5).

As discussed previously, one issue with within-participant analyses is that the training sets were smaller than in the between-participant analyses (i.e., anywhere between 6% and 27% the number of samples). To address this concern, we further conducted an analysis using a hybrid method combining the within- and between-participant analyses. Table 5 and Figure 5 show the results of the hybrid analyses. Adding between-participant data to the within-participant models increased correlations near or above .80 for two more participants (i.e., Dave and Alia), which led to the models adequately predicting session-by-session patterns for six of eight participants.

An unexpected observation from the previous analyses was that the kappa scores did not necessarily increase when correlations increased. Kappa represents within-session patterns of responding whereas correlations capture between-session patterns (e.g., immediacy, level,

trend). One potential explanation is that errors in measurement in sessions with low levels of stereotypy may deflate the mean kappa score. As an example, assume that a human observer measured a behavior for 0.3% of a session whereas the model did not detect the behavior (i.e., 0%). Despite the absolute difference being only 0.3%, the kappa score would be 0 for this session. To examine this hypothesis, we measured the correlation between the kappa score and the percentage of engagement on a session-by-session basis. Figure 6 shows an example of this correlation for Billy-Peter³. Our analyses found a positive correlation between kappa and percentage of engagement for all participants, indicating that sessions with low levels of stereotypy skewed the estimation of the kappa scores towards lower values (as the computation of the reported kappa scores involved the mean of all sessions).

Discussion

Our proof of concept produced session-by-session correlations near or above .80 for six of eight participants when using a hybrid approach, which generally produced the best outcomes. The hybrid approach may have performed best because it took advantage of each participant's individual responding while augmenting the dataset with samples from other participants. Interestingly, the removal of music during the between-participant analyses significantly improved the measures for two further participants while worsening the predictions for two others. This worsening of the results may be explained by the removal of the data from the music sessions. The algorithm trained and tested the models on less data in the sessions without music only, which could explain the reduction for some participants.

³ We chose Billy-Peter as an example because he showed the largest discrepancy between kappa and correlation scores.

Nonetheless, the results are encouraging as the high correlations observed in the hybrid analyses included both baseline and treatment sessions.

As noted in the results, the analyses often produced better estimations of between-session patterns than within-session patterns, which may be partly caused by the difficulty in estimating low levels of behavior. This result is consistent with prior research by Leblanc et al. (2020) who found that discontinuous methods of measurement produced less accurate estimates when challenging behavior occurred less frequently. Another potential explanation is that machine learning may produce systematic minor errors at the within-session level that have a limited effect at the between-session level. This type of systematic error is not unheard of in behavior analysis. One notable example is the use discontinuous recording methods. Although discontinuous methods may produce considerably different within-session patterns, between-session patterns are similar enough to make these tools useful in practice (LeBlanc et al., 2020; Meany-Daboul et al., 2007; Rapp et al., 2008; Schmidt et al., 2013). Similarly, our machine learning models preserved important between-session features used for the analysis of single-case designs such as level, trend, and immediacy while producing less consistent within-session patterns.

To our knowledge, our study is the first to use artificial intelligence algorithms to measure the duration of vocal stereotypy during sessions. Our results replicate and extend prior studies that have used machine learning to measure motor and vocal forms of stereotypy (Goodwin et al., 2011; Min & Tewfik, 2010; Rad & Furlanello, 2016; Westeyn et al., 2005). We also extend research on artificial intelligence as we studied how we can program computers to perform a task at which humans are currently better at. Notably, some of our models produced session-by-session correlations that rivaled with those produced by

discontinuous measurement methods (Leblanc et al., 2020). Despite the promising nature of our results, we consider our study as an experimental proof of concept as the session-by-session correlations remained inadequate for two participants. As researchers working in a university (non-clinical) setting, we simply did not have access to sufficient data to further improve the performance of our models. Our artificial neural network trained the models on audio data extracted from video recordings. Hence, the distance between the microphone and the child varied within and across sessions, rendering the analysis by the algorithms more challenging. In the future, we recommend that researchers use a wireless microphone positioned on the child's shirt collar, which should considerably improve measurements by increasing the power of the signal produced by the child's vocal apparatus. Moreover, this change would also facilitate the discrimination between the child's sounds and those of other individuals in the environment. If researchers continue improving the current models, the use of artificial intelligence may produce significant changes in research and practice such as the reduction of costs and the automation of certain repetitive tasks. With additional research, we can imagine the development of systems that could automatically measure target vocal behavior within research, educational and clinical contexts, freeing up time for researchers and practitioners to focus on other important activities. The utility of these models could move beyond single-case designs. Researchers could also use automated measures with large randomized samples.

Our study has two additional limitations that should be noted. First, we used a single method and set of hyperparameters to extract the audio and train our models as we lacked the computing power to conduct multiple comparison analyses. Evaluating the effects of the extraction method and hyperparameters on algorithm performance with more powerful

computers (or supercomputers) would be relevant in the future. Second, our study did not examine and compare patterns on single-case graphs. Instead, our procedures used a correlation measure similarly to a recent study examining correspondence between continuous and discontinuous measurements (see Leblanc et al., 2020). Given that we had to remove sessions due to low-quality recordings, the sessions were not necessarily consecutive, preventing a thorough single-case graph analysis. That said, the ultimate litmus test for our approach will be whether functional relations remain observable on single-case graphs when applying these algorithms in research and applied settings.

References

- Behavior Analyst Certification Board (2017). *BCBA/BcABA task list* (5th ed.).
<https://www.bacb.com/wp-content/uploads/2017/09/170113-BCBA-BCaBA-task-list-5th-ed-.pdf>
- Chebli, S. S., Martin, V., & Lanovaz, M. J. (2016). Prevalence of stereotypy in individuals with developmental disabilities : A systematic review. *Review Journal of Autism and Developmental Disorders, 3*(2), 107-118. <https://doi.org/10.1007/s40489-016-0069-x>
- Chia A. O., Hariharan, M., Yaacob, S., & Sin C. L. (2012). Classification of speech dysfluencies with MFCC and LPCC features. *Expert Systems with Applications, 39*(2), 2157-2165. <https://doi.org/10.1016/j.eswa.2011.07.065>
- DiGennaro Reed, F. D., Hirst, J. M., & Hyman, S. R. (2012). Assessment and treatment of stereotypic behavior in children with autism and other developmental disabilities : A thirty year review. *Research in Autism Spectrum Disorders, 6*(1), 422-430.
<https://doi.org/10.1016/j.rasd.2011.07.003>
- Gibbs, A. R., Tullis, C. A., Thomas, R., & Elkins, B. (2018). The effects of noncontingent music and response interruption and redirection on vocal stereotypy. *Journal of Applied Behavior Analysis, 51*(4), 899-914. <https://doi.org/10.1002/jaba.485>
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. The MIT Press.
- Goodwin, M. S., Intille, S. S., Albinali, F., & Velicer, W. F. (2011). Automated detection of stereotypical motor movements. *Journal of Autism and Developmental Disorders, 41*(6), 770-782. <https://doi.org/10.1007/s10803-010-1102-z>
- Heaton, J. (2015). *Artificial intelligence for humans, volume 3: Deep learning and neural networks*. Heaton Research Inc.

- Kumar, K., Kim, C., & Stern, R. M. (2011). Delta-spectral cepstral coefficients for robust speech recognition. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 4784-4787. <https://doi.org/10.1109/ICASSP.2011.5947425>
- Lanovaz, M. J., Giannakakos, A. R., & Destras, O. (2020). Machine learning to analyze single-case data: A proof of concept. *Perspectives on Behavior Science*, 43(1), 21-38. <https://doi.org/10.1007/s40614-020-00244-0>
- Lanovaz, M. J., Sladeczek, I. E., & Rapp, J. T. (2011). Effects of music on vocal stereotypy in children with autism. *Journal of Applied Behavior Analysis*, 44(3), 647-651. <https://doi.org/10.1901/jaba.2011.44-647>
- Leblanc, L. A., Lund, C., Kooken, C., Lund, J. B., & Fisher, W. W. (2020). Procedures and accuracy of discontinuous measurement of problem behavior in common practice of applied behavior analysis. *Behavior Analysis in Practice*, 13(2), 411-420. <https://doi.org/10.1007/s40617-019-00361-6>
- Meany-Daboul, M. G., Roscoe, E. M., Bourret, J. C., & Ahearn, W. H. (2013). A comparison of momentary time sampling and partial-interval recording for evaluating functional relations. *Journal of Applied Behavior Analysis*, 40(3), 501-514. <https://doi.org/10.1901/jaba.2007.40-501>
- Min, C. H., & Fetzner, J. (2018). Vocal stereotypy detection: An initial step to understanding emotions of children with autism spectrum disorder. *2018 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS)*, 3306-3309. <https://doi.org/10.1109/EMBC.2018.8513050>

- Min, C. H., & Fetzner, J. (2019). Training a neural network for vocal stereotypy detection. *2019 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 5451-5455. <https://doi.org/10.1109/EMBC.2019.8856626>
- Min, C. H., & Tewfik, A. H. (2010). Novel pattern detection in children with autism spectrum disorder using iterative subspace identification. *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2266-2269. <https://doi.org/10.1109/ICASSP.2010.5495885>
- Mudford, O. C., Taylor, S. A., & Martin, N. T. (2009). Continuous recording and interobserver agreement algorithms reported in the Journal of Applied Behavior Analysis (1995-2005). *Journal of Applied Behavior Analysis*, 42(1), 165-169. <https://doi.org/10.1901/jaba.2009.42-165>
- Préfontaine, I., Lanovaz, J. M., McDuff, E., McHugh, C., & Cook, J. L. (2019). Using mobile technology to reduce engagement in stereotypy: A validation of decision-making algorithms. *Behavior Modification*, 43(2), 222-245. <https://doi.org/10.1177/0145445517748560>
- Rad, N. M., & Furlanello, C. (2016). Applying deep learning to stereotypical motor movement detection in autism spectrum disorders. *2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)*, 1235-1242. <https://doi.org/10.1109/ICDMW.2016.0178>
- Rapp, J. T., Colby-Dirksen, A. M., Michalski, D. N., Carroll, R. A., & Lindenberg, A. M. (2008). Detecting changes in simulated events using partial-interval recording and momentary time sampling. *Behavioral Interventions*, 23(4), 237-269. <https://doi.org/10.1002/bin.269>

- Rapp, J. T., Swanson, G., Sheridan, S. M., Enloe, K. A., Maltese, D., Sennott, L. A., ... Lanovaz, M. J. (2013). Immediate and subsequent effects of matched and unmatched stimuli on targeted vocal stereotypy and untargeted motor stereotypy. *Behavior Modification, 37*(4), 543-567. <https://doi.org/10.1177/0145445512461650>
- Rapp, J. T., & Vollmer, T. R. (2005). Stereotypy I: A review of behavioral assessment and treatment. *Research in Developmental Disabilities, 26*(6), 527-547. <https://doi.org/10.1016/j.ridd.2004.11.005>
- Rich, E., & Knight, K. (1991). *Artificial intelligence*. McGraw-Hill Education.
- Saylor, S., Sidener, T. M., Reeve, S. A., Fetherston, A., & Progar, P. R. (2012). Effects of three types of noncontingent auditory stimulation on vocal stereotypy in children with autism. *Journal of Applied Behavior Analysis, 45*(1), 185-190. <https://doi.org/10.1901/jaba.2012.45-185>
- Schmidt, M. G., Rapp, J. T., Novotny, M. A., & Lood, E. A. (2012). Detecting changes in non-simulated events using partial interval recording and momentary time sampling: Evaluating false positives, false negatives, and trending. *Behavior Interventions, 28*(1), 58-81. <https://doi.org/10.1002/bin.1354>
- Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation coefficients: appropriate use and interpretation. *Anesthesia & Analgesia, 126*(5), 1763-1768. <https://doi.org/10.1213/ANE.0000000000002864>
- Skinner, B. F. (1951). *Science and human behavior*. The Free Press.
- Spencer, V. G., & Alkhanji, R. (2018). Response interruption and redirection (RIRD) as a behavioral intervention for vocal stereotypy: A systematic review. *Education and*

Training in Autism and Developmental Disabilities, 53(1), 33-43. <https://doi.org/10.1901/jaba.2011.44-169>

Westeyn, T., Vadas, K., Xuehai B., Starner, T., & Abowd, G. D. (2005). Recognizing mimicked autistic self-stimulatory behaviors using HMMs. *Ninth IEEE International Symposium on Wearable Computers (ISWC'05)*, 164-169.

<https://doi.org/10.1109/ISWC.2005.45>

Weston, R., Hodges, A., & Davis, T. N. (2018). Differential reinforcement of other behaviors to treat challenging behaviors among children with autism: A systematic and quality review. *Behavior Modification*, 42(4), 584-609.

<https://doi.org/10.1177/0145445517743487>

Wong, T.-T. (2015). Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. *Pattern Recognition*, 48(9), 2839-2846.

<https://doi.org/10.1016/j.patcog.2015.03.009>

Table 1*Participant Characteristics*

Participants	Age	Gender	Description of vocal stereotypy
Emile	7	M	Grunting and unintelligible vocalizations
Matt	5	M	Monosyllable sounds and repetitive singing
Dave	6	M	Humming and unintelligible vocalizations
Billy-Peter	8	M	Monosyllable sounds and acontextual giggling
Owen	7	M	Phrase or word repetitions
Dan	11	M	Phrase or word repetitions
Alia	10	F	Humming and unintelligible vocalizations
Nate	6	M	Phrase or word repetitions

Table 2*Number of Sessions and Duration of Dataset Per Participant*

Participants	Number of Sessions	Total Time (s)	With Music (s)	Without Music (s)
Emile	38	27,448	719	26,729
Matt	6	4,015	2,044	1,971
Dave	30	20,756	3,685	17,071
Billy-Peter	10	6,909	2,729	4,180
Owen	25	17,461	4,357	13,104
Dan	11	7,533	4,750	2,783
Alia	10	7,091	2,783	4,308
Nate	12	8,351	4,820	3,531
Total	142	99,564	25,887	73,677

Table 3

Between-Participant Analyses: Accuracy, Kappa, and Correlation for All Sessions and Those Without Music Sessions.

Participants	All Sessions			Sessions Without Music Only		
	Accuracy	Kappa	Correlation	Accuracy	Kappa	Correlation
Emile	0.90	0.66	0.86	0.90	0.67	0.87
Matt	0.78	0.49	0.97	0.77	0.54	0.97
Dave	0.79	0.50	0.82	0.81	0.57	0.81
Billy-Peter	0.89	0.50	0.88	0.73	0.33	0.42
Owen	0.83	0.52	0.80	0.78	0.50	0.47
Dan	0.75	0.29	0.30	0.77	0.34	-0.90
Alia	0.79	0.30	-0.37	0.87	0.52	0.78
Nate	0.71	0.33	-0.12	0.79	0.57	0.88

Table 4*Within-Participant Analyses: Accuracy, Kappa, and Correlation for Each Participant*

Participants	Accuracy	Kappa	Correlation
Emile	0.94	0.75	0.97
Matt	0.80	0.43	0.96
Dave	0.83	0.60	0.66
Billy-Peter	0.91	0.25	0.93
Owen	0.86	0.40	0.88
Dan	0.79	0.23	0.34
Alia	0.91	0.67	0.58
Nate	0.74	0.34	0.33

Table 5*Hybrid Approach: Accuracy, Kappa, and Correlation for Each Participant*

Participants	Accuracy	Kappa	Correlation
Emile	0.95	0.74	0.97
Matt	0.78	0.41	0.98
Dave	0.83	0.57	0.84
Billy-Peter	0.91	0.23	0.87
Owen	0.85	0.45	0.88
Dan	0.83	0.24	0.20
Alia	0.92	0.60	0.79
Nate	0.73	0.31	0.08

Figure 1

Visual Representation of the Design of the Artificial Neural Network

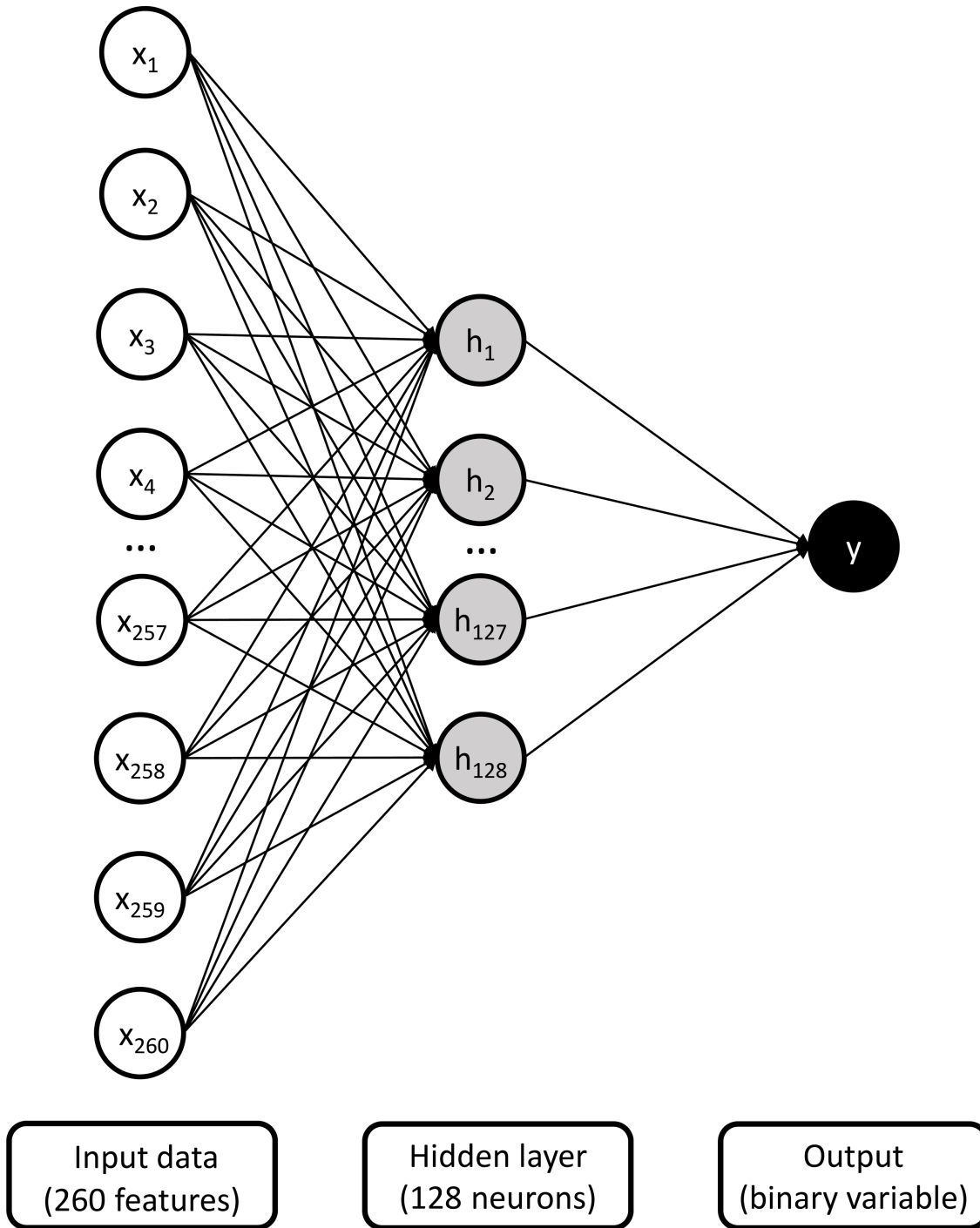


Figure 2

Between-Participant Analyses: Correlation Between the Percentages of Vocal Stereotypy Measured by the Machine Learning Algorithm and Those Measured by the Human Observer Across All Sessions for Each Participant

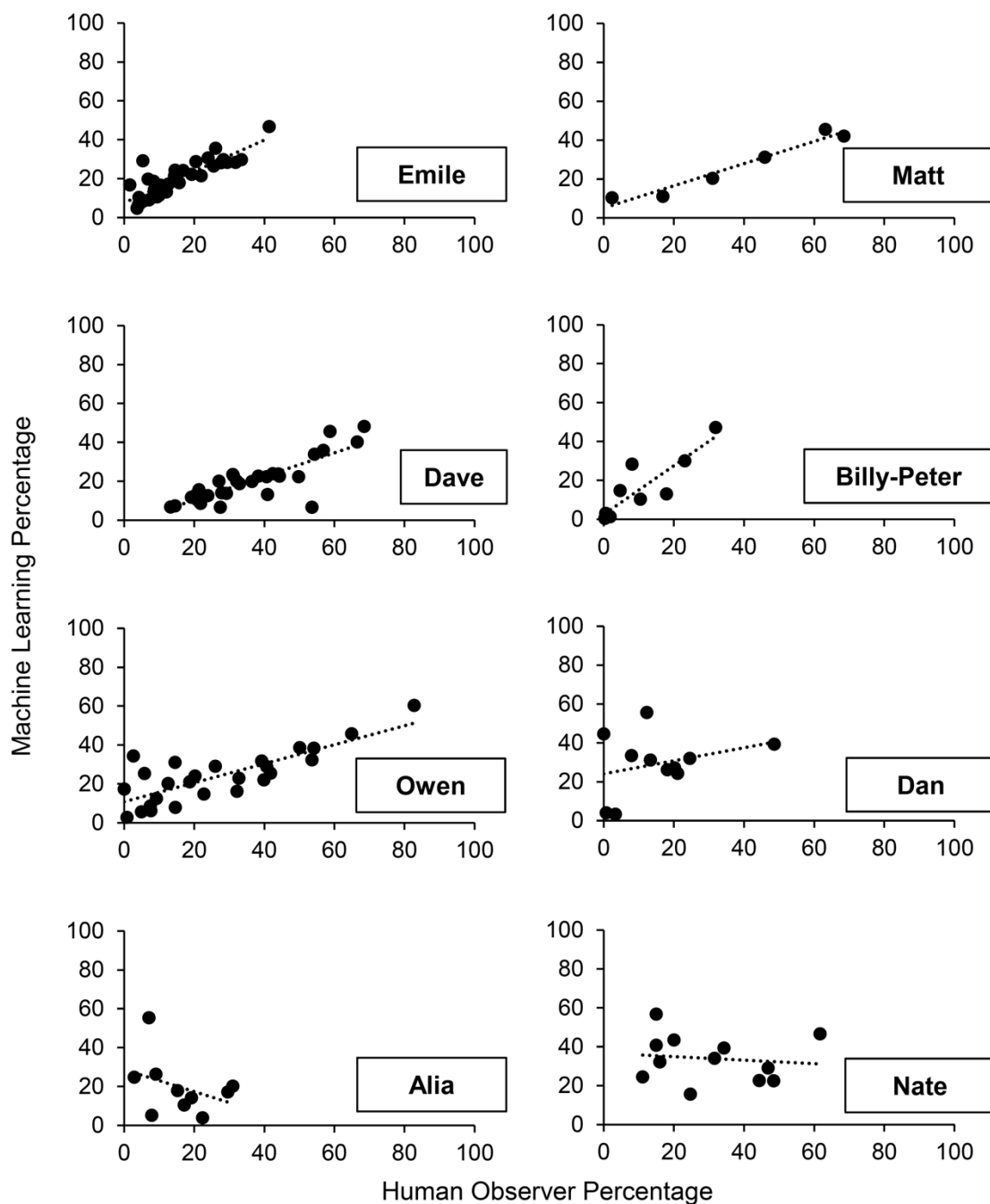


Figure 3

Between-Participant Analyses: Correlation Between the Percentages of Vocal Stereotypy Measured by the Machine Learning Algorithm and Those Measured by the Human Observer Across Sessions Without Music Only for Each Participant

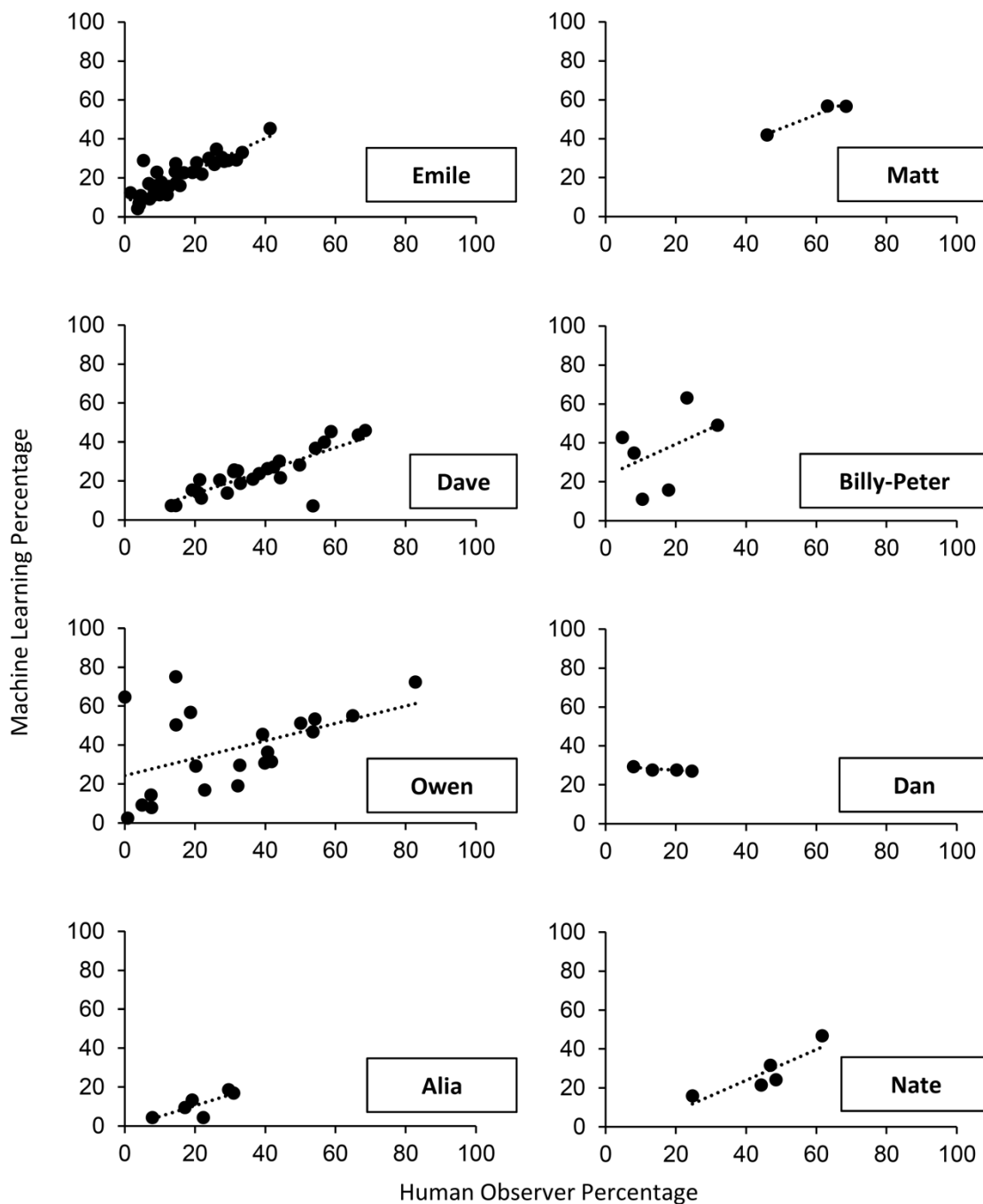


Figure 4

Within-Participant Analyses: Correlation Between the Percentages of Vocal Stereotypy Measured by the Machine Learning Algorithm and Those Measured by the Human Observer Across Sessions for Each Participant

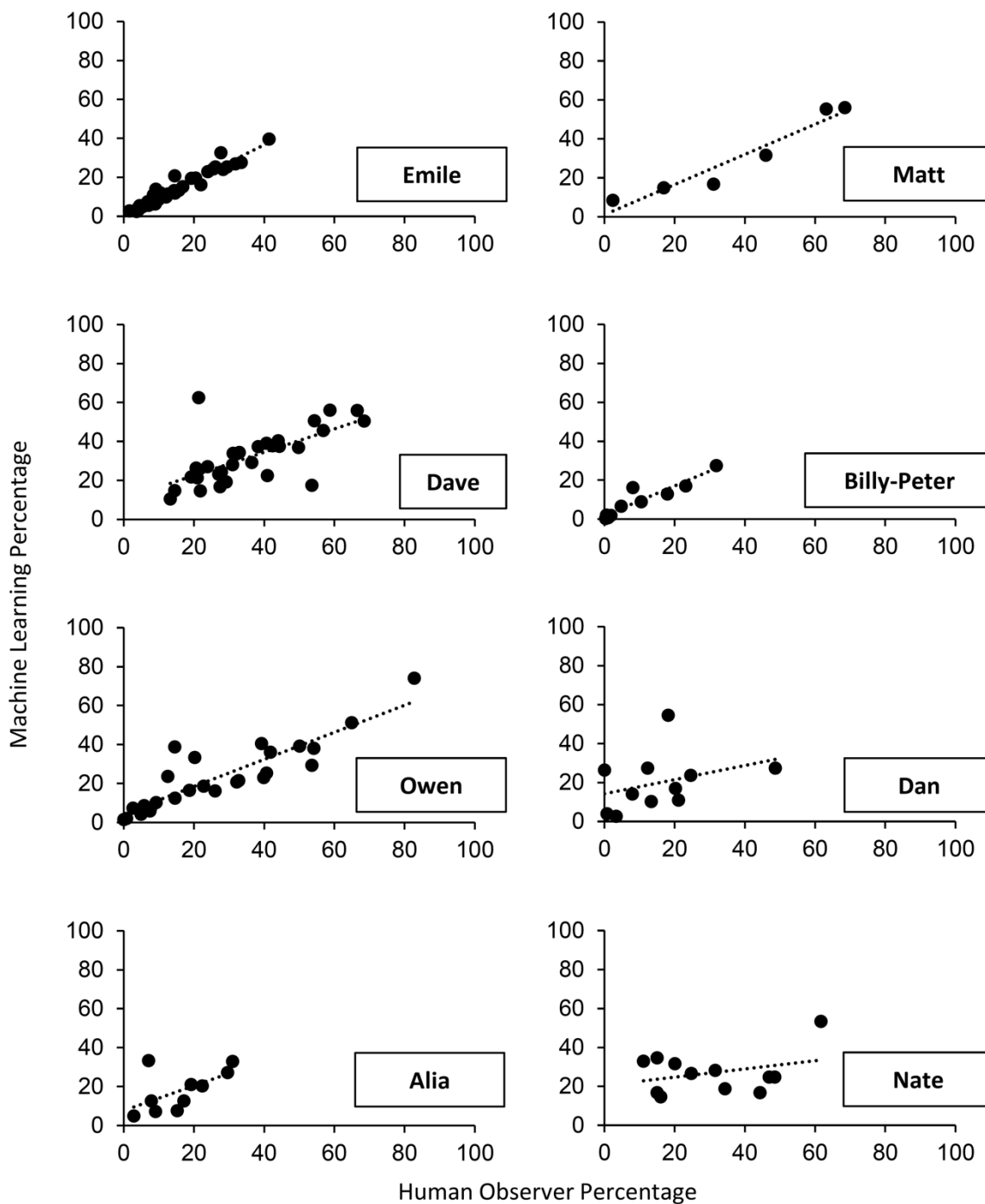


Figure 5

Hybrid Analyses: Correlation Between the Percentages of Vocal Stereotypy Measured by the Machine Learning Algorithm and Those Measured by the Human Observer Across Sessions for Each Participant

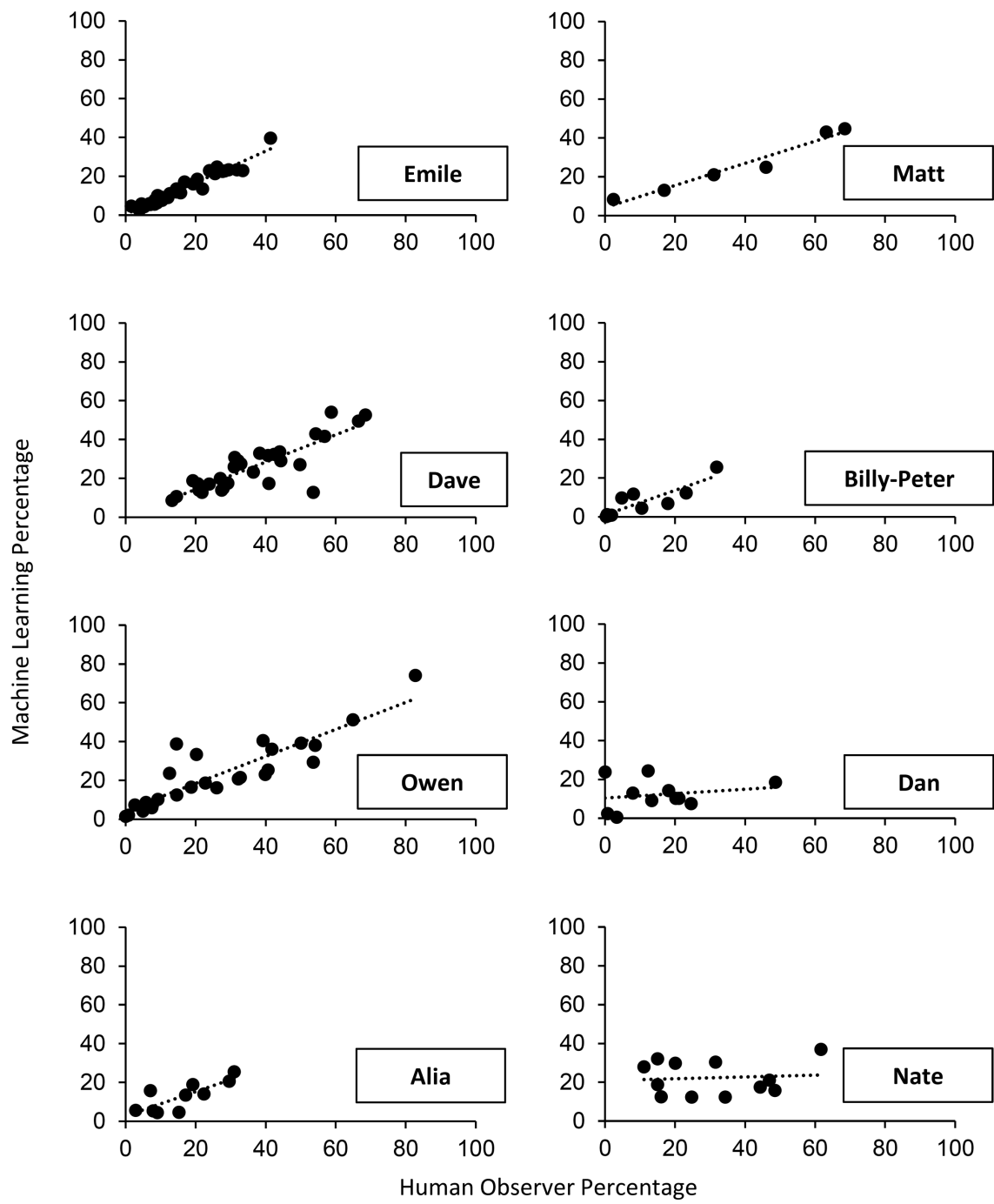
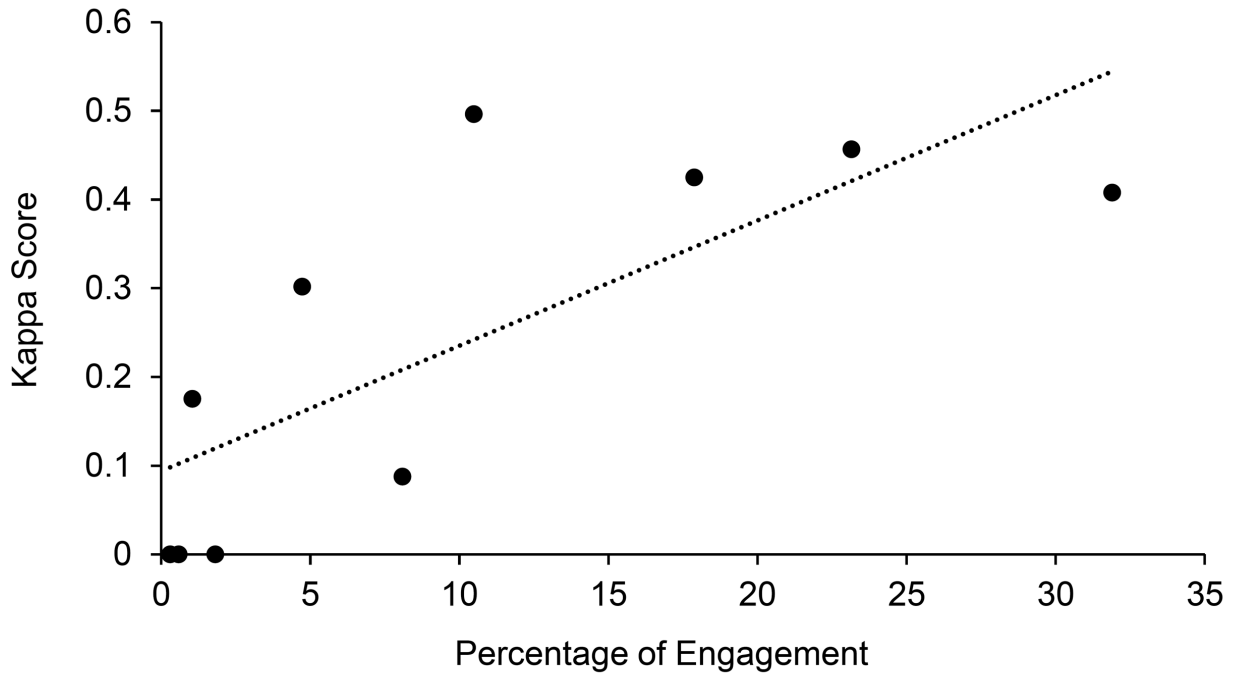


Figure 6

Correlation Between the Kappa Scores and the Percentage of Engagement in Each Session for Billy-Peter



Chapitre V : Discussion générale et conclusion

Résumé des principaux résultats empiriques

Les principaux objectifs de cette thèse doctorale étaient de : (a) déterminer l'efficacité d'une intervention comportementale pour augmenter la tolérance au port d'une ceinture cardiaque chez des enfants non verbaux ayant un TSA, (b) évaluer la relation entre le cortisol salivaire, le rythme cardiaque et des mesures d'observations directes de la stéréotypie chez des enfants minimalement verbaux ayant un TSA et (c) déterminer l'efficacité d'un algorithme d'intelligence artificielle dans l'identification de la stéréotypie vocale chez des enfants ayant un TSA.

Les résultats de mon premier article indiquent que l'intervention comportementale renforcement différentiel d'autres comportements (*differential reinforcement of other behavior*; DRO) était efficace pour augmenter la tolérance au port d'un dispositif cardiaque chez des enfants non verbaux ayant un TSA. Ces résultats mettent en lumière la pertinence des interventions comportementales pour amener ces enfants à tolérer le port de différents dispositifs médicaux (Cook et al., 2015; DeLeon et al., 2008; Richling et al., 2011).

Le deuxième article de cette thèse démontre, une fois de plus, la complexité de la relation entre le stress et la stéréotypie chez les enfants minimalement verbaux ayant un TSA. Notamment, les résultats montrent que le niveau moyen de cortisol est associé négativement à la stéréotypie globale et motrice. Ces résultats vont à contresens des résultats d'études précédentes qui ont identifié une corrélation positive entre les niveaux de cortisol et la stéréotypie (Bitsika et al., 2015; Lydon, Healy, Roche et al., 2015; Yang et al., 2015). Bien que le protocole de cette étude ne nous permette pas de nous positionner quant à la séquence causale du lien, nos résultats sont cohérents avec ceux de Gabriels et al. (2013) qui ont

déterminé que les enfants ayant un TSA qui présentent des niveaux élevés de stéréotypie, avaient des niveaux de cortisol plus faibles.

Dans cette même étude, le rythme cardiaque était associé positivement à la stéréotypie globale et motrice. Les données cardiaques sont à interpréter avec prudence, considérant qu'une élévation du niveau de stéréotypie peut entraîner une augmentation de l'activation physique de la personne qui à son tour, augmente le rythme cardiaque. Ceci dit, ces résultats tendent à soutenir l'hypothèse de l'autostimulation ou du renforcement automatique. Il est à noter qu'aucun lien n'a été trouvé entre le cortisol, le rythme cardiaque et la stéréotypie vocale. Ces résultats nous amènent à formuler l'hypothèse que différentes formes de stéréotypie pourraient répondre à des fonctions différentes chez ces enfants. Finalement, nous avons tenté de déterminer si les liens interindividuels observés suite à nos analyses statistiques étaient aussi observables au plan intra-individuel. L'examen visuel de nos figures semblait démontrer que les liens n'étaient pas présents au plan intraparticipants. Puisque nous n'avons pas assez de données pour mener une analyse inférentielle, des études futures sont nécessaires pour évaluer empiriquement la nature du lien entre ces variables à l'échelle intraindividus.

Les résultats du troisième article de la thèse montrent que les réseaux de neurones artificiels, en utilisant une approche hybride, ont produit des corrélations séance par séance de près de 0.80 pour six des huit participants. Au niveau interparticipant, nos modèles et l'observateur humain sont arrivés à des accords modérés à substantiels ($\kappa > 0.5$) pour cinq participants sur huit et ce, malgré la présence de musique en arrière-plan. Le retrait de la musique a considérablement augmenté la performance de nos algorithmes pour deux des participants, alors que l'effet inverse a été observé pour deux autres participants.

L'entraînement et l'évaluation des modèles se faisaient sur une quantité de données inférieure

pour les échantillons sans musique, ce qui peut potentiellement expliquer cet effet négatif. Finalement, nous avons appliqué nos modèles au niveau intraparticipants, ce qui veut dire que l'algorithme s'entraînait sur les données d'un participant à la fois et tentait de prédire la stéréotypie vocale de ce même participant. Étonnamment, la performance des algorithmes n'a pas augmenté au niveau moléculaire (c.-à-d., indice kappa), mais la performance du niveau molaire s'est améliorée (c.-à-d., séance par séance). Bien que la tâche qui consiste à prédire la stéréotypie vocale chez un même participant puisse sembler plus simple, la moindre quantité de données utilisées pour l'entraînement peut expliquer cette réduction de la performance au plan moléculaire.

Implications pour la recherche

Les trois articles qui constituent cette thèse ont permis de faire avancer les connaissances sur plusieurs plans. Dans un premier temps, la vaste majorité des études centrées sur le stress dans le passé incluaient des personnes ayant un TSA qui présentaient un plus haut niveau de fonctionnement. Comme mentionné dans l'introduction de cette thèse, des études montrent qu'entre 30% et 70% des enfants ayant un TSA présentent aussi une déficience intellectuelle, fréquemment associée à une absence ou à un retard significatif du langage (QI < 70; Bertrand et al., 2001; Chakrabarti & Fombonne, 2001, 2005; Fombonne, 2009). La grande hétérogénéité du TSA souligne l'importance de mener des recherches auprès des personnes qui se situent à tous les niveaux du continuum de la symptomatologie autistique. De ce fait, les résultats des deux premières études représentent une avancée dans le développement des connaissances du stress chez cette population, qui sont souvent sous-représentés dans les échantillons des études.

En reproduisant les données d'études précédentes (Cook et al., 2015; DeLeon et al., 2008; Richling et al., 2011) et en étendant ces résultats au port d'un dispositif cardiaque, le premier article de cette thèse a démontré qu'une intervention comportementale est efficace pour augmenter la tolérance au port d'une ceinture cardiaque chez des enfants non verbaux ayant un TSA. Ces résultats ont une grande implication sur le plan de la recherche. En effet, la majorité des méthodes de collectes de données physiologiques nécessite le port d'un dispositif (la conductance électrodermale, l'électrocardiogramme, l'électroencéphalogramme). Les sensibilités sensorielles fréquemment associées au TSA rendent ces personnes plus susceptibles de ne pas tolérer différents dispositifs de prises de mesures physiologiques. De plus, les personnes qui présentent un niveau de sévérité du TSA plus élevé sont plus susceptibles de présenter des sensibilités sensorielles plus importantes (Uljarević et al., 2017). Ainsi, cette caractéristique peut faire en sorte que les personnes qui ont un TSA et qui ont un niveau de fonctionnement plus bas soient représentées en moins grande proportion dans les projets de recherche qui incluent la prise de mesures physiologiques, ce qui limite le développement des connaissances chez cette population. Par ailleurs, puisque les résultats de cette étude sont une reproduction de résultats précédents, la validité externe s'en trouve augmentée. Conséquemment, il est possible d'émettre l'hypothèse que cette intervention possède un grand potentiel pour plusieurs dispositifs nécessaires à la collecte de données physiologiques. Ces résultats permettront donc d'outiller les chercheurs qui désirent mener des recherches qui incluent des variables physiologiques auprès des enfants non verbaux, et on l'espère contribuera à augmenter le volume de recherches menées auprès de cette population.

Le deuxième article de cet ouvrage est le premier, à notre connaissance, à évaluer le lien entre plusieurs variables physiologiques (cortisol salivaire et rythme cardiaque) et des

mesures d'observations directes de la stéréotypie. La discordance entre les résultats de la stéréotypie globale, motrice et vocale met de l'avant la possibilité que les différentes formes de comportements stéréotypés répondent à des besoins différents. Cette nouvelle information représente une avancée pour la recherche et devra être prise en considération pour les études futures.

Finalement, au-delà des défis associés spécifiquement à la prise de mesures physiologiques, cette thèse souligne les défis inhérents aux mesures comportementales de manière plus large et les résultats du troisième article introduisent une solution novatrice à cet enjeu. En effet, cette étude montre le potentiel des algorithmes d'intelligence artificielle dans la prise de mesures de la stéréotypie vocale chez les enfants ayant un TSA. Bien que d'autres recherches doivent être menées avant de pouvoir rendre disponibles ces algorithmes aux chercheurs, le potentiel pour la recherche est immense. À l'aide des recherches futures, on peut imaginer un système qui pourrait mesurer automatiquement la stéréotypie vocale. L'utilisation de ce système permettrait de réduire considérablement les ressources nécessaires (autant humaines que monétaires) pour mener des projets de recherche sur ce sujet, notamment lors de l'étude du stress chez les enfants minimalement verbaux. Ainsi, un volume beaucoup plus grand de recherches pourrait être menées, ce qui aurait un impact direct sur la croissance de la connaissance dans le domaine de la stéréotypie vocale chez les personnes ayant un TSA.

Implications pour la pratique

Les résultats du premier article ont des répercussions directes au plan clinique. Les sensibilités sensorielles sont fréquentes chez les personnes ayant un TSA (Grapel et al., 2015) et représentent donc un enjeu avec lequel les intervenants doivent composer dans leur pratique. Le bien-être et les besoins des enfants étant toujours au centre des interventions, la

première chose à mettre en place est l'aménagement de l'environnement afin de rendre celui-ci le plus confortable possible pour eux. Tamiser les lumières, aménager un endroit calme où l'enfant peut se retirer au besoin, mettre des coquilles à sa disposition pour réduire le bruit sont tous des moyens qui peuvent être mis en place pour augmenter l'adéquation entre les sensibilités sensorielles des enfants et leur environnement. Par contre, certaines situations nécessitent d'amener les enfants à augmenter leur tolérance face à l'élément qui les rend inconfortables (p. ex., le port de lunettes de vision, de prothèses auditives, d'un bracelet médical, de dispositifs orthodontiques). De ce fait, il importe de trouver des méthodes efficaces pour augmenter cette tolérance. Les résultats de cette étude ont montré des résultats rapides (entre 13 et 25 séances de cinq minutes), ce qui fait de cette intervention une option pertinente et envisageable pour les milieux cliniques.

Les résultats du deuxième article s'insèrent dans un contexte où les connaissances sont contradictoires dans la littérature scientifique. Or, ces résultats représentent un pas vers la clarification du lien entre le stress et les comportements stéréotypés chez les enfants minimalement verbaux ayant un TSA. La portée clinique de ces résultats est grande puisque la fonction des comportements stéréotypés est au centre des interventions qui sont mises en place à ce jour. À l'heure actuelle, plusieurs interventions comportementales dérivées de l'analyse appliquée du comportement (AAC) ont démontré leur efficacité dans la réduction des comportements stéréotypés (Lanovaz et al., 2011; Spencer & Alkhanji, 2018). La plupart des interventions comportementales sont fonctionnelles, ce qui veut dire qu'elles se basent sur la fonction du comportement. Ainsi, deux comportements qui n'ont pas la même fonction ne seront pas traités de la même manière. Il importe donc de poursuivre les recherches quant à

l'identification de la fonction de la stéréotypie chez ces enfants, ce qui pourrait avoir un impact direct sur les interventions mises en place.

Le troisième article visait à évaluer l'efficacité d'un algorithme d'intelligence artificielle dans la reconnaissance de la stéréotypie vocale chez des enfants ayant un TSA. Dans le contexte des interventions psychosociales, l'évaluation de l'efficacité des interventions est un concept central. L'évaluation est pertinente et essentielle à deux moments précis : (1) avant la mise en place de l'intervention et (2) pendant l'implantation de l'intervention. L'évaluation préintervention est essentielle pour établir le niveau de base du comportement, c'est donc cette mesure qui permettra de juger des effets de l'intervention lors des évaluations subséquentes. L'évaluation en cours d'intervention est primordiale afin d'observer rapidement les effets de l'intervention et particulièrement l'absence d'effet ou la présence d'effets négatifs. C'est cette évaluation qui permet de revoir et redéfinir rapidement les stratégies d'intervention lorsque les effets escomptés ne sont pas observés. De par leur nature, certains comportements sont plus difficiles à mesurer de manière rigoureuse que d'autres et la stéréotypie vocale en fait partie. Des données montrent que les comportements stéréotypés peuvent interférer avec le processus d'apprentissage (Lanovaz et al., 2013) et peuvent avoir des conséquences négatives sur l'intégration sociale de ces personnes (Cunningham & Schreibman, 2008). Pour ces raisons, ces comportements font fréquemment l'objet d'interventions afin de diminuer leur fréquence. Le développement d'un système qui pourrait mesurer automatiquement la stéréotypie vocale rendrait possible une évaluation rigoureuse des interventions qui ciblent ces comportements ce qui est difficile, voire impossible à faire présentement dans la majorité des contextes cliniques.

Implications pour la psychoéducation

Les psychoéducateurs sont très impliqués auprès de la clientèle ayant un TSA et leur rôle est multiple selon leur milieu de pratique. Les psychoéducateurs sont, entre autres, en charge de superviser les programmes dérivés de l'approche comportementale, d'évaluer les capacités adaptatives de cette clientèle ainsi que d'élaborer et de mettre en place des plans d'intervention. En contexte d'intervention, les psychoéducateurs sont souvent confrontés aux conséquences de la sensibilité sensorielle des enfants qui ont un TSA. Lorsque ces enfants présentent plusieurs problématiques qui semblent découler directement du profil sensoriel, une demande d'évaluation en ergothérapie est remplie, mais le volume élevé de demandes entraîne des délais dans l'accès à ces services. L'évaluation de l'efficacité de l'intervention comportementale qui a fait l'objet du premier article de cette thèse ouvre la possibilité aux psychoéducateurs de pouvoir intervenir sur certaines problématiques sensorielles, lors de la période d'attente de services en ergothérapie.

Lors de l'adoption de la loi 21 en 2012, le champ de pratique des psychoéducateurs a été redéfini et sept activités professionnelles leur ont été réservées (Ordre des Psychoéducateurs et Psychoéducatrices du Québec [OPPQ], 2020). L'évaluation des difficultés d'adaptation et des capacités adaptatives sont au cœur des actes réservés aux psychoéducateurs. La notion d'adaptation est centrale en psychoéducation et s'inscrit dans une perspective écosystémique qui implique la notion d'interactionnisme. Selon Gendreau (2001), l'adaptation est le produit de la rencontre entre le potentiel d'adaptation de la personne (PAD) et le potentiel expérientiel (PEX) qu'offre son environnement. Le PEX est un levier d'intervention riche puisque c'est un élément sur lequel le psychoéducateur peut agir pour ainsi présenter des défis adaptatifs adaptés à sa clientèle. Le psychoéducateur doit alors évaluer le niveau de convenance, qui correspond à l'écart entre le PAD et le PEX (OPPQ,

2010). L'objectif est alors d'être face à un écart qui présente un défi pour le sujet, mais qui tient aussi compte de ses forces et de ses besoins. La personne se trouve alors en état de déséquilibre adaptatif qui l'amène à se mettre en action et à acquérir de nouveaux apprentissages (OPPQ, 2010).

Une étape préalable essentielle afin de définir un niveau de convenance qui place le sujet en déséquilibre adaptatif, est la compréhension réelle des comportements auprès desquels on intervient. Les résultats du deuxième article de cet ouvrage visent à faire avancer les connaissances sur la valeur adaptative des comportements stéréotypés chez les enfants minimalement verbaux qui ont un TSA. Les résultats actuels étant contradictoires, il n'est pas possible de savoir avec certitude à quel besoin répondent ces comportements. En se basant sur l'approche comportementale, il est possible d'affirmer que si ces comportements ne répondaient à aucune fonction, ils auraient été portés à disparaître au fil des années (Skinner, 1951). Bien que plusieurs autres recherches doivent être menées pour clarifier cet aspect, les comportements stéréotypés possèdent une valeur adaptative qu'on ne comprend pas entièrement à ce jour. Comme mentionné précédemment, les psychoéducateurs sont très impliqués auprès de la clientèle ayant un TSA et la stéréotypie est centrale au diagnostic. De plus, ces comportements ont été démontrés comme ayant des impacts négatifs sur le fonctionnement social des personnes ayant un TSA (Cunningham & Schreibman, 2008; Lanovaz et al., 2013), ils peuvent alors être considérés comme un facteur de risque à l'intégration sociale et à l'acquisition de nouvelles habiletés. En ayant une compréhension élargie de leur mécanisme d'action, les psychoéducateurs pourront déterminer avec justesse le niveau de convenance entre le PAD et le PEX de ces enfants et donc de favoriser leur fonctionnement adaptatif optimal.

Huit opérations professionnelles sont propres à une pratique psychoéducative rigoureuse (Gendreau, 2001) et les résultats du troisième article de la thèse contribuent significativement à trois d'entre elles : l'observation, l'évaluation préintervention et l'évaluation postsituationnelle. L'observation est la première opération professionnelle et sert de base à l'évaluation et à la planification subséquente (Gendreau, 2001). L'observation, telle qu'originellement définie, s'actualise dans les moments de vécu éducatif partagé. Pendant le déroulement des interactions, le psychoéducateur enregistre les attitudes et les comportements significatifs du sujet. L'évaluation, pour sa part, est conçue par Gendreau (2001) en trois temps : l'évaluation en cours d'intervention, l'évaluation en retrait et l'évaluation postsituationnelle. L'évaluation en cours d'intervention « consiste à donner immédiatement un sens à la situation que l'on vient d'observer » (Gendreau, 2001). L'évaluation en retrait, pour sa part, se définit par le moment que le psychoéducateur réserve pour analyser l'ensemble des informations recueillies. Le rôle de ces deux étapes est fondamental puisque c'est en accomplissant ces évaluations avec rigueur que le psychoéducateur élabore les hypothèses sur lesquelles seront basées les interventions proposées. Finalement, l'évaluation postsituationnelle et plus particulièrement l'évaluation sommative consiste à évaluer l'efficacité de l'intervention à l'aide de mesures objectives.

L'objectivation de la mesure de la stéréotypie s'insère dans une perspective plus large d'augmentation de la rigueur de la pratique psychoéducative. Depuis plusieurs années, des efforts sont faits pour entraîner les psychoéducateurs à adopter une approche basée sur les données probantes (Larivée, 2014). Dans cette approche, le choix des interventions doit être basé sur des données scientifiques qui démontrent leur efficacité. Au-delà de la sélection de des interventions à mettre en oeuvre, l'ensemble des décisions cliniques doit être basé sur des

données rigoureuses. Pour ce faire, le professionnel doit être outillé pour mesurer le comportement d'intérêt afin d'évaluer l'efficacité de l'intervention mise en place. Les résultats du troisième article est un pas vers l'automatisation de la prise de données et vers une solution potentielle qui ne requiert pas de ressources humaines supplémentaires. La reconnaissance automatique des comportements stéréotypés permettrait donc d'enrichir considérablement les observations du psychoéducateur. En effet, pendant qu'il est en situation de vécu éducatif partagé et centre son attention sur des éléments plus subjectifs, comme les attitudes et les agissements de l'enfant de manière générale, des mesures rigoureuses et systématiques de la stéréotypie seraient enregistrées. Cette reconnaissance automatique pourrait contribuer à faire avancer les possibilités de mesures et de suivis d'interventions qui ciblent la stéréotypie. Puisque les étapes d'observation et d'évaluation représentent la base de l'ensemble des opérations professionnelles subséquentes du psychoéducateur, l'augmentation de la rigueur de ce processus aura un effet direct sur la qualité de sa pratique de manière plus générale.

Forces et limites de l'étude doctorale

Le recrutement

Le recrutement représente fréquemment un défi lors de l'implantation de projets de recherche, mais la nature de nos mesures physiologiques et les caractéristiques de l'échantillon recherché (enfants TSA minimalement verbaux) ont rendu cette étape d'autant plus difficile pour les travaux de cette thèse. Bien que plusieurs familles aient manifesté leur intérêt pour participer à la deuxième étude de la thèse, très peu d'enfants ont toléré à la fois la ceinture cardiaque et les mesures de cortisol de la deuxième étude. Nous espérions recruter six enfants, mais les contraintes n'ont permis que d'en recruter quatre.

L'échantillon

L'échantillon des deux premières études de cette thèse est constitué d'enfants ayant un TSA minimalement verbaux, ce qui représente une force puisque cette catégorie d'enfants ayant un TSA est fréquemment sous-représentée dans les études. Par contre, la taille de l'échantillon du deuxième article est une limite puisqu'elle nous a restreints au plan des analyses, ainsi que dans les inférences qu'il a été possible de tirer par rapport aux résultats.

Les protocoles expérimentaux

L'utilisation d'un devis à niveau de base multiple dans la première étude représente une force puisque le recours à un tel devis expérimental permet d'évaluer de manière fiable les changements produits par une intervention et son efficacité (Lanovaz, 2013). En revanche, le devis du deuxième article ne nous a pas permis de déterminer la séquence causale du lien entre les mesures physiologiques et les comportements stéréotypés, ce qui a limité les conclusions qui ont pu être tirées à la suite des analyses. Les études futures devront prendre en considération l'inclusion d'une notion de temporalité dans la conception du devis afin d'être en mesure de tirer des conclusions plus détaillées sur la relation entre les variables.

Intersectorialité

Les trois articles de cette thèse contiennent des mesures et des techniques qui proviennent de secteurs variés et leur combinaison représente une force propre à cet ouvrage dans le domaine de la psychoéducation. En effet, la thèse combine l'évaluation d'une intervention clinique à des mesures biologiques et à des algorithmes d'intelligence artificielle. Ce sont des mesures qui sont plus rarement utilisées dans le domaine psychoéducatif et qui ajoutent un caractère prometteur et novateur à cette thèse.

Futures études

Les résultats de cette thèse ouvrent le chemin aux études futures pour le développement des connaissances dans ce domaine. Il serait pertinent de mener des études sur le stress et les comportements stéréotypés en concevant un devis qui inclut une notion de temporalité, ce qui permettrait de déterminer la directionnalité du lien évalué. Un échantillon de plus grande taille ouvrirait la possibilité de mener des analyses statistiques plus complexes. Des analyses multiniveaux seraient pertinentes afin de déterminer quelle part des résultats obtenus est attribuable à la variance interindividus versus la variance intra-individus. Considérant l'hétérogénéité de ce trouble, on peut penser que la relation entre le stress et la stéréotypie au plan intra-individu varie d'une personne à l'autre. Faire avancer les connaissances sur le plan intra-individuel permettrait possiblement d'amener des pistes explicatives aux résultats contradictoires trouvés dans le passé. Les études futures qui s'intéresseront au stress auprès de la population ayant un TSA devraient aussi considérer avoir recours à des dispositifs qui sont plus faciles à tolérer pour cette population, comme des vêtements intelligents, qui mesurent plusieurs variables physiologiques simultanément.

Compte tenu de son potentiel énorme dans l'amélioration de la prise de mesures, l'intelligence artificielle devrait être au cœur des recherches futures. Dans un premier temps, il serait pertinent de reprendre l'évaluation d'algorithmes d'intelligence artificielle dans la reconnaissance de la stéréotypie vocale : (1) en testant les modèles avec un plus grand nombre de données, (2) en évaluant la performance de différents types d'algorithmes et (3) en évaluant l'effet de la modification des hyperparamètres à l'intérieur de ces algorithmes. De plus, lorsque la performance de l'algorithme de reconnaissance de la stéréotypie vocale sera optimisée, il serait intéressant d'explorer la possibilité d'appliquer des algorithmes d'AI à la reconnaissance d'autres comportements autistiques, comme la stéréotypie motrice.

Pour terminer, il est essentiel d'allier les forces et les connaissances spécifiques de différentes professions afin d'optimiser le développement des connaissances dans le domaine de l'autisme. Le caractère intersectoriel de cette thèse ouvre la voie à des projets de recherche interdisciplinaires. Nous vivons présentement dans une ère de changement au plan de l'implication du numérique et de la technologie dans toutes les sphères de nos vies et l'intervention psychosociale n'en fait pas exception. En plus de continuer de développer des algorithmes d'AI pour automatiser la prise de mesure, des efforts de recherche devront être mis dans l'évaluation de l'efficacité de solutions numériques pour évaluer et intervenir en autisme, mais aussi en santé mentale de manière plus large. Pour ce faire, des efforts de collaboration interdisciplinaire devront être fournis pour optimiser le développement des connaissances.

Conclusion

La présente thèse visait à répondre à trois objectifs soit : (a) déterminer l'efficacité d'une intervention comportementale pour augmenter la tolérance au port d'une ceinture cardiaque chez des enfants non verbaux ayant un TSA, (b) évaluer la relation entre le cortisol salivaire, le rythme cardiaque et des mesures d'observations directes de la stéréotypie chez des enfants minimalement verbaux ayant un TSA et (c) déterminer l'efficacité d'un algorithme d'intelligence artificielle dans l'identification de la stéréotypie vocale chez des enfants ayant un TSA. Les résultats de cette thèse contribuent à outiller les psychoéducateurs qui interviennent auprès des enfants TSA présentant des difficultés liées à des sensibilités sensorielles. De plus, cet ouvrage contribue à faire avancer les connaissances sur le stress des enfants minimalement verbaux ayant un TSA, qui étaient jusqu'à présent sous-représentés dans les recherches de ce domaine. Finalement, la présente thèse ouvre une voie peu explorée

auparavant dans le domaine psychoéducatif qui permettrait d'améliorer significativement les opérations professionnelles des psychoéducateurs qui touchent l'évaluation objective des comportements. Dans les prochaines années, il sera intéressant de suivre de près l'évolution et l'implication des nouvelles technologies dans la mesure objective des comportements et les retombées positives qu'elles auront sur la rigueur de l'évaluation psychoéducative.

Références

- Bertrand, J., Mars, A., Boyle, C., Bove, F., Yeargin-Allsopp, M., & Decoufle, P. (2001). Prevalence of autism in a United States population: The brick township, New Jersey, investigation. *Pediatrics*, *108*(5), 1155-1161. <https://doi.org/10.1542/peds.108.5.1155>
- Bitsika, V., Sharpley, C. F., Agnew, L. L., & Andronicos, N. M. (2015). Age-related differences in the association between stereotypic behaviour and salivary cortisol in young males with an Autism Spectrum Disorder. *Physiology & Behavior*, *152*, 238-243. <https://doi.org/10.1016/j.physbeh.2015.10.010>
- Chakrabarti, S., & Fombonne, E. (2001). Pervasive developmental disorders in preschool children. *Journal of the American Medical Association*, *285*, 3093-3099. <https://doi.org/10.1001/jama.285.24.3093>
- Chakrabarti, S., & Fombonne, E. (2005). Pervasive developmental disorders in preschool children: Confirmation of high prevalence. *American Journal of Psychiatry*, *162*, 1133-1141. <https://doi.org/10.1176/appi.ajp.162.6.1133>
- Cook, J. L., Rapp, J. T., & Schulze, K. A. (2015). Differential negative reinforcement of other behavior to increase wearing of a medical bracelet. *Journal of Applied Behavior Analysis*, *48*(4), 901-906. <https://doi.org/10.1002/jaba.228>
- Cunningham, A. B. et 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>
- DeLeon, I. G., Hagopian, L. P., Rodriguez-Catter, V., Bowman, L. G., Long, E. S., & Boelter, E. W. (2008). Increasing wearing of prescription glasses in individuals with mental retardation. *Journal of Applied Behavior Analysis*, *41*(1), 137-142.

- <https://doi.org/10.1901/jaba.2008.41-137>
- Fombonne, E. (2009). Epidemiology of pervasive developmental disorders. *Pediatric Research*, 65(6), 591-598. <https://doi.org/10.1203/PDR.0b013e31819e7203>
- Gabriels, R. L., Agnew, J. A., Pan, Z., Holt, K. D., Reynolds, A., & Laudenslager, M. L. (2013). Elevated repetitive behaviors are associated with lower diurnal salivary cortisol levels in autism spectrum disorder. *Biological Psychology*, 93(2), 262-268. <https://doi.org/10.1016/j.biopsycho.2013.02.017>
- Gendreau, G. (2001). *Jeunes en difficulté et intervention psychoéducative*. Béliveau éditeur.
- Grapel, J. N., Cicchetti, D. V., & Volkmar, F. R. (2015). Sensory features as diagnostic criteria for autism: Sensory features in autism. *Yale Journal of Biology and Medicine*, 88(1), 69-71.
- Lanovaz, M. J. (2013). L'utilisation de devis expérimentaux à cas unique en psychoéducation. *Revue de Psychoéducation*, 42(1), 161-183. <https://doi.org/10.7202/1061728ar>
- Lanovaz, M. J., Robertson, K., Soerono, K., & Watkins, N. (2013). Effects of reducing stereotypy on other behaviors: A systematic review. *Research in Autism Spectrum Disorders*, 7(10), 1234-1243. <https://doi.org/10.1016/j.rasd.2013.07.009>
- Lanovaz, M. J., Sladeczek, I. E., & Rapp, J. T. (2011). Effects of music on vocal stereotypy in children with autism. *Journal of Applied Behavior Analysis*, 44(3), 647-651. <https://doi.org/10.1901/jaba.2011.44-647>
- Larivée, S. (2014). Vers l'utilisation de données probantes. *La Pratique en Mouvement*, 7, 7-8. <https://www.ordrepsed.qc.ca/~~/media/pdf/Publication/MLapratiqueenmouvementno7.ashx?la=fr>

Lydon, S., Healy, O., Roche, M., Henry, R., Mulhern, T., & Hughes, B. M. (2015). Salivary cortisol levels and challenging behavior in children with autism spectrum disorder.

Research in Autism Spectrum Disorders, 10, 78-92.

<https://doi.org/10.1016/j.rasd.2014.10.020>

Ordre des Psychoéducateurs et Psychoéducatrices du Québec. (2010). *L'évaluation*

psychoéducative de la personne en difficulté d'adaptation : Lignes directrices. Repéré

à :

http://www.ordrepsed.qc.ca/~media/pdf/Publication/Evaluation_psychoeducative_2014.ashx?la=fr

Ordre des Psychoéducateurs et Psychoéducatrices du Québec. (2020). *Activités réservées au*

psychoéducateur. Repéré à : [https://www.ordrepsed.qc.ca/fr/grand-public/activites-](https://www.ordrepsed.qc.ca/fr/grand-public/activites-reservees-au-psychoeducateur/)

[reservees-au-psychoeducateur/](https://www.ordrepsed.qc.ca/fr/grand-public/activites-reservees-au-psychoeducateur/)

Richling, S. M., Rapp, J. T., Carroll, R. A., Smith, J. N., Nystedt, A., & Siewert, B. (2011).

Using noncontingent reinforcement to increase compliance with wearing prescription prostheses. *Journal of Applied Behavior Analysis, 44*(2), 375-379.

<https://doi.org/10.1901/jaba.2011.44-375>

Skinner, B. F. (1951). *Science and human behavior*. The Free Press.

Spencer, V. G., & Alkhanji, R. (2018). Response interruption and redirection (RIRD) as a

behavioral intervention for vocal stereotypy: A systematic review. *Education and*

Training in Autism and Developmental Disabilities, 53(1), 33-43.

Yang, C.-J., Tan, H.-P., Yang, F.-Y., Wang, H.-P., Liu, C.-L., He, H.-Z., Sang, B., Zhu, X.-

M., & Du, Y.-J. (2015). The cortisol, serotonin and oxytocin are associated with

repetitive behavior in autism spectrum disorder. *Research in Autism Spectrum Disorders*, 18, 12-20. <https://doi.org/10.1016/j.rasd.2015.07.002>

Uljarević, M., Baranek, G., Vivanti, G., Hedley, D., Hudry, K., & Lane, K. (2017).

Heterogeneity of sensory features in autism spectrum disorder: Challenges and perspectives for future research. *Autism Research*, 10(5), 703-710.

<https://doi.org/10.1002/aur.1747>