

Université de Montréal

Dépôt final

**Mesures des préférences sexuelles envers les enfants : intégration des stimuli auditifs et
virtuels**

par

Shawn Marschall-Lévesque

Département de Psychologie

Faculté des arts et des sciences

Thèse présentée

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

en psychologie recherche et intervention

option psychologie clinique

mai 2017

© Shawn Marschall-Lévesque, 2017

Résumé

La présence d'intérêts sexuels envers des enfants est l'un des facteurs principaux lors du diagnostic de trouble pédophilique et de la prédiction du risque de la récidive sexuelle. Néanmoins, l'évaluation objective de cet intérêt subjectif demeure difficile et nécessite l'utilisation de différents outils dont le plus rigoureux est la pléthysmographie pénienne (PPG). Cet outil, qui mesure les réponses érectiles des participants, permet l'évaluation de plusieurs types d'intérêts sexuels (déviants ou non). Pour ce faire, la PPG dépend des stimuli utilisés (*p. ex.*, visuels ou auditifs), ce qui constitue à la fois une force et une faiblesse, car certains des stimuli actuellement disponibles génèrent régulièrement des résultats invalides. De plus, des normes éthiques plus rigoureuses encadrant l'utilisation d'images réelles d'enfants limitent l'usage de certains stimuli. Par ce projet doctoral, dans le cadre de deux études, nous visons à diminuer le nombre élevé de profils invalides en attaquant la problématique sur deux fronts.

Dans la première étude, des stimuli auditifs standards largement utilisés en PPG seront agencés à des images de personnages générés par ordinateur dans le but de créer une modalité audiovisuelle capable de produire de plus grandes amplitudes péniennes que les stimuli auditifs ou visuels présentés seuls. L'objectif est de réduire le nombre de profils invalides et permettre une discrimination efficace des intérêts sexuels. Pour évaluer leur efficacité, les réponses érectiles de deux groupes d'adultes (*c.-à-d.* un groupe d'agresseurs sexuels d'enfants et un groupe de comparaison) ont été mesurées et comparées durant la présentation de trois blocs de stimuli (auditifs, visuels et audiovisuels).

Dans la deuxième étude, ces nouveaux stimuli visuels et audiovisuels seront utilisés lors de l'enregistrement des données pupillométriques et pléthysmographiques (PPG) dans le but d'obtenir une seconde mesure objective capable de discriminer la présence d'intérêts sexuels. Pour ce faire, les résultats pupillométriques des participants ont été enregistrés et comparés à leurs résultats pléthysmographiques.

Les résultats de la première étude soutiennent l'utilisation des stimuli audiovisuels en faisant usage de personnages générés par ordinateur pour produire de plus fortes réponses péniennes et ainsi augmenter le nombre de profils valides obtenus lors de l'évaluation PPG. De plus, les résultats démontrent que les stimuli audiovisuels sont équivalents aux stimuli auditifs pour classer les participants dans leurs groupes respectifs. Les résultats de la seconde étude démontrent les impacts positifs de l'utilisation de la pupillométrie en complémentarité aux résultats de la PPG. Il est alors prouvé qu'ensemble ces deux mesures sont capables de catégoriser efficacement les participants des deux groupes tout en montrant que les stimuli audiovisuels utilisés sont une alternative éthique et bénéfique aux stimuli visuels conventionnels (photographies d'enfants). Ces deux études renforcent l'importance de l'utilisation de nouvelles technologies dans le but d'améliorer les outils déjà à notre disposition, et ce, dans le respect de normes éthiques rigoureuses.

Mots-clés : pléthysmographie pénienne; pupillométrie; agresseurs d'enfants; intérêts sexuels déviants; stimuli virtuels.

Abstract

The presence of sexual interests towards children remains one of the foremost factors predicting the risk of sexual recidivism and one of the diagnosis criteria for pedophilic disorder. Nevertheless, assessing these subjective interests objectively remains difficult and requires the use of objective tools including penile plethysmography (PPG). This tool uses erectile responses to assess a variety of sexual interests (deviant or not). To accomplish this, PPG depends on the ability of stimuli (*e.g.*, visual or audio) to generate erectile responses. This dependence constitutes one of PPG's strengths and weaknesses, since rigorous ethical norms are applied to the use of real images of children and therefore, some of the stimuli currently used, have difficulty in generating valid results caused by a significant rates of invalid PPG profiles. This doctoral research is an attempt to improve the number of valid profiles by attacking this problematic on two fronts.

In the first study, in order to reduce the number of invalid profile, we combined standard audio stimuli with images of computer-generated characters in order to create a new set of audiovisual stimuli capable of producing stronger penile responses than both the audio or visual stimuli presented individually. Erectile responses from two groups of participants (sexual offenders against children and non-offenders) were recorded and compared, while they were presented with three blocks of stimuli (audio, visual and audiovisual). For the second study, we attempted to show that these new audiovisual stimuli could be used in a conjuncture with pupillometry to obtain a second objective measure of sexual interests. This was done by recording and comparing pupillometric and plethysmographic data PPG from the same groups of participants while they were presented with the audiovisual stimuli.

Results from the first study support the efficacy of the audiovisual stimuli using computer generated characters to generate stronger penile responses than both other modalities, while reducing the number of invalid PPG profiles when compared to the original audio stimuli. Furthermore, results demonstrate that the new audiovisual stimuli are as good as the standard audio stimuli at categorizing participants in their appropriate group. Results for the second study support the use of pupillometry as a complementary measure to PPG with both measures capable of adequately categorizing participants in their appropriate groups. Results also suggest that the audiovisual stimuli using computer generated characters are good and ethical alternative to conventional visual stimuli. These two studies also emphasise the importance of new technologies as a means of enhancing tools already known, whilst maintaining rigorous ethical norms.

Keywords: Penile Plethysmography; Pupillometry; Child Molesters; Deviant Sexual Interests; Virtual Stimuli.

Table des matières

RÉSUMÉ	III
ABSTRACT	V
TABLE DES MATIÈRES	VII
LISTE DES TABLEAUX	X
LISTE DES FIGURES	XI
LISTE DES ABRÉVIATIONS	XII
REMERCIEMENTS	XIV
CHAPITRE I - INTRODUCTION	1
CONTEXTE THÉORIQUE	2
Pléthysmographie pénienne	4
Limites de la pléthysmographie pénienne	18
Évaluation visuelle de l'intérêt sexuelle	22
OBJECTIFS/HYPOTHÈSES DE LA RECHERCHE	30
RÉFÉRENCES	33
CHAPITRE II - PREMIER ARTICLE	50
ABSTRACT	52
INTRODUCTION	53
Visual Stimuli	53
Audio Stimuli.....	54
Comparing the Different Modalities.....	55
Penile Plethysmography Limitation.....	56
Computer-Generated Characters.....	57
Current Study	58
METHOD	58
Participants.....	58

Measures	60
Research Design and Procedure.....	63
Statistical Analyses and Data Processing	64
RESULTS	65
Perceived age of stimuli.....	65
Increase in MM Responses	66
Discriminant Validity of Stimuli	67
Classificatory Accuracy of Modalities	69
DISCUSSION	71
Limitations and Conclusion	74
ACKNOWLEDGEMENTS	76
REFERENCES	77
CHAPITRE III - DEUXIÈME ARTICLE	87
ABSTRACT	89
INTRODUCTION	90
Measuring Sexual Interests	91
Pupillometry.....	93
Computer-Generated Characters.....	95
METHOD	96
Participants.....	96
Measure and Data Processing	97
Research Design and Procedure.....	102
RESULTS	103
Discriminant validity	103
Classification Accuracy	106
Criterion validity.....	110
DISCUSSION	113
Limitations, Future Studies and Conclusion.....	115
ACKNOWLEDGMENTS	118
REFERENCES	119

CHAPITRE IV - DISCUSSION ET CONCLUSION.....	129
RETOUR SUR LA PROBLÉMATIQUE.....	130
OBJECTIFS ET PRINCIPAUX RÉSULTATS.....	132
Résultats.....	133
INTÉGRATION DES RÉSULTATS	135
Implications théoriques.....	138
Implications cliniques et perspective d'avenir.....	141
Limites	144
Futures recherches	147
CONCLUSION.....	149
RÉFÉRENCES.....	I

Liste des tableaux

CHAPITRE II – PREMIER ARTICLE	50
Table 1 Participants Sociodemographic Characteristics	60
Table 2 Percentage of Valid PPG Profiles by Millimeter Change and Modality	65
Table 3 Mean Age Given to Virtual Characters by Participants (With Standard Deviations) and Percentage Rates of Accurately Identifying Targeted Age.....	65
Table 4 Interaction Between Stimuli Type and Group	67
Table 5 AUC Obtained for all Modalities, With and Without a 2.5 mm Validity Cut-Off Point and Corresponding Deviance Score.....	71
CHAPITRE III – DEUXIÈME ARTICLE	87
Table 1 Mean luminance values (with standard deviations) for the seven stimuli with maximum and minimum lumens values for images within each stimulus	102
Table 2 Decomposed interaction for stimuli X measure (with means and standard deviations	105
Table 3 Decomposed interaction for stimuli X group (with means and standard deviations	106
Table 4 AUC for all four measures and for the combined Indices with corresponding cut-off scores using either Youdens' index or maximized Sp.....	107
Table 5 Spearman correlations between pupillometric, plethysmographic and combined indices.	108
Table 6 AUC for both PPG and PUP measures with corresponding cut-off Scores using either Youdens' index or maximized Sp.....	109
Table 7 Spearman correlations between SSPI variables and visual and audiovisual (standardized) PPG results.....	110
Table 8 Spearman correlations between SSPI variables and visual and audiovisual (standardized) PUP results.....	110
Table 9 Spearman correlations between standardized PPG and PUP results.	111

Liste des figures

CHAPITRE II – PREMIER ARTICLE	50
Figure 1 Censured Adult female stimulus.....	63
Figure 2 PPG Standardized Results by Stimuli, Modality of Presentation, and Group (with Standard Errors).....	68
CHAPITRE III – DEUXIÈME ARTICLE	87
Figure 1 Participants overall standardized results by stimuli and group.	104

Liste des abréviations

ATSA	Association for the treatment of sexual abusers
AUC	Aire sous la courbe / Area under the curve
IPPM	Institut Philippe-Pinel de Montréal
PGO	Personnage généré par ordinateur
PPG	Pléthysmographie pénienne / Penile plethysmography
PUP	Pupillométrie / Pupillometry
ROC	Receiver operating characteristic analyses
SD	Écart type / Standard deviation
SE	Sensibilité / Sensitivity
SP	Spécificité / Specificity

À Lucie Marschall et Jean-Claude Lévesque

Remerciements

Mon expérience doctorale à l'Université de Montréal fut une source motivante de stimulation, de joie, de frustration et d'accomplissement. Ces cinq années m'ont permis de rencontrer et de travailler avec plusieurs personnes qui ont eu un impact marquant sur mon développement académique, professionnel et personnel. Je profite donc de cette section pour remercier plusieurs de ces personnes.

Je veux tout d'abord remercier mes directeurs de recherche, Joanne-Lucine Rouleau et Patrice Renaud, sans qui ce projet n'aurait pas été possible. Ils m'ont non seulement permis de développer mes compétences et mon autonomie en recherche, mais m'ont aussi offert plusieurs expériences cliniques qui ont su modeler ma compréhension de la psychologie.

Je tiens aussi à remercier l'ensemble de l'équipe de recherche à l'Institut Philippe-Pinel de Montréal. Ils ont su être une source de support, de compétences et de camaraderie au cours de ces années. Leur présence a rendu mon séjour à l'IPPM non seulement productif, mais plaisant.

Je remercie aussi Jean R. Séguin et Natalie Castellanos Ryan, du GRIP, qui ont facilité mes premières expériences en recherche, soutenu mes premières publications et qui sont demeurés des sources de support moral et professionnel depuis plus de six ans.

Non lié à la recherche, mais tout de même central durant mon cheminement académique, je tiens à remercier mes collègues doctorants qui m'ont accompagné dans ce parcours en rendant ce chapitre de ma vie plus agréable et humain.

Je réserve un grand remerciement à mes parents et ma famille qui m'ont supporté inconditionnellement. Sans eux, mes études n'auraient pas été possibles. Dans le même ordre

d'idées, je tiens à remercier ma conjointe, Sophie Rodrigue, qui a été une source de conseils, d'encouragements et de support immuable depuis les débuts de mes études universitaires.

Chapitre I - Introduction

Contexte théorique

La présence d'intérêts sexuels déviants demeure à ce jour le meilleur prédicteur unique du risque de récidive chez les agresseurs sexuels d'enfants, prédisant jusqu'à 32 % de la variance de la récidive sexuelle (Hanson et Bussière, 1998). La présence d'intérêts et de fantasmes sexuels envers les enfants est aussi l'un des facteurs clés permettant l'établissement d'un diagnostic de pédophilie ou de trouble pédophilique (American Psychiatric Association, 2000, 2013; DSM-IV-TR et DSM-5). Néanmoins, la présence de fantasmes déviants, critère A de ce diagnostic, peut s'avérer difficile à évaluer. En effet, il est courant pour les individus répondant d'emblée au critère B (*c.-à-d.* un individu ayant passé à l'acte) et au critère C (*c.-à-d.* être âgé de 16 et plus et être âgé d'au moins 5 ans de plus que la victime) de nier la présence de ces intérêts sexuels par peur de s'incriminer davantage ou d'aggraver leur situation. Ainsi, il apparaît nécessaire d'avoir divers outils pour évaluer efficacement la présence d'intérêts sexuels envers les enfants afin d'être en mesure de protéger la population générale, tout en offrant les ressources et les traitements nécessaires aux individus les plus à risque, limitant ainsi leur risque de récidive.

Avant de mesurer la présence possible d'intérêts sexuels envers les enfants, il est préalablement nécessaire de définir ce qu'est un intérêt ou une excitation sexuelle. Singer (1984) postule que l'excitation sexuelle peut être expliquée par l'agencement de trois composantes ou phases distinctes : la réponse esthétique, la réponse d'approche et la réponse génitale. Dans l'ordre, la réponse esthétique se caractérise par un sentiment plaisant d'attirance sexuelle qui se déclenche au niveau cognitif. Cette réaction arrive généralement en réponse à un stimulus visuel (*p. ex.*, visage) ou auditif (*p. ex.*, voix) jugé attrayant. Les composantes physiologiques d'une réponse esthétique peuvent être désignées par un maintien de l'attention

et du regard sur le stimulus d'intérêt et par l'expression vocale d'un intérêt envers ce stimulus. De son côté, la réponse d'approche se produit lorsque l'intérêt esthétique d'ordre cognitif se transforme en comportements moteurs visant le déplacement vers le stimulus sexuel ou tout rapprochement vers celui-ci. Cette recherche se définit donc par un rapprochement physique vers le stimulus et un désir cognitif d'entrer en contact avec celui-ci (*p. ex.*, le toucher, la saisie ou être palpé par cet individu). Finalement, la réponse génitale se caractérise par une activation physiologique automatique. Celle-ci peut être distinguée par une augmentation du rythme cardiaque, de la rapidité de la respiration et de la sudation. Néanmoins, Singer (1984) postule que le meilleur moyen d'évaluer cette troisième phase consiste à mesurer la réponse génitale d'un individu. Il est à noter que ces trois phases sont considérées distinctes, puisque l'une d'elles peut être observée en l'absence d'indices des deux autres. En ce sens, une réponse esthétique pourrait être observée en l'absence d'une réponse génitale et une réponse génitale pourrait être observée en l'absence d'une réponse d'approche. Il apparaît ainsi qu'en mesurant au moins une des phases de ce modèle durant la présentation de stimuli sexuels variés (incluant des enfants et des adultes), il serait possible d'évaluer la présence d'intérêts sexuels envers ces stimuli et de graduer l'intensité ou l'ordre de ces préférences (critère A du diagnostic de pédophilie). En ce sens, la littérature suggère différentes méthodes pour évaluer ces phases incluant la pléthysmographie pénienne (PPG), le temps d'observation (*viewing time*) et la pupillométrie (PUP; Gaither et Plaud, 1997; Kalmus et Beech, 2005; Lalumière et Harris, 1998).

Pléthysmographie pénienne.

La PPG, déjà utilisée depuis plus de 60 ans (Freund, 1963, 1965, 1976), consiste en la mesure des réponses érectiles des participants durant la présentation de divers stimuli. Cette méthode, aussi appelée phallométrie, vise à provoquer une réponse esthétique et génitale en ayant recours à des stimuli visuels et auditifs. Malgré la stimulation des phases esthétiques et génitales, la PPG se concentre exclusivement sur l'évaluation de la réponse génitale. Cette mesure, originalement développée par Kurt Freud durant les années 50, a rapidement été utilisée pour évaluer les intérêts sexuels préférentiels envers les enfants (Akerman et Beech, 2011; Proulx, 1993), les intérêts sexuels envers la violence (Launay, 1999; Laws, Hanson, Osborn, et Greenbaum, 2000), et fait partie de l'évaluation du risque de récidive chez les individus ayant déjà commis des agressions sexuelles envers des enfants (Clift, Rajlic, et Gretton, 2009; Hanson et Bussiere, 1998; Hanson et Morton-Bourgon, 2005; Malcolm, Andrews, et Quinsey, 1993).

Même si les intérêts sexuels déviants, mesurés par la PPG sont, lorsque pris individuellement, les meilleurs prédicteurs dynamiques de la récidive sexuelle, ils doivent parallèlement à des prédicteurs statiques, être associés à d'autres facteurs de risque dynamiques pour une évaluation plus complète du risque de récidive sexuelle. (Hanson et Morton-Bourgon, 2009).

Dans un premier temps, la STATIC-99 et plus récemment la Static 99-R évaluent dix éléments centrés sur trois domaines liés au passé de la personne évaluée, soit les antécédents criminels, les caractéristiques de la victime et l'historique des rapports entre l'agresseur et la victime (Hanson et Thornton, 2000; Harris et al., 2003, Hanson, Babchishin, Helmus, et Thornton, 2013).

Ensuite, l'évaluation des facteurs de risque dynamiques incluant les intérêts sexuels déviants, se fait à l'aide d'échelles actuarielles incluant notamment pour la déviance sexuelle : la STABLE-2007 et l'AIGU-2007 (Hanson et Thornton, 2000; Harris et al., 2003, Hanson et Morton-Bourgon, 2009). Respectivement, la STABLE-2007 évalue treize items incluant : l'influence sociale, la capacité à demeurer dans une relation stable, l'identification émotive aux enfants, l'hostilité à l'égard des femmes, le rejet social, l'absence d'empathie envers autrui, la présence de gestes impulsifs, les faibles aptitudes cognitives pour la résolution de problèmes, les émotions négatives, la présence d'une forte libido, la présence de préoccupations personnelles, le recours au sexe comme mécanisme d'adaptation, les intérêts sexuels déviants et le manque de coopération dans le cadre de la surveillance. Finalement, l'AIGU-2007 se base sur sept facteurs dynamiques pour évaluer le risque de récidive sexuelle ou avec violence et le risque de récidive générale à court terme. Ces facteurs sont : l'accès aux victimes, l'effondrement émotionnel, l'effondrement des supports sociaux, l'hostilité, la toxicomanie, les préoccupations sexuelles et le rejet de surveillance. Ces trois outils actuariels présentent aussi l'avantage de pouvoir être utilisés conjointement pour obtenir une évaluation plus fiable du risque de récidive (Hanson, Harris, Scott, et Helmus, 2007). Ainsi, l'évaluation du risque de récidive sexuelle implique plusieurs facteurs validés parallèlement à celui des intérêts sexuels déviants.

La PPG existe sous deux formes de procédures distinctes, soit la PPG volumétrique ou la PPG circonférentielle (Kuban, Barbaree, et Blanchard, 1999). La PPG volumétrique fait usage d'un cylindre scellé hermétiquement à l'aide d'un joint d'étanchéité placé sur le pénis du participant. Les variations du volume du pénis engendrées lors de la tumescence sont alors évaluées en mesurant les variations dans la pression de l'air contenu dans le cylindre. La PPG

circonférentielle, quant à elle, fait usage d'une jauge en caoutchouc, remplie de mercure et dans laquelle, un courant électrique circule continuellement. Une fois placée autour du pénis, cette jauge est étirée par l'action de la tumescence produisant des changements de la conductance électrique du mercure. Ces variations de conductance peuvent alors être converties en variation de la circonférence du pénis. Malgré le fait que la PPG volumétrique soit plus sensible que la PPG circonférentielle, cette dernière est plus répandue étant donnée sa facilité d'utilisation. De plus, ces deux mesures présentent une bonne validité de convergence ($r > 0,80$) lorsque les amplitudes péniennes obtenues par les participants en PPG circonférentielle sont égales ou supérieures à 2,5 mm. Pour plus d'informations sur ces deux mesures ainsi que la concordance entre celles-ci, il est conseillé de lire Kuban et al., (1999).

Dans ces deux cas, la PPG fonctionne par l'entremise d'une procédure à deux étapes. La première étape consiste en la présentation d'une variété de stimuli (*p. ex.*, visuels, auditifs, incluant des enfants, des adultes, des stimuli sexuels et non sexuels) durant laquelle on enregistre les réponses érectiles du participant. La deuxième étape comprend la comparaison des résultats obtenus en fonction du type de stimuli présentés (pédophilique ou non pédophilique; Harris, Rice, Quinsey, Chaplin, et Earls, 1992; Kalmus et Beech, 2005; Lalumière et Harris, 1998; Rosen et Keefe, 1978). En ce sens, la réponse érectile la plus élevée est utilisée comme substitut de l'intérêt sexuel le plus fort. Cette procédure dépend donc largement des stimuli utilisés et de leur capacité à générer une réponse génitale. Ces stimuli se regroupent généralement en deux grandes catégories, c'est-à-dire les stimuli auditifs et les stimuli visuels.

Stimuli auditifs. Cette modalité de stimuli est généralement composée de courtes vignettes, mettant en scène un homme ayant des contacts sexuels et non sexuels avec des

hommes, des femmes et des enfants. Ces courtes histoires sont généralement décrites avec différents niveaux de violence variant d'aucune violence à l'utilisation de coercition et d'abus physique pour avoir un contact sexuel. Ces stimuli permettent d'évaluer la présence d'intérêts sexuels envers les enfants et l'évaluation du risque de récidive sexuelle (Proulx et al., 1997; Quinsey, Rice, et Harris, 1995). De plus, ils permettent la discrimination entre les individus ayant déjà commis une agression sexuelle envers des enfants et ceux n'ayant jamais commis de telles actions (Murphy, Haynes, Stalgaitis, et Flanagan, 1986; Thornton et Laws, 2009). L'avantage, à ce jour, des stimuli auditifs est qu'ils permettent l'évaluation d'intérêts sexuels envers la violence, l'humiliation et la coercition (Chaplin, Rice et Harris, 1995; Harris, Lalumière, Seto, Rice et Chaplin, 2012; Harris et al., 1992; Lalumière et Harris, 1998; Launay, 1999; Laws, Hanson, Osborn et Greenbaum, 2000; Marshall, Barbaree et Christophe, 1986; Proulx, 1989; Quinsey et Chaplin, 1988). À cette fin, certains ensembles de stimuli auditifs (*p. ex.*, Quinsey et Chaplin, 1988) font usage de vignettes d'agression sexuelle contenant différents niveaux de violence, de coercition et d'humiliation, permettant ainsi d'établir si un patient présente un intérêt envers ce type d'action. Certaines recherches ont aussi démontré que les stimuli auditifs présentant un haut niveau de coercition étaient les mieux validées pour l'évaluation des intérêts sexuels envers la violence et la coercition (Lalumière et Harris, 1998). Ces bandes auditives peuvent même inclure des scénarios d'agression physique envers un enfant en l'absence de comportements sexuels pour distinguer l'apport de la violence aux résultats du test.

Stimuli visuels. Les stimuli visuels, quant à eux, sont utilisés depuis les débuts de l'évaluation phallométrique (Freund, 1963, 1965). Cette modalité de présentation est généralement composée d'images réelles ou synthétiques présentant des enfants et des adultes

de différents groupes d'âge et de niveaux de nudité. Ces stimuli sont un bon prédicteur de la récurrence sexuelle et sont capables de discriminer entre les agresseurs sexuels d'enfants et les non-agresseurs (Barbaree et Marshall, 1988; Malcolm, Andrews et Quinsey, 1993; Murphy et al., 1986; Nagayama Hall, 1991; Quinsey et al., 1995). Leur particularité consiste en leur capacité de discriminer plus efficacement l'âge préférentiel de leurs victimes (Harris et al., 1992; Lalumière et Harris, 1998; Launay, 1999; Malcolm et al., 1993; Marshall et al., 1986; Proulx, 1989). Les images d'enfant ont aussi l'avantage d'être les plus sensibles pour l'évaluation de l'âge et du sexe préférentiel des agresseurs sexuels par rapport aux autres modalités de stimuli (Lalumière et Harris, 1998; Laws, 2009). La modalité visuelle a grandement évolué au cours des années et a été utilisée sous plusieurs formes (*c.-à-d.* photo monochrome, photo en couleur, montage de différentes photos pour former une personne synthétique, personnages virtuels) et avec différents contenus (*c.-à-d.* individu avec ou sans vêtement(s) et avec ou sans arrière-plan(s)). Toutefois pour des raisons éthiques, depuis neuf ans, l'utilisation de photographies d'enfants aux fins d'évaluations n'est plus permise au Canada. Certains considèrent que l'utilisation de réelles images d'enfants constitue une victimisation potentielle et une atteinte aux droits fondamentaux de ces enfants (Abel, Huffman, Warberg et Holland, 1998; Kalmus et Beech, 2005; Launay, 1994). Il semblerait que cette interdiction ait, entre autres, promu l'utilisation des stimuli auditifs tout en engendrant l'intérêt pour d'autres types de stimuli visuels.

Personnages générés par ordinateur. Dans le but de pallier la perte de stimuli visuels classiques (*c.-à-d.* images d'enfants), certains chercheurs ont favorisé le développement de stimuli ou de personnages virtuels générés par ordinateur (PGO; Chartier et al., 2006; Dombert et al., 2013; Goyette, Rouleau et Renaud, 2012; Renaud et al., 2011; Trottier,

Rouleau et Renaud, 2013; Zappalà et al., 2013). Ces PGO sont modelés pour respecter les caractéristiques morphologiques d'enfants aux stades I et II et d'adultes au stade V de l'échelle de développement physiologique de Tanner et Whitehouse (1976). Cette échelle des transformations physiologiques vers la puberté comporte cinq niveaux ou stades variant de prépubère (Stade I) à adulte (Stade V). Ces stades sont utilisés pour caractériser les seins des filles, les pénis et les testicules des garçons, de même que le développement de la pilosité pubienne des deux sexes. Par exemple, au stade I, les filles et les garçons ne présentent aucun poil pubien, les filles n'ont pas de seins et ne présentent aucune élévation des mamelons, de plus leurs aréoles sont petits et plats. Au même stade, les garçons ne présentent aucun changement quant à la taille de leur pénis, leur scrotum ou leurs testicules. Les changements commencent seulement à partir du stade II, stade auquel les deux sexes présentent quelques poils pigmentés et droits. Le corps des filles présentent également l'apparition du bourgeon mammaire, ainsi qu'une augmentation du diamètre et un bombement des aréoles. Pour une description détaillée des stades de Tanner et Whitehouse (1976), le lecteur peut consulter Marshall et Tanner (1969, 1970).

Les PGO présentant des enfants aux stades I et II et des adultes au stade V sont ceux qui ont démontrés leur efficacité à discriminer les agresseurs sexuels d'enfants des non-agresseurs, et ce, de manière comparable ou supérieure à ce qui est possible avec les stimuli auditifs standards (Dombert et al., 2013; Goyette et al., 2012; Renaud et al., 2002; Renaud et al., 2010; Trottier et al., 2013; Trottier et al., 2015). Les PGO ont aussi été plus largement utilisés dans le domaine de la recherche sur la sexualité pour distinguer d'autres types de préférences sexuelles (*p. ex.*, hétérosexuel, homosexuel, émotionnel, morphologique; Chartier et al., 2006; Dennis, Rouleau, Renaud, Nolet et Saumur, 2014; Zappalà et al., 2013) ou pour

identifier la présence de patrons visuels distincts lors de l'observation de stimuli visuels sexuels (Trottier et al., 2014). Ces stimuli permettent de remettre de l'avant les stimuli visuels sans pour autant enfreindre les règles d'éthique plus rigoureuses que celles présentes durant les années 60. Actuellement, la littérature scientifique disponible sur la PPG suggère l'utilisation de différentes modalités de stimuli pour arriver à une discrimination efficace des intérêts sexuels. Néanmoins, il existe une certaine variabilité à l'intérieur de cette modalité. En effet, il existe différents ensembles de PGO qui ont été créés au cours des années et ceux-ci ne présentent pas tous les mêmes capacités discriminantes. Par exemple, les recherches démontrent que les PGO comprenant des expressions émotionnelles présentent une meilleure capacité de discrimination entre les participants d'un groupe d'agresseurs sexuels et les participants d'un groupe de comparaison que les PGO neutres (Dennis, Rouleau et Renaud., 2015). Similairement, l'une des premières études sur les PGO n'a pas su démontrer leur capacité discriminante (Renaud et al., 2009). De plus, les PGO modelés pour représenter des adolescents ne permettent pas la discrimination entre les groupes d'agresseurs sexuels d'enfant et le groupe de sujets provenant de la population générale (Goyette et al., 2012). Finalement, le développement initial de ces stimuli s'avère complexe, dispendieux et long. Il est donc nécessaire de déterminer quelles modalités ou combinaisons de modalités seraient à privilégier lors de l'évaluation plétysmographique (PPG) afin de faire le choix le plus judicieux possible.

Comparaison d'amplitudes péniennes obtenus aux différentes modalités. Plusieurs modalités de stimuli ont été suggérées et utilisées au cours des années pour susciter l'intérêt sexuel de participants (*c.-à-d.* vidéos, images, scénarios auditifs, textes érotiques, fantasmes, la combinaison de scénarios auditifs et d'images, stimuli générés par ordinateur, immersion virtuelle; Abel, Barlow, Blanchard et Mavissakalian, 1975; Abel, Blanchard et Barlow, 1981;

Goyette et al., 2012; Julien et Over, 1988; Trottier et al., 2014). Néanmoins, peu d'études ont directement tenté de comparer ces modalités les unes aux autres.

Une telle étude, menée auprès d'individus présentant des intérêts sexuels problématiques et comparant la modalité visuelle, auditive, fantasmes et vidéo a démontré que seule la modalité vidéo se distinguait des autres en générant de plus fortes réponses érectiles (Abel et al., 1981). Similairement, Julien et Over (1988) ont comparé l'efficacité de cinq modalités à provoquer une réponse érectile chez un groupe de participants hétérosexuels. Dans cette étude, les réponses pénienues ont été enregistrées avec la présentation d'une vidéo érotique, d'images sexuelles générées à partir de la vidéo, d'un texte décrivant l'activité présentée dans la vidéo, d'une version auditive du texte et d'une situation où le participant devait imaginer des relations sexuelles entre un homme et une femme adulte à l'aide de phrases indicées. Les résultats ont démontré que la modalité vidéo (film érotique) générait des réactions pénienues significativement plus fortes que les autres modalités. Finalement, une étude de Gaither et Plaud (1997) a démontré que les sons présents dans une vidéo érotique contribuaient significativement à augmenter la réponse érectile des participants, et ce, en comparaison à la même vidéo sans son. Ces résultats sont corroborés par d'autres recherches comparant la modalité vidéo à la modalité auditive et démontrant que celle-ci suscite la plus forte réaction à la fois chez des participants hétérosexuels (Sakheim, Barlow, Beck et Abrahamson, 1985) et homosexuels (Abel et al., 1975; Sakheim et al., 1985). De plus, ces études ont aussi établi qu'individuellement les stimuli auditifs (*c.-à-d.* histoires) et visuels standards (*c.-à-d.* photographies) ne se distinguaient pas significativement les uns des autres (Abel et al., 1975; Julien et Over, 1988; Sakheim et al., 1985).

La nature plus récente des PGO fait en sorte que peu d'études ont pu comparer leur capacité à générer des réponses érectiles à celles des autres modalités (*c.-à-d.* vidéo, visuel standard, fantasmes). Néanmoins, les études disponibles montrent que les PGO sont comparables aux stimuli auditifs quant aux patrons de réponses péniennes qu'ils génèrent, suggérant qu'ils sont similaires aux stimuli visuels standards (Chartier et al., 2006; Dombert et al., 2013; Goyette et al., 2012; Laws et Gress, 2004; Zappalà et al., 2013).

En somme, l'évaluation pléthysmographique ayant recours à des vidéos semble avoir la plus forte probabilité de susciter une réaction pénienne. L'évaluation ayant recours aux stimuli auditifs ou visuels (virtuels) génère des réponses sexuelles similaires, mais moindres que les vidéos. Il semblerait donc que des stimuli suscitant plus d'une modalité sensorielle seraient à privilégier afin de générer de plus fortes réponses péniennes. Par contre, il est important de noter que l'utilisation de stimuli vidéo pour l'évaluation des agresseurs d'enfants est déconseillée pour des raisons éthiques et pour leur tendance à générer un effet plafond chez les participants exposés (*c.-à-d.* érection complète à tous les stimuli), réduisant ainsi la capacité discriminante de l'évaluation (Marshall et Fernandez, 2000).

Cette idée d'utiliser un agencement de stimuli visuels et auditifs est déjà appliquée dans plusieurs laboratoires. Ces deux modalités sont fréquemment présentées de manière séquentielle (Clift, Rajlic et Gretton, 2009; Goyette et al., 2012; Hall, Hirschman et Oliver, 1995; Laws et al., 2000; Marshall et al., 1986; Trottier et al., 2013) ou concurrente (Blanchard, Klassen, Dickey, Kuban et Blak, 2001; Blanchard et al., 2007; Blanchard et al., 2012; Blanchard, Kuban et al., 2009; Blanchard, Lykins et al., 2009; Farrall et Card, 1988; Freund et Blanchard, 1989; Freund, Watson, Dickey et Rienzo, 1991; McAnulty et Adams, 1990). Ces études présentent cependant des limites puisqu'aucune d'entre elles ne semble

avoir comparé l'efficacité de l'utilisation de ces deux types de stimuli à l'utilisation d'un seul. En ce sens, malgré des données significatives quant à l'influence des modalités de présentation sur la force de l'amplitude pénienne obtenue, il n'est pas clair si la combinaison de modalités engendre une capacité discriminante supérieure à celle obtenue lorsqu'on présente ces mêmes modalités séparément.

À cet effet, une récente méta-analyse comparant les différentes modalités de stimuli (auditive, visuel, audiovisuel, généré par ordinateur et vidéos) lors de la présentation de stimuli pédophiliques a démontré que l'ensemble des modalités étaient comparables les unes aux autres dans leur capacité à discriminer entre des groupes cliniques et normatifs. Seule une distinction est ressortie démontrant que les personnages générés par ordinateur étaient supérieurs aux vidéos quant à leur capacité discriminante (McPhail et al., 2017).

Traitement des données phallométriques et capacité discriminante. Peu importe la modalité utilisée, les amplitudes péniennes brutes obtenues lors d'une évaluation phallométrique doivent être traitées et analysées. Or, la manière de traiter les données brutes obtenues demeure en soi une source de controverse dans le domaine de la PPG (Earls, Quinsey, et Castonguay, 1987; McPhail et al., 2017). Au moins trois types de traitements sont présents dans la littérature : aucun traitement, conversion des données brutes en pourcentage d'une pleine érection et standardisation des résultats bruts par score z . À ces traitements sont ensuite ajoutés les calculs d'index de déviance qui sont présents pour chacune des méthodes.

Dans les trois cas, les amplitudes péniennes brutes sont obtenues en soustrayant les niveaux de base des participants (détumescence) à l'amplitude maximale obtenue à chacun des stimuli. À ce jour, la littérature suggère que les données brutes en soi n'offrent aucune

information pertinente pour l'évaluation des intérêts sexuels (Wilson, Abel, Coyne, et Rouleau, 1992). En ce sens, une érection de 3 mm obtenus à un stimulus en particulier ne permet aucune interprétation clinique des intérêts d'un participant. Les résultats doivent, au minimum, être comparés les uns aux autres, ce qui est accompli en calculant les indices de déviance. Ces derniers sont obtenus en divisant l'amplitude pénienne la plus élevée aux stimuli comportant des enfants par l'amplitude pénienne la plus élevée aux stimuli comportant un adulte. Ensuite, cet indice est utilisé pour évaluer la préférence relative des participants envers les enfants (enfants/adulte ou déviant/non déviant). Par exemple, si un participant obtient un score de 1, cela signifie qu'il a réagi de manière identique aux stimuli d'enfants et d'adulte. Alors que si un participant obtient un score supérieur à 1, cela indique des réactions plus fortes aux stimuli d'enfants qu'aux stimuli d'adultes (Lalumière et Harris, 1998). La principale critique retenue contre l'interprétation de données brutes porte sur le haut niveau de variabilité intra-sujet et inter-sujet que procure cette mesure. Un exemple de variabilité intra-sujet serait que pour un même participant une amplitude pénienne de 3 mm soit obtenue à un stimulus d'enfant fille au début de l'évaluation, qu'une amplitude de 7 mm soit obtenue pour un même type de stimulus à la fin de l'évaluation ou qu'une amplitude de 1 mm soit obtenu au même stimulus au cours d'une seconde évaluation. Ces variations importantes concernant les amplitudes obtenues par un même participant au même type de stimulus compliquent l'interprétation des résultats ainsi que leur validité, particulièrement lorsqu'il s'agit de données brutes. De son côté, la variabilité inter-sujet fait plutôt état du fait qu'une amplitude pénienne de 5 mm pour un participant pouvant atteindre une amplitude pénienne maximale de 25 mm n'est pas représentative d'un même résultat pour un autre participant dont les amplitudes

péniennes peuvent atteindre un maximum de 50 mm. C'est entre autres sur ce principe que se base la prochaine méthode de traitement des données PPG.

La seconde méthode, qui tente de palier les critiques mentionnées précédemment, consiste à transformer les résultats obtenus en un pourcentage de la pleine érection du participant (Earls et al., 1987; Howes, 1995; Murphy et Barbaree, 1994; Murphy, DiLillo, Haynes, et Sterre, 2001). Dans le meilleur scénario, une pleine érection est obtenue à chacune des évaluations phallométriques soit par l'entremise des stimuli standard ou par l'utilisation de stimuli d'échauffement (p. ex., vidéo pornographique adulte) précédant l'évaluation clinique. Ce score de pleine érection est ensuite utilisé pour calculer le pourcentage d'érection obtenu par le participant à chacun des stimuli de l'évaluation clinique : amplitude pénienne obtenue à un stimulus divisé par l'amplitude pénienne représentant la pleine érection multipliée par 100. Néanmoins, cette amplitude de pleine érection peut être difficile à obtenir (Howes, 2003). Pour cette raison, différents laboratoires utilisent des scores approximatifs d'une pleine érection pour calculer les pourcentages d'érection obtenus (Hunter et Goodwin, 1992). Certains auront recours à l'amplitude la plus élevée obtenue au cours de l'évaluation, à une valeur standard, à une estimation offerte par le participant, ou encore à un pourcentage d'augmentation du niveau de base attendu (p. ex., 30% à 40%) pour estimer une pleine érection (Howes, 2003; Hunter et Goodwin, 1992; Lundeen, Manohar, et Parks, 1983). Par ailleurs, une étude normative a analysé les résultats de 724 hommes ayant obtenu une pleine érection démontrant une augmentation moyenne de 32,57 mm avec écart type de 8,76 mm. (Howes, 2003). Ces auteurs suggèrent donc que dans 95 % des cas, une amplitude pénienne de 47 mm sera représentative d'une pleine érection. Ces résultats sont comparables à ceux obtenus dans une autre étude normative démontrant qu'en moyenne, un homme en Amérique

du Nord obtient une pleine érection à 27,94 mm (Jamison et Gebhard, 1988). De plus, Howes, (2003) suggère d'utiliser le score de 9,4 mm comme amplitude standard d'une érection représentant au moins 20% de la capacité érectile maximale. Il est à noter que les auteurs recommandent d'utiliser ces résultats normatifs pour calculer le pourcentage d'érection seulement lorsque l'amplitude d'une pleine érection est inconnue. Une fois ces pourcentages obtenus pour chacune des catégories de stimuli, il est alors possible de calculer un indice de déviance de manière identique à ce qui a été décrit pour les résultats bruts. La méthode du pourcentage de pleine érection a l'avantage de prendre en compte une partie de la variabilité intra-sujet de même que la variabilité inter-sujet dans les réactions pénienues (Able, 1976). Ainsi, une amplitude brute de 3 mm obtenue par un individu présentant une pleine érection à 25 mm (12% d'une pleine érection) est plus facilement comparable au même changement chez un individu ayant une pleine érection à 50 mm (6% d'une pleine érection). Par contre, cela est seulement possible lorsque la valeur représentant la pleine érection est connue, ce qui représente une limite importante à cette méthode.

Finalement, la troisième méthode recommandée consiste à normaliser les amplitudes pénienues brutes à chacun des stimuli en appliquant une transformation par score z . Cette transformation permet d'obtenir un indice de la force de réaction d'un participant à un stimulus en comparaison à ses réactions à l'ensemble des stimuli (c.-à-d., la moyenne de ses réactions; Blanchard, Klassen, Dickey, Kuban, et Blak, 2001; Earls et al., 1987). L'avantage de cette méthode est qu'elle permet de contrôler pour la variation intra-sujet et inter-sujet en normalisant les scores bruts, en les rendant comparables les uns aux autres, et en les plaçant sur une échelle identique pour les comparer avec ceux obtenus par d'autres participants (Earls et al., 1987; Freund, Scher, Racansky, Campbell, et Heasman, 1986; Harris, Rice, Quinsey,

Chaplin, et Earls, 1992; Quinsey, Chaplin, et Carrigan, 1979). Cette transformation fait alors en sorte que chaque participant devient sa propre échelle de comparaison, comme c'était le cas avec le pourcentage d'une érection, sans pour autant avoir besoin de connaître l'amplitude associée à une pleine érection. Une fois les résultats standardisés, ceux-ci peuvent être utilisés pour calculer un indice de déviance en soustrayant le score z le plus élevé obtenu à la catégorie de stimuli adultes au score z le plus élevé obtenu à la catégorie enfant. Puisque la moyenne des scores z est de 0, un résultat de 0 à cet indice indique une égalité entre les résultats obtenus pour les adultes et les enfants, alors que tout score supérieur à 0 indique des résultats plus élevés obtenus à l'une des catégories enfant (Harris et al., 1992; Lalumière et Harris, 1998; McPhail et al., 2017).

Même si certaines études démontrent que cette dernière méthode est la mieux validée des trois (Harris et al., 1992; Lalumière et Harris, 1998) une critique demeure quant à sa capacité de discriminer la présence d'intérêts sexuels envers des enfants lorsque l'ensemble des résultats obtenus est de faible amplitude. C'est pourquoi certains auteurs critiquent l'action de normaliser les résultats qui ne suivent pas une courbe normale. Cette action aurait pour effet de générer une distorsion des données augmentant les risques de détections de faux positifs (Barbaree et Mewhort, 1994; Howes, 2003). Cependant, cette critique ne semble pas supportée par l'ensemble des chercheurs de ce domaine (Lalumière et Harris, 1998). Certains auteurs suggèrent plutôt que les résultats ayant une faible amplitude ne sont pas problématiques en soi, mais le deviennent lorsque les amplitudes pénienues aux stimuli sexuels (déviants ou non) sont inférieures aux amplitudes pénienues aux stimuli neutres qui sont créés pour ne pas générer de réponse sexuelle (Freund, Watson, et Rienzo, 1988; Lalumière et Harris, 1998). Cette notion de faible taux de réponse et de seuil de validité

minimale, sera abordée à nouveau dans la section portant sur les limites de la pléthysmographie.

En somme, on peut en conclure que les trois méthodes mentionnées précédemment ont été comparées les unes aux autres avec des résultats variables. Par exemple, certaines recherches montrent que l'utilisation du pourcentage de pleine érection est supérieure à la standardisation des résultats par score z (Barbaree et Mewhort, 1994) alors que d'autres études suggèrent le contraire (Earls et al., 1987). La méta-analyse de McPhail et ses collègues (2017) clarifie la littérature en démontrant qu'une fois les données aberrantes contrôlées, l'utilisation d'indices de déviations, obtenue avec des scores z est supérieure aux autres formes de discrimination incluant les données brutes, les pourcentages de pleine érection, ainsi que les indices de déviations dérivées de ces résultats.

Limites de la pléthysmographie pénienne.

Depuis son développement, la PPG a eu son lot de critiques concernant ses caractéristiques psychométriques. Ces critiques se résument généralement, mais non exclusivement par : la facilité associée à la falsification des résultats, le manque de standardisation des stimuli et des protocoles utilisés dans les différents laboratoires, la faible capacité test-retest de la mesure, les difficultés associées à la prédiction de la récurrence sexuelle, le haut taux de profils invalides et par la restriction de l'utilisation de certains stimuli (Akerman et Beech, 2012; Fernandez, 2002; Laws, 2009; Marshall et Fernandez, 2000; Odesho, 2004; Renaud et al, 2012; Wilson et Miner, 2016). Ces faiblesses ont même poussé certains auteurs à décrire la PPG comme une forme d'art plutôt qu'une science (Laws, 2009). La première de ces critiques fait état de la facilité avec laquelle des sujets peuvent influencer

leurs résultats psychométriques en employant différents stratagèmes tels que la distraction, l'évitement ou le recours à un stimulus aversif (Card et Farrall, 1990; Marshall et Fernandez, 2000). Malgré plusieurs études s'étant penché sur le sujet et différentes stratégies employées afin de tenter de s'assurer que les participants portent effectivement attention aux stimuli employés, cette faiblesse demeure toujours présente (Quinsey et Chaplin, 1988; Renaud et al., 2012; Trottier et al., 2013).

Le manque de standardisation des protocoles fait référence à un éventail de caractéristiques qui peuvent varier d'un laboratoire à un autre : l'appareillage, la durée de présentation des stimuli, le type de stimuli employés et le nombre de catégories par stimuli, le traitement des données brutes (Laws, 2009; Marshall, 2013). Par exemple, certains laboratoires ont recours à des stimuli visuels, auditifs ou à des combinaisons des deux. Une large variabilité existe d'ailleurs à l'intérieur des mêmes modalités de stimuli. Certains laboratoires ayant par exemple recours à des stimuli visuels peuvent employer des photographies d'enfants, alors que d'autres peuvent employer des stimuli généraux par ordinateur sans que ceux-ci n'aient jamais été comparés. Malgré la reconnaissance de cette limite et un appel à la standardisation des stimuli en PPG, cette faiblesse demeure toujours présente (Akerman et Beech, 2012; Laws, 2009; Marshall et Fernandez, 2000; McPhail et al., 2017).

La capacité test-retest fait état de la fiabilité de la mesure et évalue sa capacité à obtenir des résultats similaires entre deux passations ou plus avec les mêmes participants. Ces passations sont généralement distancées d'une semaine à quelques mois. Ainsi, la PPG montre une grande variabilité quant à sa fiabilité test-retest avec des corrélations variant de $r = 0,53$ à $r = 0,79$ (Kalmus et Beech, 2005; Marshall, 2006; Marshall, 2013; Marshall et Fernandez, 2000;

Renaud et al. 2010). Lalumière et Harris (1998) stipulent que ces faibles résultats test-retest sont possiblement causés par un effet d'apprentissage qui permettrait aux participants d'apprendre à influencer les résultats ou en générant un effet d'habitation. De plus, certaines recherches soutiennent que la PPG aurait une faible capacité prédictive post-traitement (Lalumière et Harris, 1998; Rice, Quinsey, et Harris, 1991). Néanmoins, la récente méta-analyse de McPhail et al. (2017) a non seulement démontré que la majorité des stimuli employés (visuels, auditifs ou combinés) présentent une bonne capacité discriminante, mais elle soutient aussi que la PPG pré et post-traitement permettrait de prédire les risques de récidive chez les individus ayant commis une agression sexuelle envers des enfants extrafamiliaux de sexe féminin et masculin.

Il est aussi pertinent d'aborder plus en détail le haut taux de profils invalides présents en PPG, sujet sur lequel se penche en partie la présente thèse. Tel que décrit précédemment, la capacité de la PPG de distinguer les individus présentant un intérêt sexuel envers des enfants de ceux qui n'en ont pas repose sur les stimuli utilisés durant l'évaluation. C'est de cette dépendance que provient une partie de la faiblesse de la PPG. En effet, malgré une bonne spécificité (*c.-à-d.* capacité à catégoriser correctement les individus sans intérêts déviants), la PPG présente des limites sur le plan de sa sensibilité (*c.-à-d.* capacité d'identifier les individus avec des intérêts déviants) comme le rapporte certaines études (*c.-à-d.* une sensibilité variant entre 41 % et 61 %; Blanchard, Klassen, Dickey, Kuban et Blak, 2001; Freund et Blanchard, 1989; Malcolm et al., 1993). Cette sensibilité peut être mise en lien avec un faible taux de réponses enregistré chez certains participants. Par exemple, jusqu'à 34 % des participants évalués avec la PPG circonférentielle ne présenteront pas de réponses d'une assez grande amplitude pour être considérées valides et donc interprétables (Akerman et Beech, 2012;

Marshall, Barbaree et Christophe, 1986). Malgré le fait que plusieurs raisons, aient été mises de l'avant pour expliquer le faible taux de réponse, comme la falsification des résultats, l'anxiété et le manque d'intérêt suscité par les stimuli, ces limites demeurent toujours problématiques dans le domaine de l'évaluation des intérêts sexuels (Kuban, Barbaree, et Blanchard, 1999; Marshall, 2013; Marshall et Fernandez, 2000; Wilson et Miner, 2016). Par mesure préventive et surtout par crainte de générer des faux positifs, plusieurs laboratoires déclarent invalides les résultats démontrant de faibles taux de réponse (Barbaree, Marshall, et Lanthier, 1979; Marshall et Fernandez, 2000). C'est pourquoi plusieurs études suggèrent qu'un score valide devrait représenter au moins 10 % d'une érection complète (Lykins et al., 2010; Malcolm et al., 1993; Marshall et al., 1986; Proulx et al., 1997). Afin de simplifier la procédure, certains laboratoires font plutôt usage d'une amplitude minimale pour évaluer la validité des résultats. Ce score peut varier de 1 mm à 3 mm avec l'amplitude minimale la plus répandue fixé à 2,5 mm (Association for the Treatment of Sexual Abusers [ATSA], 2001; Earls, 1983; Marshall, 2013). Ce 2,5 mm correspond au changement minimal en PPG circonférentielle pour obtenir des résultats comparables (*c.-à-d.* $r > .80$) en PPG volumétrique (Kuban et al., 1999). De plus, Lykins et ses collègues (2010) ont démontré que l'utilisation d'un score minimal de 10% d'une érection complète permet d'améliorer la fidélité test-retest. Il s'avère donc nécessaire d'utiliser des stimuli ayant la capacité de susciter de fortes réactions péniennes pour générer des résultats valides chez les individus ou d'ajouter une seconde mesure capable de pallier cette faible sensibilité. Il est à noter qu'il existe cependant un débat dans la littérature quant à l'utilité d'avoir recours à un seuil minimal de changement de circonférence pour interpréter la validité des données (Murphy et al., 2015). C'est pourquoi certains laboratoires interprètent toutes les données obtenues (Harris, Rice, Quinsey, Chaplin,

et Earls, 1992) tandis que d'autres utilisent la valeur obtenue au stimulus neutre comme seuil de validité (p.ex., Seto et Lalumière, 2001). Enfin, d'autres laboratoires comme le nôtre utilisent un seuil de validité minimale à 2.5 mm. En vertu de ce débat, l'utilisation d'un seuil de validité minimale ne peut être écarté.

Finalement, une limite plus récente, abordée à plusieurs reprises dans le premier chapitre évoque les restrictions éthiques associées à l'utilisation de la PPG. Plus précisément, l'utilisation de photographies d'enfants est découragée ou même proscrite lors d'évaluation PPG par crainte de victimiser les enfants qui y sont présentés (Abel et al., 1998; Kalmus et Beech, 2005; Launay, 1994). Cette limite encourage donc le développement de nouveaux stimuli (PGO) et de mesures (PUP) éthiquement acceptables présentant une bonne capacité de classification.

En somme, lors de l'utilisation de l'évaluation phallométrique il importe d'avoir recours aux stimuli permettant l'obtention de fortes amplitudes péniennes et permettant une bonne discrimination. De plus, les données brutes obtenues devraient être analysées en ayant recours aux méthodes les mieux validées, soient les indices de déviance calculés à partir des amplitudes péniennes standardisées (scores z).

Évaluation visuelle de l'intérêt sexuelle.

Tel que mentionné au début du chapitre, la PPG se concentre exclusivement sur l'évaluation de la phase génitale de l'intérêt sexuel (3^e phase), mais stimule fort probablement les deux autres phases par l'entremise des stimuli utilisés. Une mesure complémentaire évaluant la première phase (esthétique) ou la deuxième phase (approche) serait donc utile afin d'obtenir une évaluation plus juste ou du moins complémentaire de cette réponse génitale. À

cette fin, différentes mesures de la phase esthétique ont été proposées au cours des années incluant le temps de réaction à une image, plus communément appelée *viewing time* (Rosenzweig, 1942; Wright & Adams, 1994) et la mesure de la dilatation de la pupille (Kalmus et Beech, 2005).

Tout d'abord introduite au domaine de la mesure des intérêts sexuels par Rosenzweig (1942), le *viewing time* se base, selon certains auteurs sur le principe d'interférence créé par une image sexuelle attirante pour tenter de mesurer indirectement la phase esthétique de l'excitation sexuelle (Wilson et Miner, 2016). La mesure consiste généralement en l'assignation d'une tâche : cibler un point blanc sur une image (Wright et Adams, 1994) ou évaluer l'attraction d'une image (Abel, Jordan, Hand, Holland, et Phipps, 2001) lors de la présentation d'un stimulus d'interférence (c.à.d., un stimulus sexuellement attrayant). Le *viewing time* évalue donc le temps pris par le participant pour accomplir la tâche de manière similaire à ce qui est fait dans une tâche de Stroop (MacLeod, 1991). L'aspect discriminatif de la mesure provient des images qui sont employées comme source d'interférence pour accomplir la tâche. Ces images présentent une variété d'individus (vêtus ou non) de sexe différent et qui varient en âge (enfants et adultes). Les auteurs postulent que plus le stimulus présenté est sexuellement attrayant, plus l'interférence qu'il génère sera grande et plus le temps de réaction (*viewing time*) sera long. En comparant les temps de réaction à différents types de stimuli (enfants et adultes), il devient alors possible de statuer quant aux intérêts sexuels des participants se soumettant au test (Abel, Jordan, Hand, Holland, et Phipps, 2001; Wright et Adams, 1994). Cette mesure a été popularisée au début des années 2000 par l'outil commercial *Abel Assessment for Sexual Interest (AASI; 2001)* de la compagnie *Abel Screening Incorporated*. Cet outil a démontré son efficacité à évaluer les intérêts sexuels des

participants envers les adultes et les enfants (Abel et al., 2001; Kalmus et Beech, 2005) avec des résultats comparables à ceux obtenus en PPG (Harris, Rice, Quinsey et Chaplin, 1996). De plus, cette mesure a démontré une bonne consistance interne (Krueger, Bradford et Glancy, 1998).

Néanmoins, le *viewing time* présente des limites non négligeables dont trois qui sont maintenant abordées. La première concerne l'aspect indirect et transparent de la mesure qui la rend susceptible au contrôle des résultats (Fischer et Smith, 1999; Kalmus et Beech, 2005). En ce sens, l'AASI dépend du fait que le participant ait conscience de l'image, la reconnaisse, l'évalue cognitivement et juge qu'elle est attrayante ou non, et enfin qu'il choisisse d'utiliser sa main pour coter l'image sur un appareil physique. Suivant l'étape de prise de conscience qui s'apparenterait davantage à la phase esthétique le participant peut exercer un contrôle sur sa réponse et influencer ses résultats. Le second désavantage, plus fondamental, fait état de l'absence de compréhension des mécanismes soutenant *viewing time* (Wilson et Miner, 2016). Par exemple, il n'est toujours pas clair si le délai dans la réponse est engendré par la distraction que génère l'image sexuelle ou par l'effort cognitif des agresseurs sexuels d'enfants pour refouler cet intérêt. Le dernier désavantage, moins incommode, mais toutefois présent est que cette mesure peut être difficilement pairée à une autre mesure des intérêts sexuels. Les mesures physiologiques des intérêts sexuels telles que la PPG ou l'oculométrie (Goyette et al., 2012; Trottier et al., 2013) de même que les nouvelles technologies comme l'imagerie par résonance magnétique fonctionnelle (IRMF; Green, Chapman, Kalaska, et Lepore, 2011) ou l'électroencéphalographie (EEG; Waismann, Fenwick, Wilson, Hewett, et Lumsden, 2003) dépendent du fait que chaque stimulus soit présenté pour la même période de temps et sans distraction. À l'opposé, les mesures de type

viewing time, comme l'AASI ou l'Affinity (Glasgow, Osborne, et Croxen, 2003) dépendent du fait que les images ne sont pas observées pour la même période de temps et qu'elles servent de distraction. Il apparaît donc qu'une mesure plus directe de la phase esthétique fonctionnant sur des principes de présentation similaires aux autres mesures physiologiques, serait à prioriser.

Similairement au *viewing time*, il est proposé dans la littérature que la pupillométrie (PUP) serait elle aussi une mesure de la phase esthétique de l'intérêt sexuel (Kalmus et Beech, 2005). Cette mesure, qui évalue les variations de diamètre de la pupille, a été introduite au domaine de la sexualité dans les années 60 et 70 (Hess, Beaver et Shroul, 1975; Hess, Seltzer et Shlien, 1965). Les études préliminaires ont alors suggéré que la PUP pourrait être utilisée pour discriminer et évaluer la préférence de divers intérêts sexuels présents chez des participants. Ces recherches ont connu un certain succès en démontrant qu'il était possible d'utiliser les résultats PUP pour distinguer les hommes homosexuels et hétérosexuels, les hommes et les femmes hétérosexuelles, et les stimuli sexuels et non sexuels (Aboyoun et Dabbs, 1998; Atwood et Howell, 1971; Hess et Polt, 1960; Hess et al., 1965; Nunnally, Knott, Duchnowski et Parker, 1967; Peavler et McLaughlin, 1967; Simms, 1967).

Néanmoins, la méthodologie des premières études et les hypothèses les supportant ont été hautement critiquées résultant en un abandon de cette mesure physiologique pendant plusieurs années (Janisse, 1973; Zuckerman, 1971). Ce fut uniquement vers les années 2000 que la pupillométrie a refait son apparition de manière plus soutenue dans la littérature. Ces nouvelles études présentant un contrôle méthodologique plus rigoureux (*c.-à-d.* contrôle de la luminosité, amélioration des technologies et amélioration substantielle du traitement des données) ont démontré, chez des participants normatifs (hommes et femmes), une plus grande dilatation de la pupille lors de la présentation de stimuli correspondant à leur(s) orientation(s)

sexuelle(s) qu'à ceux ne correspondant pas à leur orientation sexuelle (*c.-à-d.* aux stimuli hétérosexuels pour les participants hétérosexuels et aux stimuli homosexuels pour les participants homosexuels; Rieger et al., 2015; Rieger et Savin-Williams, 2012; Rieger, Savin-Williams, Chivers et Bailey, 2015). Cependant, ces études plus récentes ont toujours fait usage de stimuli présentant seulement des adultes. À notre connaissance, seulement une étude récente a tenté de présenter des stimuli d'enfants et d'adultes à un groupe d'hommes et de femmes hétérosexuels sans problématique sexuelle lors d'une évaluation pupillométrique. Leurs résultats prometteurs démontrent une dilatation significativement plus grande de la pupille pour les hommes lors de la présentation d'images de femmes, et ce, en comparaison avec des images d'hommes et d'enfants. Similairement, les femmes ont présenté une plus forte dilatation de la pupille aux images d'hommes qu'aux images de femmes et d'enfants (Attard-Johnson, Bindemann et Ciardha, 2016). Néanmoins, les progrès sur cette mesure ont été parsemés de critiques et la méthodologie l'entourant a rapidement dû être modifiée.

Limites et évolution de la pupillométrie. Plusieurs auteurs ont affirmé que les résultats obtenus par Hess et ses collaborateurs (1965, 1975) étaient, entre autres, le résultat d'un manque de contrôle dans les stimuli utilisés (Janisse, 1973; Zuckerman, 1971). Ces critiques concernaient plus particulièrement le manque de contrôle quant à la luminescence des images qui étaient utilisées en PUP. L'exemple le plus notable de ce manque de contrôle fut le recours à une image grise comme stimulus neutre et d'images complexes avec des contrastes de couleurs comme stimuli sexuels, sans aucun contrôle pour la différence de luminescence entre les stimuli (Hess et al., 1965). Aujourd'hui, non seulement ce type de faiblesses méthodologiques a largement diminué dans les études plus récentes, mais les hypothèses soutenant la méthodologie ont aussi évolué en fonction des nouvelles données disponibles.

Initialement, Hess et ses collaborateurs (1965, 1975) ont proposé une compréhension simple des résultats PUPs voulant qu'une dilatation pupillaire soit le résultat d'un intérêt envers un stimulus et qu'une contraction pupillaire soit le résultat d'un désintérêt ou dédain envers un stimulus. C'était donc la valence, positive ou négative, que suscitait un stimulus chez un participant qui était responsable de la contraction ou de la dilatation de la pupille. Cette hypothèse, infirmée au cours des années subséquentes (Janisse, 1973), avait l'avantage d'être simple et présentait le potentiel de la pupillométrie en tant que mesure de discrimination des intérêts sexuels. En ce sens, notre compréhension de la PUP a été modifiée et raffinée au cours des années en prenant compte des nouvelles données disponibles. Plusieurs chercheurs, de différents domaines, en sont venus à considérer que la dilatation de la pupille était plutôt associée à la force de l'émotion suscitée par un stimulus et non à la valence de l'émotion suscitée par ce stimulus (Janisse, 1973). De telle sorte qu'une image suscitant une forte réaction émotionnelle, qu'elle soit positive ou négative, attirante ou repoussante, sera associée à une plus grande dilatation de la pupille. Similairement, la dilatation sera plus importante lorsque le stimulus présenté requiert un effort cognitif de traitement de l'information (Bradley, Miccoli, Escrig et Lang, 2008; Mathôt et Van der Stigchel, 2015).

Or, si la dilatation de la pupille est associée à la force de l'émotion ou la force de la stimulation et non à sa valence, comment utiliser cette mesure pour discriminer des intérêts sexuels comme démontré précédemment (Attard-Johnson et al., 2016; Rieger et al., 2015; Rieger et Savin-Williams, 2012; Rieger, Savin-Williams, Chivers et Bailey, 2015). Une des réponses possibles à cette question se trouve dans les recherches plus fondamentales sur la fonction de la dilatation de la pupille. Par exemple, Bradley et ses collaborateurs (2008) stipulent que la dilatation et la constriction automatique de la pupille permettent, entre autres,

un meilleur équilibre entre la sensibilité visuelle et l'acuité visuelle. Ainsi, en se dilatant, la pupille permet à une plus grande quantité de lumière d'atteindre la rétine et augmente les chances que des éléments périphériques (*p. ex.*, ressources) soient captés et interprétés par le cerveau (sensibilité). À l'opposé, la constriction de la pupille diminue la quantité de lumière à atteindre la rétine, mais augmente le niveau de détails perceptibles et améliore la vision en profondeur (spécificité). Bradley et ses collaborateurs (2008) lient cette réaction automatique à la proposition d'Aston-Jones et Cohen (2005) stipulant qu'il existe deux modes d'utilisation de la pupille, l'exploration et l'exploitation. Par exemple, lors d'une phase d'exploration, l'objectif principal consiste à être le plus alerte possible afin de trouver des ressources pertinentes (*p. ex.*, nourriture ou partenaire sexuelle). Il est alors avantageux que la pupille soit la plus dilatée possible ou la plus sensible possible à des stimuli potentiels. À l'opposé, lorsqu'un individu est concentré sur l'exécution d'une tâche spécifique, la dilatation est plus faible et l'importance est accordée à une meilleure précision de la vue ou spécificité. En ce sens, lorsqu'un individu est confronté à une image générant une plus grande réaction émotionnelle, ou dans notre cas à une plus forte réaction esthétique, la pupille devrait se dilater pour augmenter la sensibilité de l'œil et augmenter les chances de percevoir d'autres stimuli importants. Conséquemment, la présentation d'un stimulus sexuellement attrayant devrait générer une réponse esthétique automatique caractérisée par une dilatation de la pupille plus grande que celle obtenue pour un stimulus moins sexuellement attrayant. De plus, puisque cette réaction est automatique, il est probable qu'elle soit hautement circonscrite dans le temps, se produisant dans les premières secondes de présentation d'un stimulus. Une fois le stimulus interprété cognitivement, l'individu subirait alors les autres changements physiologiques associés à l'intérêt sexuel décrit précédemment, permettant l'évaluation

pléthysmographique. Contrairement au *viewing time*, le fonctionnement de la PUP permettrait une évaluation plus directe de la phase esthétique et pourrait être combinée à une évaluation PPG standard.

Agencement des données pupilométriques et pléthysmographiques.

Les données obtenues à l'aide de la PUP et de la PPG permettraient donc d'évaluer les intérêts sexuels d'un individu, que ces intérêts soient considérés déviants ou non, en mesurant chacune une phase complémentaire du modèle de Singer (1984). La PUP rendrait ainsi possible la mesure de la première phase de l'intérêt sexuel (phase esthétique) qui est considérée par plusieurs comme étant cognitive et automatique (Kalmus et Beech, 2005; Trottier et al., 2015). En revanche, la PPG permettrait l'évaluation de la phase génitale, soit la dernière phase de l'intérêt sexuel (Singer 1984). Ensemble ces deux mesures rendraient la prédiction de la récurrence sexuelle possible puisque la présence d'intérêt sexuel déviant demeure un facteur central et commun dans la majorité des théories explicatives de l'agression sexuelle (McPhail et al., 2017). Par exemple, l'intérêt sexuel et les phases de l'excitation sexuelle s'inscrivent plus largement dans la théorie intégrée de l'agression sexuelle (Ward et Beech, 2006). Selon cette théorie, l'agression sexuelle se comprend dans une perspective évolutive et développementale dans laquelle trois facteurs (biologique, écologique et neuropsychologique) interagissent pour modéliser l'intérêt sexuel, les fantasmes, les émotions et les comportements sexuels d'un individu. En ce sens, les prédispositions génétiques et neurologiques d'un individu sont actualisées par des facteurs écologiques (proximaux et distaux) tels que les facteurs sociaux, les circonstances personnelles d'un individu et les facteurs environnementaux. Une fois actualisés, ils agissent sur les fonctions neuropsychologiques (perception, motivation, prise de décisions) et expliquent le passage à

l'acte. Un des avantages non négligeables de cette théorie est son aspect dynamique par lequel les intérêts sexuels déviants ou non peuvent être maintenus, exacerbés et même diminués par les facteurs écologiques présents dans la vie d'un individu. Ainsi, cette théorie permet de prendre en compte les recherches démontrant qu'un stimulus n'est pas naturellement sexuellement attrayant, mais le devient par différents processus d'apprentissage qui associent ce stimulus à une réponse sexuelle et l'intègre dans le processus dynamique décrit précédemment (Spiering, Everaerd et Laan, 2004).

En somme, la pupillométrie et la pléthysmographie pénienne sont deux procédures qui tentent de prendre une image statique des fonctions neuropsychologiques d'un individu qui se trouve dans un contexte dynamique. Ces méthodes y arrivent en ayant recours à des stimuli qui peuvent par apprentissage devenir des représentations des intérêts et des désirs d'un individu. Par sa théorie de l'intérêt sexuel, Singer (1984) établit des étapes distinctes dans l'excitation sexuelle et suggère différents marqueurs physiologiques permettant de mesurer les facteurs neuropsychologiques. Finalement, les études mentionnées dans les sections précédentes offrent des indices quant aux caractéristiques favorables (stimulations multimodales) pour activer les fonctions neuropsychologiques d'un individu et permettre la mesure des intérêts sexuels chez un individu.

Objectifs/Hypothèses de la recherche

Notre première hypothèse est que la présentation des personnages générés par ordinateur, jumelée à des histoires auditives, permettent la création de stimuli éthiquement responsables n'enfreignant pas les lois ou les conventions canadiennes capables de susciter davantage de réactions aux phases esthétique et génitale de l'intérêt sexuel tout en réduisant le

nombre de profils invalides en PPG. Parallèlement, en ajoutant la pupillométrie à la pléthysmographie pénienne standard, il est attendu que l'obtention d'une mesure physiologique complémentaire et moins invasive permettra de mieux discriminer la présence d'intérêts sexuels envers des enfants.

Pour ce faire, cette recherche doctorale est divisée en deux études, répondant à deux questions distinctes. La première étude (chapitre II) a pour but d'examiner si l'utilisation de stimuli audiovisuels incorporant des PGO et des histoires auditives résultera en une modalité suscitant de plus fortes amplitudes péniennes, un plus grand nombre de profils valides et une meilleure validité discriminante que l'utilisation de PGO ou de stimuli auditifs présentés individuellement. La deuxième étude (chapitre III) a pour but d'examiner si cette nouvelle modalité audiovisuelle peut être utilisée avec la PUP pour discriminer les individus ayant commis une agression sexuelle envers des enfants d'individus de la population générale. Il est attendu que les participants ayant déjà commis une agression sexuelle envers les enfants présentent une plus forte dilatation de la pupille lors de la présentation de stimuli présentant des enfants que durant la présentation de stimuli présentant des adultes. Similairement, les participants n'ayant jamais commis d'agression sexuelle envers des enfants devraient réagir plus fortement aux stimuli d'adultes qu'aux stimuli d'enfants. Si cette distinction dans les patrons de réponses est maintenue, il est attendu que les résultats PUP puissent être utilisés pour établir des indices d'intérêts sexuels similaires à ce qui est possible avec les résultats de PPG, permettant ainsi de discriminer les individus entre ces deux groupes. Afin d'accomplir ces deux études sans avoir à exposer les participants deux fois aux mêmes stimuli, les deux mesures (PPG et PUP) nécessaires pour les deux expérimentations ont été prises simultanément lors de la passation des stimuli auditifs, visuels et audiovisuels. Enfin, il est à

noter qu'afin de faciliter la lecture de la thèse, les références utilisées seront placées à la fin de leur chapitre respectif (I à IV).

Références

- Abel, G., Barlow, D., Blanchard, E. et Mavissakalian, M. (1975). Measurement of sexual arousal in male homosexuals: Effects of instructions and stimulus modality. *Archives of Sexual Behavior*, 4(6), 623-629. doi:10.1007/BF01544270
- Abel, G., Blanchard, E. B. et Barlow, D. H. (1981). Measurement of sexual arousal in several paraphilias: The effects of stimulus modality, instructional set and stimulus content on the objective. *Behaviour Research and Therapy*, 19(1), 25-33.
doi:http://dx.doi.org/10.1016/0005-7967(81)90109-1
- Abel, G., Huffman, J., Warberg, B. et Holland, C. (1998). Visual Reaction Time and Plethysmography as Measures of Sexual Interest in Child Molesters. *Sexual Abuse: A Journal of Research and Treatment*, 10(2), 81-95. doi:10.1023/a:1022063214826
- Abel, G., Jordan, A., Hand, C. G., Holland, L. A., & Phipps, A. (2001). Classification models of child molesters utilizing the Abel Assessment for sexual interest TM. *Child Abuse & Neglect*, 25(5), 703-718.
- Aboyoun, D. C. et Dabbs, J. M. (1998). The Hess pupil dilation findings: Sex or novelty? *Social Behavior and Personality: an international journal*, 26(4), 415-419.
doi:10.2224/sbp.1998.26.4.415
- Akerman, G. et Beech, A. R. (2012). A systematic review of measures of deviant sexual interest and arousal. *Psychiatry, Psychology and Law*, 19(1), 118-143.
doi:10.1080/13218719.2010.547161
- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.

- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.
- Association for the Treatment of Sexual Abusers. (2001). *Practice standards and guidelines for the members of the Association for the Treatment of Sexual Abusers*. Beaverton, OR: Author.
- Aston-Jones, G. et Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annu. Rev. Neurosci.*, 28, 403-450.
- Attard-Johnson, J., Bindemann, M. et Ciardha, C. Ó. (2016). Pupillary Response as an Age-Specific Measure of Sexual Interest. *Archives of Sexual Behavior*, 45(4), 855-870. doi:10.1007/s10508-015-0681-3
- Atwood, R. W. et Howell, R. J. (1971). Pupillometric and personality test score differences of female aggressing pedophiliacs and normals. *Psychonomic Science*, 22(2), 115-116. doi:10.3758/bf03332522
- Barbaree, H. E. et Marshall, W. L. (1988). Deviant sexual arousal, offense history, and demographic variables as predictors of reoffense among child molesters. *Behavioral Sciences & the Law*, 6(2), 267-280. doi:10.1002/bsl.2370060209
- Barbaree, H., Marshall, W., et Lanthier, R. (1979). Deviant sexual arousal in rapists. *Behaviour Research and Therapy*, 17(3), 215-222.
- Barbaree, H., et Mewhort, D. (1994). The effects of the z-score transformation on measures of relative erectile response strength: A re-appraisal. *Behaviour Research and Therapy*, 32(5), 547-558.

- Blanchard, R., Klassen, P., Dickey, R., Kuban, M. E. et Blak, T. (2001). Sensitivity and specificity of the phallometric test for pedophilia in nonadmitting sex offenders. *Psychological Assessment, 13*(1), 118-126. doi:10.1037/1040-3590.13.1.118
- Blanchard, R., Kolla, N., Cantor, J., Klassen, P., Dickey, R., Kuban, M. et Blak, T. (2007). IQ, handedness, and pedophilia in adult male patients stratified by referral source. *Sexual Abuse: A Journal of Research and Treatment, 19*(3), 285-309. doi:10.1007/s11194-007-9049-0
- Blanchard, R., Kuban, M., Blak, T., Klassen, P., Dickey, R. et Cantor, J. (2012). Sexual attraction to others: A comparison of two models of alloerotic responding in men. *Archives of Sexual Behavior, 41*(1), 13-29. doi:10.1007/s10508-010-9675-3
- Blanchard, R., Kuban, M. E., Blak, T., Cantor, J. M., Klassen, P. E. et Dickey, R. (2009). Absolute versus relative ascertainment of pedophilia in men. *Sexual Abuse: A Journal of Research and Treatment, 21*(4), 431-441. doi:10.1177/1079063209347906
- Blanchard, R., Lykins, A., Wherrett, D., Kuban, M., Cantor, J., Blak, T., ... Klassen, P. (2009). Pedophilia, hebephilia, and the DSM-V. *Archives Sexual Behavior, 38*(3), 335-350. doi:10.1007/s10508-008-9399-9
- Bradley, M. M., Miccoli, L., Escrig, M. A. et Lang, P. J. (2008). The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology, 45*(4), 602-607. doi:10.1111/j.1469-8986.2008.00654.x
- Card, R. D., & Farrall, W. (1990). Detecting faked penile responses to erotic stimuli: A comparison of stimulus conditions and response measures. *Annals of sex research, 3*(4), 381-396.

- Chaplin, T. C., Rice, M. E. et Harris, G. T. (1995). Salient victim suffering and the sexual responses of child molesters. *Journal of Consulting and Clinical Psychology*, 63(2), 249-255. doi:10.1037/0022-006X.63.2.249
- Chartier, S., Renaud, P., Bouchard, S., Proulx, J., Rouleau, J. L., Fedoroff, P. et Bradford, J. P. (2006). Sexual preference classification from gaze behavior data using a multilayer perceptron. *Annual Review of CyberTherapy and Telemedecine*, 4, 149-157.
- Clift, R. J. W., Rajlic, G. et Gretton, H. M. (2009). Discriminative and predictive validity of the penile plethysmograph in adolescent sex offenders. *Sexual Abuse: A Journal of Research and Treatment*, 21(3), 335-362. doi:10.1177/1079063209338491
- Dennis, E., Rouleau, J.-L., Renaud, P., Nolet, K. et Saumur, C. (2014). A pilot development of virtual stimuli depicting affective dispositions for penile plethysmography assessment of sex offenders. *The Canadian Journal of Human Sexuality*, 23(3), 200-208. doi:10.3138/cjhs.2529
- Dombert, B., Mokros, A., Brückner, E., Schlegl, V., Antfolk, J., Bäckström, A. et Santtila, P. (2013). The virtual people set: developing computer-generated stimuli for the assessment of pedophilic sexual interest. *Sexual Abuse: A Journal of Research and Treatment*, 25(6), 557-582. doi: 10.1177/1079063212469062
- Earls, C. M., Quinsey, V. L., et Castonguay, L. G. (1987). A comparison of three methods of scoring penile circumference changes. *Archives of Sexual Behavior*, 16(6), 493-500.
- Farrall, W. R. et Card, R. D. (1988). Advancements in physiological evaluation of assessment and treatment of the sexual aggressor. *Annals of the New York Academy of Sciences*, 528(1), 261-273. doi: 10.1111/j.1749-6632.1988.tb50869.x

- Fernandez, Y. (2002). Phallometric testing with sexual offenders against female victims: Limits to its value. *Paper presented at the Forum on Corrections Research*.
- Fischer, L., & Smith, G. (1999). Statistical Adequacy of the Abel Assessment for Interest in Paraphilias. *Sexual Abuse: A Journal of Research and Treatment*, 11(3), 195- 205.
doi:10.1023/A:1021308224123
- Freund, K. (1963). A laboratory method for diagnosing predominance of homo-or hetero-erotic interest in the male. *Behaviour Research and Therapy*, 1(1), 85-93.
doi:10.1016/0005-7967(63)90012-3
- Freund, K. (1965). Diagnosing heterosexual pedophilia by means of a test for sexual interest. *Behaviour Research and Therapy*, 3(4), 229-234. doi:10.1016/0005-7967(65)90031-8
- Freund, K. (1976). Assessment of anomalous erotic preferences in situational impotence. *Journal of sex & marital therapy*, 2(3), 173-183.
- Freund, K. et Blanchard, R. (1989). Phallometric diagnosis of pedophilia. *Journal of Consulting and Clinical Psychology*, 57(1), 100-105. doi:10.1037/0022-006X.57.1.100
- Freund, K., Scher, H., Racansky, I., Campbell, K., et Heasman, G. (1986). Males disposed to commit rape. *Archives of Sexual Behavior*, 15(1), 23-35.
- Freund, K., Watson, R., Dickey, R. et Rienzo, D. (1991). Erotic gender differentiation in pedophilia. *Archives of Sexual Behavior*, 20(6), 555-566. doi:10.1007/BF01550954
- Freund, K., Watson, R., et Rienzo, D. (1988). Signs of feigning in the phallometric test. *Behaviour Research and Therapy*, 26(2), 105-112.
- Gaither, G. A. et Plaud, J. J. (1997). The effects of secondary stimulus characteristics on men's sexual arousal. *The Journal of Sex Research*, 34(3), 231-236.

doi:10.1080/00224499709551890

Glasgow, D. V., Osborne, A., & Croxen, J. (2003). An assessment tool for investigating paedophile sexual interest using viewing time: An application of single case methodology. *British Journal of Learning Disabilities*, 31(2), 96-102.

Goyette, M., Rouleau, J.-L. et Renaud, P. (2012). *Évaluation des intérêts sexuels auprès d'agresseurs sexuels d'enfants par la vidéooculographie et la présentation de stimuli générés par ordinateur* (doctoral dissertation).

Repéré à <http://hdl.handle.net/1866/8714>

Green, A., Chapman, C., Kalaska, J., & Lepore, F. (2011). Real-time functional magnetic imaging—brain—computer interface and virtual reality: promising tools for the treatment of pedophilia. *Enhancing Performance for Action and Perception: Multisensory Integration, Neuroplasticity and Neuroprosthetics*, 263.

Gress, C. L. (2005). Viewing time measures and sexual interest: Another piece of the puzzle. *Journal of Sexual Aggression*, 11(2), 117-125.

Hall, G. C. N., Hirschman, R. et Oliver, L. L. (1995). Sexual arousal and arousability to pedophilic stimuli in a community sample of normal men. *Behavior Therapy*, 26(4), 681-694. doi:10.1016/S0005-7894(05)80039-5

Hanson, R. K., Babchishin, K. M., Helmus, L., et Thornton, D. (2013). Quantifying the relative risk of sex offenders: Risk ratios for Static-99R. *Sexual Abuse: A Journal of Research and Treatment*, 25, 482-515. doi: 10.1177/1079063212469060

- Hanson, R. K. et Bussière, M. T. (1998). Predicting relapse: A meta-analysis of sexual offender recidivism studies. *Journal of Consulting and Clinical Psychology*, 66(2), 348-362. doi:10.1037/0022-006X.66.2.348
- Hanson, R. K., Harris, A. J., Scott, T. L., & Helmus, L. (2007). *Assessing the risk of sexual offenders on community supervision: The Dynamic Supervision Project* (Vol. 5, No. 6). Ottawa, Ontario: Public Safety Canada.
- Hanson, R. K. et Morton-Bourgon, K. E. (2005). The Characteristics of Persistent Sexual Offenders: A Meta-Analysis of Recidivism Studies. *Journal of Consulting and Clinical Psychology*, 73(6), 1154-1163. doi:10.1037/0022-006X.73.6.1154
- Hanson, R. K., & Morton-Bourgon, K. E. (2009). The accuracy of recidivism risk assessments for sexual offenders: a meta-analysis of 118 prediction studies: American Psychological Association.
- Hanson, R. K., & Thornton, D. (2000). Improving risk assessments for sex offenders: a comparison of three actuarial scales. *Law and Human Behavior*, 24(1), 119.
- Harris, G., Lalumière, M., Seto, M., Rice, M. et Chaplin, T. (2012). Explaining the Erectile Responses of Rapists to Rape Stories: The Contributions of Sexual Activity, Non-Consent, and Violence with Injury. *Archives of Sexual Behavior*, 41(1), 221-229. doi:10.1007/s10508-012-9940-8
- Harris, G. T., Rice, M. E., Quinsey, V. L., Chaplin, T. C. et Earls, C. (1992). Maximizing the discriminant validity of phallometric assessment data. *Psychological Assessment*, 4(4), 502-511. doi:10.1037/1040-3590.4.4.502

- Harris, G. T., Rice, M. E., Quinsey, V. L., Lalumière, M. L., Boer, D., & Lang, C. (2003). A multisite comparison of actuarial risk instruments for sex offenders. *Psychological Assessment, 15*(3), 413.
- Hess, E. H., Beaver, P. W. et Shrouf, P. E. (1975). Brightness contrast effects in a pupillometric experiment. *Perception & Psychophysics, 18*(2), 125-127.
doi:10.3758/BF03204099
- Hess, E. H. et Polt, J. M. (1960). Pupil size as related to interest value of visual stimuli. *Science, 132*(3423), 349-350. doi:10.1126/science.132.3423.349
- Hess, E. H., Seltzer, A. L. et Shlien, J. M. (1965). Pupil response of hetero-and homosexual males to pictures of men and women: A pilot study. *Journal of Abnormal Psychology, 70*(3), 165. doi:10.1037/h0021978
- Howes, R. J. (2003). Circumferential change scores in phallometric assessment: Normative data. *Sexual Abuse, 15*(4), 365-375.
- Hunter Jr, J. A., et Goodwin, D. W. (1992). The clinical utility of satiation therapy with juvenile sexual offenders: Variations and efficacy. *Annals of sex research, 5*(2), 71-80.
- Imhoff, R., Schmidt, A. F., Nordsiek, U., Luzar, C., Young, A. W., & Banse, R. (2010). Viewing time effects revisited: Prolonged response latencies for sexually attractive targets under restricted task conditions. *Archives of Sexual Behavior, 39*(6), 1275-1288.
- Israel, E., & Strassberg, D. S. (2009). Viewing time as an objective measure of sexual interest in heterosexual men and women. *Archives of Sexual Behavior, 38*(4), 551-558.
- Janisse, M. P. (1973). Pupil size and affect: A critical review of the literature since 1960. *Canadian Psychologist/Psychologie canadienne, 14*(4), 311-329.

doi:10.1037/h0082230

Jamison, P. L., et Gebhard, P. H. (1988). Penis size increase between flaccid and erect states: an analysis of the Kinsey data. *Journal of Sex Research, 24*(1), 177-183.

Julien, E. et Over, R. (1988). Male sexual arousal across five modes of erotic stimulation. *Archives of Sexual Behavior, 17*(2), 131-143. doi:10.1007/BF01542663

Kalmus, E. et Beech, A. R. (2005). Forensic assessment of sexual interest: A review. *Aggression and Violent Behavior, 10*(2), 193-217. doi:10.1016/j.avb.2003.12.002

Krueger, R. B., Bradford, J. M., & Glancy, G. D. (1998). Interest—A Brief Description. *J Am Acad Psychiatry Law, 26*(2).

Kuban, M., Barbaree, H. E. et Blanchard, R. (1999). A comparison of Volume and Circumference Phallometry: Response Magnitude and Method Agreement. *Archives of Sexual Behavior, 28*(4), 345-359. doi:10.1023/a:1018700813140

Lalumière, M. L. et Harris, G. T. (1998). Common questions regarding the use of phallometric testing with sexual offenders. *Sexual Abuse: A Journal of Research and Treatment, 10*(3), 227-237. doi: 10.1023/A:1021369904984

Launay, G. (1994). The phallometric assessment of sex offenders: Some professional and research issues. *Criminal Behaviour and Mental Health, 4*(1), 48-70.
doi:10.1002/cbm.1994.4.1.48

Launay, G. (1999). The phallometric assessment of sex offenders: an update. *Criminal Behaviour and Mental Health, 9*(3), 254-274. doi:10.1002/cbm.317

Laws, D. R. (2009). Penile plethysmography: Strengths, limitations, innovations. *Cognitive approaches to the assessment of sexual interest in sexual offenders, 7-30.*

- Laws, D. R. et Gress, C. L. Z. (2004). Seeing things differently: The viewing time alternative to penile plethysmography. *Legal and Criminological Psychology, 9*(2), 183-196.
doi:10.1348/1355325041719338
- Laws, D. R., Hanson, R. K., Osborn, C. A. et Greenbaum, P. E. (2000). Classification of Child Molesters by Plethysmographic Assessment of Sexual Arousal and a Self-Report Measure of Sexual Preference. *Journal of Interpersonal Violence, 15*(12), 1297-1312.
doi:10.1177/088626000015012004
- Letourneau, E. J. (2002). A comparison of objective measures of sexual arousal and interest: Visual reaction time and penile plethysmography. *Sexual Abuse, 14*(3), 207-223.
- Lundeen, G., Manohar, M., et Parks, C. (1983). Systemic distribution of blood flow in swine while awake and during 1.0 and 1.5 MAC isoflurane anesthesia with or without 50% nitrous oxide. *Anesthesia et Analgesia, 62*(5), 499-512.
- Lykins, A. D., Cantor, J. M., Kuban, M. E., Blak, T., Dickey, R., Klassen, P. E., et Blanchard, R. (2010). Sexual arousal to female children in gynephilic men. *Sexual Abuse, 22*(3), 279-289.
- Malcolm, P. B., Andrews, D. et Quinsey, V. L. (1993). Discriminant and predictive validity of phallometrically measured sexual age and gender preference. *Journal of Interpersonal Violence, 8*(4), 486-501. doi:10.1177/088626093008004004
- Marshall, W. L. (2006). Clinical and research limitations in the use of phallometric testing with sexual offenders. *Sexual Offender Treatment, 1*(1), 1-18.

- Marshall, W. L. (2013). Phallometric Assessments of Sexual Interests: An Update. *Current Psychiatry Reports*, 16(1), 428. doi:10.1007/s11920-013-0428-6
- Marshall, W. L., Barbaree, H. E. et Christophe, D. (1986). Sexual offenders against female children: Sexual preferences for age of victims and type of behaviour. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, 18(4), 424-439. doi: 10.1037/h0079966
- Marshall, W. L. et Fernandez, Y. M. (2000). Phallometric testing with sexual offenders: Limits to its value. *Clinical Psychology Review*, 20(7), 807-822. doi:10.1016/S0272-7358(99)00013-6
- Marshall, W. A., & Tanner, J. M. (1969). Variations in pattern of pubertal changes in girls. *Archives of Disease in Childhood*, 44(235), 291.
- Marshall, W. A., & Tanner, J. M. (1970). Variations in the pattern of pubertal changes in boys. *Archives of Disease in Childhood*, 45(239), 13-23.
- Mathôt, S. et Van der Stigchel, S. (2015). New Light on the Mind's Eye: The Pupillary Light Response as Active Vision. *Current Directions in Psychological Science*, 24(5), 374-378. doi:10.1177/0963721415593725
- McAnulty, R. et Adams, H. (1990). Patterns of sexual arousal of accused child molesters involved in custody disputes. *Archives of Sexual Behavior*, 19(6), 541-556. doi:10.1007/bf01542464
- MacLeod, C. M. (1991). Half a century of research on the Stroop effect: an integrative review. *Psychological Bulletin*, 109(2), 163.

- McPhail, I. V., Hermann, C. A., Fernane, S., Fernandez, Y. M., Nunes, K. L., & Cantor, J. M. (2017). Validity in Phallometric Testing for Sexual Interests in Children: A Meta-Analytic Review. *Assessment*, 1073191117706139.
- Murphy, W., Haynes, M., Stalgaitis, S. et Flanagan, B. (1986). Differential sexual responding among four groups of sexual offenders against children. *Journal of Psychopathology and Behavioral Assessment*, 8(4), 339-353. doi:10.1007/BF00960731
- Murphy, L., Ranger, R., Fedoroff, J. P., Stewart, H., Dwyer, R. G., & Burke, W. (2015). Standardization of penile plethysmography testing in assessment of problematic sexual interests. *The Journal of Sexual Medicine*, 12(9), 1853-1861.
- Nagayama Hall, G. (1991). Sexual arousal as a function of physiological and cognitive variables in a sexual offender population. *Archives of Sexual Behavior*, 20(4), 359-369. doi:10.1007/bf01542616
- Nunnally, J. C., Knott, P. D., Duchnowski, A. et Parker, R. (1967). Pupillary response as a general measure of activation. *Perception & Psychophysics*, 2(4), 149-155. doi:10.3758/BF03210310
- Odeshoo, J. R. (2004). Of penology and perversity: The use of penile plethysmography on convicted child sex offenders. *Temp. Pol. et Civ. Rts. L. Rev.*, 14, 1.
- Peavler, W. S. et McLaughlin, J. P. (1967). The question of stimulus content and pupil size. *Psychonomic Science*, 8(12), 505-506. doi:10.3758/BF03331723
- Proulx, J. (1989). Sexual preference assessment of sexual aggressors. *International Journal of Law and Psychiatry*, 12(4), 275-280. doi:10.1016/0160-2527(89)90019-8

- Proulx, J. (1993). L'évaluation des préférences sexuelles. Dans J. Aubut (dir.), *Les agresseurs sexuels : Théorie, évaluation et traitement* (p. 98-106). Montréal, Canada: De la Chenelière.
- Proulx, J., Pellerin, B., Paradis, Y., McKibben, A., Aubut, J. et Ouimet, M. (1997). Static and dynamic predictors of recidivism in sexual aggressors. *Sexual Abuse: A Journal of Research and Treatment*, 9(1), 7-27. doi: 10.1007/BF02674889
- Quinsey, V. L. et Chaplin, T. C. (1988). Penile responses of child molesters and normals to descriptions of encounters with children involving sex and violence. *Journal of Interpersonal Violence*, 3(3), 259-274. doi:10.1177/0886260888003003001
- Quinsey, V. L., Chaplin, T. C., et Carrigan, W. F. (1979). Sexual preferences among incestuous and nonincestuous child molesters. *Behavior Therapy*, 10(4), 562-565.
- Quinsey, V. L., Rice, M. E. et Harris, G. T. (1995). Actuarial prediction of sexual recidivism. *Journal of Interpersonal Violence*, 10(1), 85-105.
- Renaud, P., Rouleau, J.-L., Granger, L., Barsetti, I. et Bouchard, S. (2004) Measuring Sexual Preferences in Virtual Reality: A Pilot Study. *CyberPsychology & Behavior*, 5(1), 1-9. doi:10.1089/109493102753685836.
- Renaud, P., Rouleau, J.-L., Proulx, J., Trottier, D., Goyette, M., Bradford, J. P. et Bouchard, S. (2010). Virtual characters designed for forensic assessment and rehabilitation of sex offenders: standardized and made-to-measure. *Journal of Virtual Reality and Broadcasting*, 7, 1860-2037. doi:10.20385/1860-2037/7.2010.5
- Rice, M. E., Quinsey, V. L., & Harris, G. T. (1991). Sexual recidivism among child molesters released from a maximum security psychiatric institution. *Journal of*

Consulting and Clinical Psychology, 59(3), 381-386. doi:10.1037/0022-006X.59.3.381

Rieger, G., Cash, B. M., Merrill, S. M., Jones-Rounds, J., Dharmavaram, S. M. et Savin-Williams, R. C. (2015). Sexual arousal: The correspondence of eyes and genitals. *Biological psychology*, 104, 56-64. doi:10.1016/j.biopsycho.2014.11.009

Rieger, G. et Savin-Williams, R. C. (2012). The eyes have it: Sex and sexual orientation differences in pupil dilation patterns. *PloS one*, 7(8), e40256. doi:10.1371/journal.pone.0040256

Rieger, G., Savin-Williams, R. C., Chivers, M. L. et Bailey, J. M. (2015). Sexual arousal and masculinity-femininity of women. *Journal of Personality and Social Psychology*, 111(2), 265-283. doi: 10.1037/pspp0000077

Rosen, R. C. et Keefe, F. J. (1978). The Measurement of Human Penile Tumescence. *Psychophysiology*, 15(4), 366-376. doi:10.1111/j.1469-8986.1978.tb01393.x

Rosenzweig, S. (1942). The photoscope as an objective device for evaluating sexual interest. *Psychosomatic Medicine*, 4(2), 150-158.

Sakheim, D. K., Barlow, D. H., Beck, J. G. et Abrahamson, D. J. (1985). A comparison of male heterosexual and male homosexual patterns of sexual arousal. *Journal of Sex Research*, 21(2), 183-198. doi:10.1080/00224498509551257

Seto., & Lalumière. (2001). A Brief Screening Scale to Identify Pedophilic Interests Among Child Molesters. *Sexual Abuse: A Journal of Research and Treatment*, 13(1), 15-25. doi:10.1023/A:1009510328588

- Simms, T. M. (1967). Pupillary response of male and female subjects to pupillary difference in male and female picture stimuli. *Perception & Psychophysics*, 2(11), 553-555.
doi:10.3758/BF03210265
- Singer, B. (1984). Conceptualizing sexual arousal and attraction. *Journal of Sex Research*, 20(3), 230-240. doi:10.1080/00224498409551222
- Spiering, M., Everaerd, W., & Laan, E. (2004). Conscious Processing of Sexual Information: Mechanisms of Appraisal. *Archives of Sexual Behavior*, 33(4), 369-380.
doi:10.1023/B:ASEB.0000028890.08687.94
- Tanner, J. M. et Whitehouse, R. H. (1976). Clinical longitudinal standards for height, weight, height velocity, weight velocity, and stages of puberty. *Archives of Disease in Childhood*, 51(3), 170-179. doi:10.1136/adc.51.3.170
- Thornton, D. et Laws, D. R. (Eds.). (2009). *Cognitive Approaches to the Assessment of Sexual Interest in Sexual Offenders*. United-Kingdom: John Wiley & Sons Ltd.
- Trottier, D., Rouleau, J.-L. et Renaud, P. (2013). *Les apports de la réalité virtuelle et de la vidéo-oculographie à l'évaluation des intérêts sexuels* (Doctoral dissertation). Repéré à <http://hdl.handle.net/1866/9191>
- Trottier, D., Renaud, P., Rouleau, J.-L., Goyette, M., Saumur, C., Boukhalfi, T. et Bouchard, S. (2015). Erratum to: Using immersive virtual reality and anatomically correct computer-generated characters in the forensic assessment of deviant sexual preferences. *Virtual Reality*, 19(3), 303-303. doi:10.1007/s10055-015-0277-1

- Trottier, D., Rouleau, J.-L., Renaud, P. et Goyette, M. (2014). Using eye tracking to identify faking attempts during penile plethysmography assessment. *The Journal of Sex Research, 51*(8), 946-955. doi:10.1080/00224499.2013.832133
- Waismann, R., Fenwick, P. B., Wilson, G. D., Hewett, T. D., & Lumsden, J. (2003). EEG responses to visual erotic stimuli in men with normal and paraphilic interests. *Archives of Sexual Behavior, 32*(2), 135-144.
- Ward, T., & Beech, A. (2006). An integrated theory of sexual offending. *Aggression and violent behavior, 11*(1), 44-63. doi:<https://doi.org/10.1016/j.avb.2005.05.002>
- Ward, T., Polaschek, D., & Beech, A. R. (2006). *Theories of sexual offending*. John Wiley & Sons.
- Wilson, K. M., Abel, G. G., Coyne, B. J., et Rouleau, J.-L. (1992). Sex guilt and paraphilic behavior. *Annals of sex research, 5*(3), 161-170.
- Wilson, R. J., et Miner, M. H. (2016). Measurement of male sexual arousal and interest using penile plethysmography and viewing time *Treatment of sex offenders* (pp. 107-131): Springer.
- Wright, L. W., & Adams, H. E. (1994). Assessment of sexual preference using a choice reaction time task. *Journal of Psychopathology and Behavioral Assessment, 16*(3), 221-231.
- Zappalà, A., Antfolk, J., Bäckström, A., Dombert, B., Mokros, A. et Santtila, P. (2013). Differentiating sexual preference in men: Using dual task rapid serial visual presentation task. *Scandinavian Journal of Psychology, 54*(4), 320-327. doi:10.1111/sjop.12050

Zuckerman, M. (1971). Physiological measures of sexual arousal in the human. *Psychological Bulletin*, 75(5), 297. doi :10.1037/h0030923

Chapitre II - Premier article

**Increasing valid profiles in phallometric assessment of sex offenders with child victims:
combining the strengths of audio stimuli and synthetic characters**

Article accepté pour publication dans

Archives of Sexual Behavior

Shawn Marschall-Lévesque, candidat au Ph. D.

Joanne-Lucine Rouleau, Ph. D.

Patrice Renaud, Ph. D.

Abstract

Penile plethysmography (PPG) is a measure of sexual interests that relies heavily on the stimuli it uses to generate valid results. Ethical considerations surrounding the use of real images in PPG have further limited the content admissible for these stimuli. To palliate this limitation, the current study aims to combine audio and visual stimuli by incorporating computer-generated characters to create new stimuli capable of accurately classifying sex offenders with child victims, while also increasing the number of valid profiles. Three modalities (audio, visual, and audiovisual) were compared using two groups (15 sex offenders with child victims and 15 non-offenders). Both the new visual and audiovisual stimuli resulted in 13% increase in the number of valid profiles at 2.5 mm, when compared to the standard audio stimuli. Furthermore, the new audiovisual stimuli generated a 34% increase in penile responses. All three modalities were able to discriminate between the two groups by their responses to the adult and child stimuli. Lastly, sexual interest indices for all three modalities could accurately classify participants in their appropriate groups, as demonstrated by ROC curve analysis (i.e., Audio AUC = .81, 95% CI [.60, 1.00]; Visual AUC = .84, 95% CI [.66, 1.00] and Audiovisual AUC = .83, 95% CI [.63, 1.00]). Results suggest that computer-generated characters allow accurate discrimination of sex offenders with child victims and can be added to already validated stimuli to increase the number of valid profiles. The implications of audiovisual stimuli using computer-generated characters and their possible use in PPG evaluations are also discussed.

KEY WORDS: sex offenders with child victims; penile plethysmography; computer-generated characters; audio stimuli; visual stimuli.

Introduction

Phallometric assessment or penile plethysmography (PPG) has long been a standard assessment tool of sexual interests (Gaither & Plaud, 1997; Kalmus & Beech, 2005; Lalumière & Harris, 1998; Quinsey & Chaplin, 1988; Zuckerman, 1971). It is used to evaluate atypical sexual interest towards children, an important risk factor for sexual recidivism (Laws, 1984; Malcolm, Andrews, & Quinsey, 1993; Marshall & Fernandez, 2000). This procedure, which uses variations in penile tumescence to ascertain sexual interests towards varying stimuli, is mostly done using one of two validated methods: volumetric PPG or circumferential PPG. As the name suggests, volumetric PPG measures changes in the overall volume size of the participant's penis. This is done by enclosing the penis in a cylinder using a cuff to ensure that the device is airtight and then measuring changes in air pressure as the penis expands. This method is the most sensitive of the two procedures (Kuban, Barbaree, & Blanchard, 1999). Circumferential PPG usually uses a mercury filled rubber stress gauge placed around the penis to measure changes in penile circumference. Although not as precise as the volumetric procedure, circumferential PPG is commonly used for its simplicity (for a comprehensive comparison between both procedures, see Kuban et al. [1999]). Both of these procedures rely on a wide range of sexual stimuli to induce penile tumescence in order to assess the presence of sexual interests. These stimuli can generally be divided in two categories: visual stimuli and audio stimuli.

Visual Stimuli.

Visual stimuli have been used since the 1950s (Freund, 1963) and generally consist of children and adults of different ages in different stages of undress. They enable the

discrimination between sex offenders with child victims and non-offenders, allow the assessment of sexual interests towards children (Murphy, Haynes, Stalgaitis, & Flanagan, 1986), and are a good predictor of sexual recidivism (Barbaree & Marshall, 1988; Malcolm et al., 1993; Quinsey, Rice, & Harris, 1995; Rice, Quinsey, & Harris, 1991). In addition, visual stimuli accurately specify preferred ages of victims (Harris, Rice, Quinsey, Chaplin, & Earls, 1992; Lalumière & Harris, 1998; Launay, 1999; Malcolm et al., 1993; Marshall, Barbaree, & Christophe, 1986; Proulx, 1989). Unfortunately, ethical limitations surrounding the use of real images of children, including the risk of victimizing the children by using their images while they cannot ethically consent, have reduced their use over the years which has giving rise to the use of audio stimuli (Abel, Huffman, Warberg, & Holland, 1998; Kalmus, & Beech, 2005).

Audio Stimuli.

Audio stimuli have been used for several decades and are widely employed in PPG laboratories. They are generally presented as short vignettes narrated in the second person. They describe sexual and non-sexual interactions between an adult male and different characters (i.e., male and female adults and children) with varying levels of consent. Audio stimuli allow for accurate discrimination between sex offenders with child victims and non-offenders (Murphy et al., 1986; Thornton & Laws, 2009). They can assess the presence of various sexual interests, and are a good predictor of sexual recidivism (Proulx et al., 1997; Quinsey et al., 1995). Furthermore, the story base aspect of audio stimuli allows the evaluation of sexual interests towards coercion, violence, and humiliation (Chaplin, Rice, & Harris, 1995; Harris, Lalumière, Seto, Rice, & Chaplin, 2012; Harris et al., 1992; Lalumière & Harris, 1998; Launay, 1999; Laws, Hanson, Osborn, & Greenbaum, 2000; Marshall et al., 1986; Proulx,

1989; Quinsey & Chaplin, 1988). This literature suggests that both visual and audio stimuli can be used to accurately evaluate sexual interests. However, it is necessary to compare the two in order to determine whether one should be prioritized.

Comparing the Different Modalities.

Many different stimuli modalities have been compared over the years with regard to their capacity to generate strong penile responses (Abel, Barlow, Blanchard, & Mavissakalian, 1975; Abel, Blanchard, & Barlow, 1981; Goyette, 2012; Julien & Over, 1988; Kalmus & Beech, 2005; Trottier, Rouleau, Renaud, & Goyette, 2014). One study comparing four modalities (i.e., videotapes, slides, audiotapes, and fantasies) on a group of 48 males (seeking evaluation and/or treatment for their patterns of sexual arousal) found that the video modality generated the strongest penile responses compared to the other three, which did not distinguish themselves from one another (Abel et al., 1981). A study by Julien and Over (1988) arrived at a similar conclusion while comparing five modalities (i.e., videos, images, audio tracks, written texts, and fantasies). The video modality showed the best results, the fantasies obtained the worst results, and the three others showed moderate results. Finally, Gaither and Plaud (1997) found that the audio cues in sexual video clips (i.e., the sound in the clip) contributed to generating significantly greater penile responses than silent clips. These studies suggest that stimuli using more than one source of stimulation at a time (audio and visual) could generate stronger penile responses.

Several laboratories and clinics have come to this conclusion and already use two or more modalities either sequentially (Clift, Rajlic, & Gretton, 2009; Goyette, 2012; Hall, Hirschman, & Oliver, 1995; Laws et al., 2000; Marshall et al., 1986; Trottier, 2013) or concurrently (Blanchard, Klassen, Dickey, Kuban, & Blak, 2001; Blanchard et al., 2007;

Blanchard et al., 2012; Blanchard, et al., 2009; Blanchard, et al., 2009; Farrall & Card, 1988; Freund & Blanchard, 1989; Freund, Watson, Dickey, & Rienzo, 1991; McNulty & Adams, 1990). The importance accorded to the modalities generating the strongest penile response is mostly explained by the need for significant and valid results in PPG assessments. With that said, it is this reliance on stimuli capable of generating strong responses that constitutes one of PPG's weaknesses.

Penile Plethysmography Limitation.

In order to achieve a differential assessment of sexual interests, PPG relies on what can be described as a two-step procedure: (a) The presentation of stimuli (i.e., child and adult) while recording penile sexual responses; (b) the comparison of results according to the type of stimuli presented (child vs adult) to ascertain the presence of sexual interests towards children (Harris et al., 1992; Kalmus & Beech, 2005; Lalumière & Harris, 1998; Rosen & Keefe, 1978). The second step is therefore dependent on the availability of valid results generated by the first step. Valid results using circumferential PPG (mercury in rubber strain gauges method; Rosen & Keefe, 1978) have generally been defined as penile changes of at least 10% of a full erection (Malcolm et al., 1993; Marshall et al., 1986; Proulx et al., 1997). To simplify the assessment, some researchers have used a minimum recorded difference to assess the validity of results. This validity cut-off point can range from a 1-millimetre difference to a 3-millimetre difference (Association for the Treatment of Sexual Abusers [ATSA], 2001; Earls, 1983). A study by Jamison and Gebhard (1988) found that the average North-American penis shows an increase of 27.94 mm between its flaccid and fully erect state. It is therefore reasonable to conclude that an average penile circumference change between 2.5 and 3 mm can be considered valid (i.e., the 10% difference suggested in previous studies). This 2.5 mm

threshold is supported by Kuban et al. (1999) who demonstrated a good convergent validity between circumferential and volumetric plethysmography (i.e., $r > .80$) for participants obtaining a minimum base line change of 2.5 mm. However, this cut-off point can be difficult to achieve with up to 34% of PPG profiles being considered invalid (Akerman & Beech, 2011; Barker & Howell, 1992; Farrall & Card, 1988; Kalmus & Beech, 2005; Marshall et al., 1986; Murphy et al., 1986; Quinsey & Chaplin, 1988; Stinson, Becker, & Sales, 2008). This overall dependence on the stimuli to generate sexual responses and valid results is one of PPG's limitations (Kalmus & Beech, 2005), which underlines the need of having stimuli capable of producing stronger penile responses. The literature suggests that these types of stimuli can be created by combining audio and visual stimuli to create a coherent set of stimuli. Due to the ethical limitations or restrictions associated to photographic stimuli, the use of synthetic stimuli would have to be prioritized. Luckily, one such type of stimuli already exists in the form of computer-generated characters.

Computer-Generated Characters.

Computer-generated characters have been in use for several years, however available data on these characters is relatively new (Goyette, 2012; Renaud et al., 2010; Trottier et al., 2015; Trottier et al., 2014). Although preliminary PPG results did not allow for accurate discrimination between offenders and non-offenders (Renaud et al., 2009), further studies showed that these computer-generated characters were as good if not better than the standard audio stimuli in distinguishing between sexual offenders with child victims and non-offenders (Dombert et al., 2013; Goyette, 2012; Renaud et al., 2010; Trottier, 2013; Trottier et al., 2015). Computer generated characters have also been used in the field of sexuality to distinguish between homosexual participants and heterosexual participants (Zappalà et al., 2013) or to

help identify faking attempts in PPG (Trottier et al., 2014). Finally, they have been used to examine patients' varying visual or penile reactions to different types of characters (Chartier et al., 2006; Dennis, Rouleau, Renaud, Nolet, & Saumur, 2014). One of their strengths is that once they are created, they can be animated and added to different contexts. They can be made to follow storylines and perform different sexual acts without creating or using human victims (ethical shortcoming of real images).

Current Study.

The current study attempts to show that computer-generated characters can be combined with the standardized audio stimuli to create a coherent audiovisual modality that can accurately discriminate between sex offenders with child victims and non-offenders and that can increase penile circumference change, therefore increasing the number of valid PPG profiles. This new modality will be compared to the audio and visual modalities presented individually.

Method

Participants.

Two groups of participants were assessed in the current study. The sex offender group was comprised of 15 adult males; they were recruited either following their court ordered clinical evaluation at the Institut Philippe-Pinel de Montréal (IPPM) or from hand distributed fliers during treatment maintenance sessions for sexual offenders. It was made clear to them that results from the experiment were confidential and had no legal implications. In order to be included in the study, participants had to have been convicted of or have admitted to having had sexual contact with a child who was 13 years old or younger while being at least 16 years

old and at least five years older than their victim (American Psychiatric Association, 2000, 2013). Participants with confirmed erectile dysfunction were not permitted to participate in the study, which eliminated one potential participant. Fourteen of the fifteen participants were familiar with the PPG procedure, having already completed a court order evaluation in the past 10 years.

The comparison group was also comprised of 15 adult males; they were recruited using an advertisement in a local newspaper. Details on the study were provided via telephone, and the date of the evaluation was scheduled at the same time. None of the participants in the comparison group reported having had sexual contact with a child who was 13 years old or younger while being at least 16 years old or reported having ever had pedophilic interests or fantasies. Furthermore, participants in the comparison group reported no history of sexual assault, neither as the perpetrator or victim. None of the participants suffered from erectile dysfunction. Eleven of the fifteen participants in the comparison group were familiar with the PPG procedure, having already participated in a research project utilizing PPG in the past 10 years.

Using a socio-demographic self-report questionnaire, participants from both groups were compared on their demographic characteristics including their age, preferred sexual orientation (0 = homosexual or 1 = heterosexual), relationship status (single = 0 or in a relationship = 1), personnel annual income (9 point Likert scale ranging from less than 10 000 dollars = 1 to more than 45 000 dollars = 9), and level of education (years of schooling completed). There were no significant or medium to large effect size differences between the clinical and comparison group on age, sexual orientation, and relationship status. The sexual

offenders were less wealthy and less educated than the comparison group, as characterized by medium to large effect size differences in both cases (see Table 1.).

Table 1
Participants Sociodemographic Characteristics.

Characteristics	Comparison		Sexual Offenders		<i>F</i>	<i>d</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
Age	48.0	11.7	47.5	14.9	.01	.04
Sexual Orientation	.9	.4	.9	.4	.01	.04
Annual Income	4.2	3.3	2.1	2.2	4.00	.76
Relationship Status	.2	.4	.2	.4	.01	.04
Level of education	14.1	3.3	11.9	3.4	3.18	.67

Note. Participants were mostly heterosexual (score of 1); the comparison group had a mean annual income of 20 000\$ to 25 000\$ while the child molester group had a mean annual income of 10 000\$ to 15 000\$; most participants were single at the time of the study (score of 0).

Measures.

Penile plethysmograph. PPG data were obtained for three different stimuli modalities (i.e., audio, visual, and audiovisual) using circumferential plethysmography with a mercury-in-rubber strain gauge. Penile responses were recorded for 120 seconds for each stimulus (90 seconds presentation and 30 seconds without any stimulation). As the rubber gauge was stretched in response to participants penile tumescence, the mercury contained within the gauge was also stretched resulting in variations in its electrical conductance. The differences in voltage were then converted to millimeter changes by the *prefest dataPac* and computer software *PrefTest Pro Suite*. Circumference changes were obtained by computing the difference between maximal penile response and base (or resting) penile state for each stimulus. Circumference scores were then converted into ipsative *Z* scores per each participant

and for each modality separately (i.e., the Z scores were generated within each modality and not across modalities). These Z scores were used to palliate the participants' individual physiological characteristics, which can make comparing mm differences between subjects difficult (e.g., penis size and overall penile responses; Earls, Quinsey, & Castonguay, 1987; Harris et al., 1992). Sexual interests' indices were calculated for each modality by subtracting the Z score on the adult stimulus from the highest Z score on a child stimulus (Harris et al., 1992; Seto, 2007). Raw mm scores were also used to compare mm circumference changes across modalities (i.e., within-subject differences).

Audio modality. Eight French standardized audio vignettes were used to assess sexual interests towards children and adults. These vignettes show high specificity ($Sp = .95$) and overall acceptable sensitivity ($Se = .67$; Barsetti, Earls, Lalumière, & Bélanger, 1998). They were translated and validated from the Quinsey and Chaplin (1988) set. The original English set also showed clear discrimination between sex offenders with child victims and non-offenders (for more information on the original [English] set see Fernandez [2002]). The eight-story set that was used for this study was comprised of: (a) neutral one, (b) neutral two, (c) consenting adult female, (d) consenting adult male, (e) sexual assault of a girl without violence (GNV), (f) sexual assault of a boy without violence (BNV), (g) sexual assault of a girl with violence (GV) and (h) sexual assault of a boy with violence (BV). Vignettes were approximately 120 seconds long: 90 second story followed by 30 seconds of silence. The neutral stimuli described: a) an adult male ordering food at a restaurant and b) an adult male visiting a zoo. The consenting adult vignettes contained a description of a consenting sexual relation between adults (i.e., male-female and male-male) in a relationship. The two other categories were differentiated by the level of violence or coercion used in order to attain the

sexual interaction with the child. The sexual assault without violence category contained a passive child victim, while the sexual assault with violence category contained severe physical violence and coercion towards the victim. Child victims described in the story were between 8 and 12 years old. The vignettes were all narrated in the second person by an adult male in order to allow the participant to imagine himself as being the protagonist of the story.

Visual modality. For each of the eight audio vignettes, a visual six-image storyboard was created depicting the sexual and non-sexual interactions taking place in the stories. Computer-generated characters were created by a custom in-house program designed to generate virtual characters of both sexes, modelled to follow the Tanner and Whitehouse (1976) stages of puberty development. This program was also used in previous studies (e.g., Trottier et al., 2015). All synthetic characters were created to correspond to the child being described in the stories (hair color, eyes, clothing, skin color, emotions). The protagonist in the story was generated to be a white male with an average to strong body build. The characters were then added to a virtual environment also corresponding to the audio descriptions (e.g., park, automobile, zoo). Once they were animated in the proper position to enact the sexual and non-sexual activity being described, a high resolution still image was taken from the perspective of the protagonist (i.e., generating a first-person view). Six of these images were made for each story / stimulus (i.e., in accordance with the storyboard), for a total of 48 images for the entire set. Together, these images created eight coherent stories similar to what one would perceive in a picture book. Images were shown in sequential order, and the length of presentation was identical to the one used for the audio modality. For a censored example of a consenting adult female stimulus, see Figure 1.

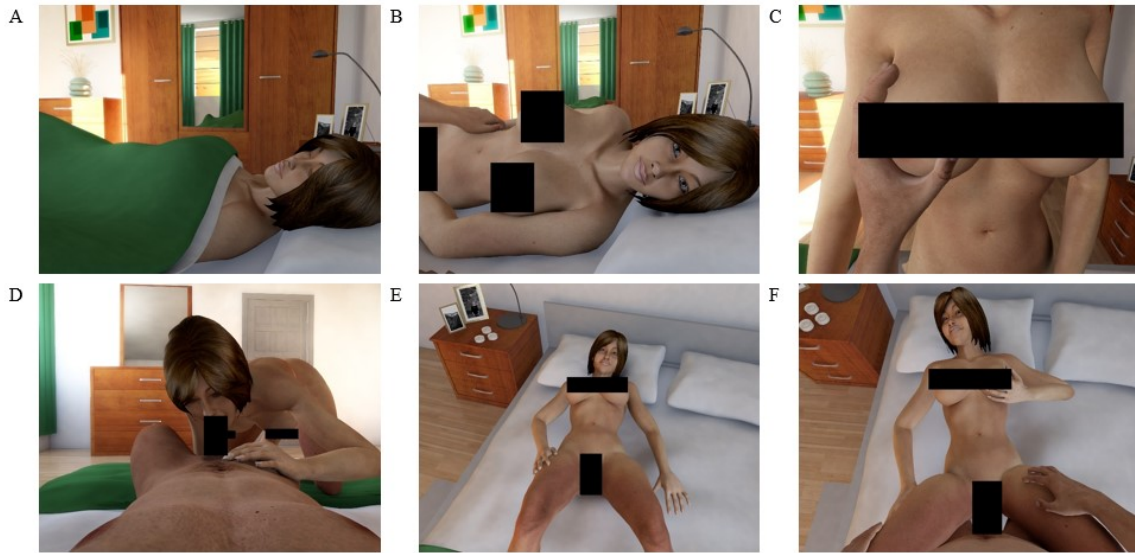


Figure 1
Censored Adult female stimulus.

Audiovisual modality. The audiovisual modality consisted of the superposition of the audio and visual stimuli. Hence, images were shown while the audio stimulus was playing, resulting in 90 seconds of stimulation per vignette followed with 30 without any visual or audio presentation.

Research Design and Procedure.

The entire procedure lasted approximately two and a half hours. At arrival, the laboratory equipment was presented to the participants in order for them to acclimatize to the experimental setting. After signed consent was obtained, participants were invited to complete the sociodemographic questionnaire and were led to the cubicle in which the PPG test was to be conducted. Participants were instructed on how to place the strain gauge on the shaft of their penis. For the entirety of the experimentation, participants were alone in a cubicle while the experimenter was in an adjacent section of the room. The participant and the researcher

were able to interact vocally during the entire procedure. The audio components of the evaluation were presented via headphones and the visual components were presented via a rear-mounted projector. A 2 X 3 block design with two independent groups (i.e., clinical and comparison) was utilized for the stimuli presentation (i.e., audio, visual, and audiovisual stimuli).

Counterbalancing was used for both the order of the block presentation and stimulus presentation within the blocks. Both counterbalancing measures were done so as to control for the order of presentation, since each participant saw and heard all three blocks (i.e., audio, visual, and audiovisual). Once all three presentation blocks were completed, participants were asked to evaluate the age of each character presented in the stimuli. Each participant was compensated 75 Canadian dollars for their participation. This project was approved by both the Institut Philippe-Pinel and the Université du Québec en Outaouais ethics board.

Statistical Analyses and Data Processing.

Although significance levels are provided for all results, we reported and interpreted data for all analyses showing medium to large effect sizes (i.e., $d \geq .50$ or $\eta^2 p \geq .06$). Data for several participants (see Table 2) had to be removed while conducting ROC curve analysis because they did not reach a minimum penile circumference change of 2.5 millimeters for any of the stimuli presented. These low responders did not differ from the rest of the participants on group affiliation (offenders or non-offenders), age, sexual orientation, annual income, relationship status, or level of education in terms of significance or in terms of effect size.

Table 2

Percentage of Valid PPG Profiles by Millimeter Change and Modality.

Modality	Valid at 2.5 mm (%)	Valid at 3 mm (%)
Audio	20/30 (67)	18/30 (60)
Visual	24/30 (80)	24/30 (80)
Audiovisual	24/30 (80)	24/30 (80)

Results

Perceived age of stimuli.

The computer-generated characters were modelled according to the Tanner and Whitehouse (1976) stages of puberty development and were created to be physically proportional to either children in Tanner Stage II (8 to 12 years old) or adults in Tanner Stage V (around 25 years old). We evaluated if participants could accurately identify the age of the characters presented. This was done by observing participants' accuracy percentage (see Table 3). Overall, participants correctly identified the Tanner stage II characters' ages ($M = 8.97$, $SD = 1.6$) as well as the Tanner Stage V characters' ages ($M = 27.58$, $SD = 3.80$). These results lend support to the notion that these characters were designed to detect pedophilic interests.

Table 3

Mean Age Given to Virtual Characters by Participants (With Standard Deviations) and Percentage Rates of Accurately Identifying Targeted Age.

Stimuli	M	SD	Accuracy of age (%)
Adult F	27.3	3.9	100
Adult M	27.8	3.7	100
GNV	8.8	1.5	83
BNV	8.6	1.6	77
GV	9.1	1.6	90
BV	9.3	1.6	87

Note. The means and standard deviations are presented in years; Accuracy represents the percentage of participants who correctly identified the adults as being over 18 years old and the children as being aged between 8 and 12 years old.

Increase in MM Responses.

One of the main objectives of the study was to examine whether the combination of the two most used stimulus modalities would generate larger responses than both modalities taken separately. This was tested using a 3 X 2 mixed repeated measures ANOVA with the group variable (sexual offenders and non-offenders) as the between-subjects factor and the modality variable (audio, visual, and audiovisual) as the within-subjects factors.

Raw scores were collapsed within each modality to obtain the highest mm change generated by each of the 30 participants for all three modalities. As expected, no substantial or significant interaction effects were found between group and modality, $F(2, 56) = .63, p = .538, \eta_p^2 = .02$. Analysis did show a large main effect for modality, $F(2,56) = 5.17, p = .009, \eta_p^2 = .16$. Pair sampled *t*-tests using Bonferroni correction ($\alpha = .025$) further demonstrated that there was no difference between the audio modality ($M = 9.49, SD = 10.01$) and the visual modality ($M = 9.06, SD = 7.22$), $t(29) = 0.33, p = .738, d = 0.13$). However, there was a larger effect size difference between the audiovisual modality ($M = 12.65, SD = 10.31$) and both the audio, $t(29) = 2.62, p = .014, d = 0.97$, and visual modality, $t(29) = 3.16, p = .004, d = 1.17$.

We then examined if these increases in penile responses resulted in a greater number of valid profiles using a 2.5 and 3 mm cut off point (see Table 2). Paired sample *t*-tests were used to compare the standard audio modality to both the visual and audiovisual modalities. Participants had more valid profiles (i.e., 2.5 mm and with large effect sizes) when presented with the visual ($M = .80, SD = .41$) or audiovisual ($M = .80, SD = .41$) stimuli than they did when presented with audio stimuli ($M = .67, SD = .48$), $t(29) = 2.11, p = .043, d = .80$; $t(29) = 2.11, p = .043, d = .80$. Similar results were found using the 3 mm cut-off point with both the visual ($M = .80, SD = .41$) and the audiovisual ($M = .80, SD = .41$) modalities, generating

more valid profiles than the audio modality ($M = .60, SD = .50; t(29) = 2.69, p = .012, d = 1.00; t(29) = 2.69, p = .012, d = 1.00$).

Discriminant Validity of Stimuli.

A mixed method repeated measures ANOVA was used to assess the effect of group, modality, and stimulus type as well as the interaction between stimulus type and group. The group variable (sex offenders and non-offenders) was used as the between-subjects factor while modality (audio, visual, and audiovisual) and stimuli type (Neutral, Adult, GNV, BNV, GV, BV) were used as the within-subjects factors.

As expected, there was a large interaction between stimulus type and group, $F(5, 140) = 5.73, p < .001, \eta_p^2 = 0.17$. This interaction was decomposed using a simple t -tests with Bonferroni correction ($\alpha = .001$). Results are presented in Table 4.

Table 4
Interaction Between Stimuli Type and Group.

Stimuli	Comparison		Sexual Offenders		t -test	d
	M	SD	M	SD		
Neutral	-.29	.53	-.49	.63	-.96	.36
Adult	1.48	.55	.57	.66	-4.07*	1.54
GNV	-.26	.30	.05	.65	1.71	.65
BNV	-.02	.52	.48	.68	2.26	.85
GV	-.27	.38	-.11	.54	.96	.36
BV	-.37	.38	.04	.65	2.08	.78

Note. Degrees of freedom remain constant (28); * = $p < .05$

As expected, the comparison group responded more to adult stimuli than the offender group and the offender group responded more to the child stimuli than the comparison group. The triple interaction (i.e., stimuli, modality, and group) only showed a small effect size difference that did not reach significance $F(10, 280) = .92, p = .488, \eta_p^2 = 0.03$. This suggests

that participants presented similar patterns of arousal to stimuli in all three modalities, which is to be expected. For instance, one would not expect a participant from the non-offender group to show strong arousal to the BNV stimulus in the audio modality and then show weak responses to the same BNV stimulus in the visual modality. For a graphical presentation of results by modality, group and stimulus type, see Figure 2.

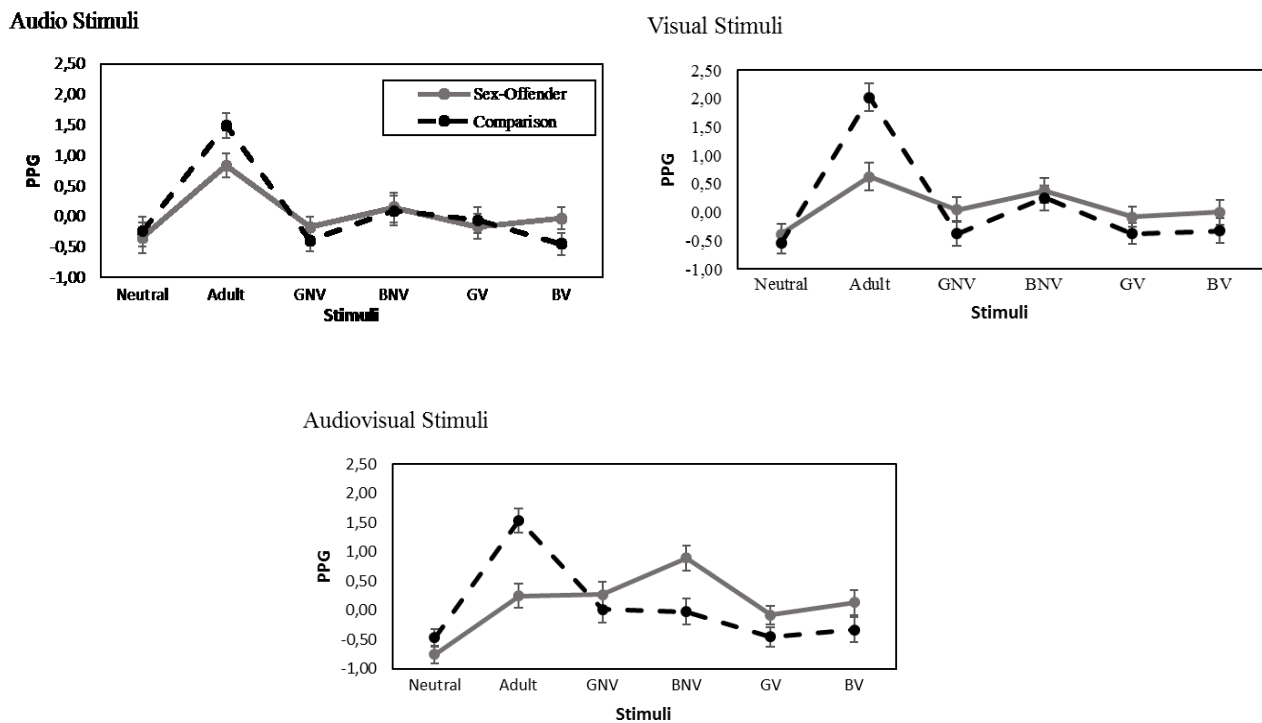


Figure 2
PPG Standardized Results by Stimuli, Modality of Presentation, and Group (with Standard Errors).

There was also a large main effect for stimulus type, $F(5, 140) = 21.51, p < .001, \eta_p^2 = .43$. This effect was decomposed using paired sample t -Tests with Bonferroni correction ($\alpha = .01$). Results showed that participants reacted more strongly to the consenting adult stimuli ($M = 1.02, SD = .76$) than to the Neutral ($M = -.65, SD = .48$), the GNV ($M = -.08, SD = .52$), the GV ($M = -.34, SD = .35$) and the BV stimuli ($M = -.34, SD = .54$). This main effect should be

interpreted with caution since a significant and large interaction effect was found previously showing that participants from both groups reacted differently according to stimuli type.

Classificatory Accuracy of Modalities.

A more rigorous way of evaluating criterion validity of a specific test consists of using Receiver Operating Characteristics (ROC) curves to calculate the specificity (Sp) and sensitivity (Se) at different cut-off points of a classification variable (i.e., ipsative sexual interest indices). This allows us to calculate the area under the curve (AUC) for each modality. In the context of this study, specificity refers to the capacity of the test to correctly identify a non-offender as belonging to the comparison group. Conversely, sensitivity refers to the test's ability to correctly identify a sexual offender with a child victim as belonging to the clinical group (Marshall & Fernandez, 2003). In essence, we are measuring the ability of the test to accurately identify group membership and not its ability to detect particular sexual interests. This is done because we cannot be certain that all participants in the offender group present pedophilic interests and that all participants in the comparison group do not present pedophilic interests. For the first analysis, we used ipsative sexual interest indices for all participants in order to obtain the real classification accuracy of all three modalities and not an artificially increased classification accuracy obtained by choosing a subgroup of participants. The second test was done using only valid profiles (i.e., at least 2.5 mm difference).

When all participants are included, the classification accuracy of both the visual (AUC = .76, $p = .017$) and the audiovisual (AUC = .85, $p = .001$) modalities were significant and respectively showed acceptable and excellent discrimination (Hosmer & Lemeshow, 2004). However, the classification accuracy of the audio modality did not reach significance (AUC = .68, $p = .093$). When only participants with results of 2.5 mm or more were included in the

analysis, the accuracy of the audio modality increased and reached significance ($AUC = .81, p = .026$) and the accuracy of both the visual ($AUC = .84, p = .013$) and audiovisual ($AUC = .83, p = .017$) stimuli remained significant. Using this 2.5 mm validity cut-off point, all three modalities are considered as having excellent discrimination. All AUC results as well as their confidence intervals are presented in Table 5. The AUC results for the audio modality obtained using the 2.5 mm validity cut-off point (i.e., $Sp = .91; Se = .63$) are similar to the results obtained from the original validation study ($Sp = .95; Se = .67$; Barsetti et al., 1998), suggesting that the audio stimuli performed as they were intended. The visual and audiovisual modalities were further statistically compared to the audio modality using the Hanley and James (1982, 1983) method. The differences did not reach statistical significance for either the visual $\chi^2(1, N = 30) = -1.07, p = .137$ or the audiovisual $\chi^2(1, N = 30) = -1.64, p = .102$ modalities, suggesting that both new modalities are at least as good as the audio modality in terms of classification accuracy.

In order to use all three modalities clinically, appropriate deviant cut-off points have to be determined. Two methods are generally recommended to achieve these clinical cut-off points. The first consists of using the optimized cut-off point (i.e., ipsative sexual interest indices compromising between Sp and Se) and the second consists of using the score which maximizes specificity (Youden's Index; Perkins & Schisterman, 2006). The cut-off scores for clinical use with corresponding Se and Sp are presented in Table 5.

Table 5

AUC Obtained for all Modalities, With and Without a 2.5 mm Validity Cut-Off Point and Corresponding Deviance Score.

Modality	<i>n</i>	AUC	95% CI	Youden's index			Maximized Sp		
				Se	Sp	Cut-off Score	Se	Sp	Cut-off Score
Audio	30	.68	.49 - .87	.67	.60	-.55	.33	.93	.76
Visual	30	.76*	.58 - .87	.67	.87	.43	.20	.93	2.13
Audiovisual	30	.85**	.70 - .99	.73	.93	.81	.73	.93	.81
Audio	20	.81*	.60 - 1.00	.63	.91	.78	.63	.91	.78
Visual	24	.84*	.66 - 1.00	1.00	.55	-1.94	.63	.91	-.51
Audiovisual	24	.83*	.63 - 1.00	.88	.82	-.53	.63	.91	.88

Note. *n* = 30 signifies that all participants were included in the analysis; *n* = 20 or *n* = 24 signifies that the 2.5 mm cut-off point was used; CI = Confidence Interval; Youden's index corresponds to the best compromises between Se and Sp with the according sexual interest indices cut-off scores; Maximized Sp corresponds to Sp between .90 and .95 with according Se and sexual interest indices cut-off scores; * = *p* < .05. ** = *p* < .01.

Discussion

The current study examined if the combination of computer-generated characters and audio stimuli would result in new multimodal stimuli capable of accurately discriminating between sexual offenders with child victims and non-offenders. It also tested if this new modality could generate greater penile circumference changes and greater numbers of valid PPG profiles than both other modalities. The new audiovisual modality resulted in a 33% (3.16 mm) increase in tumescence, while both the visual and audiovisual modalities showed a 13% (*n* = 4) increase in the number of valid (2.5 mm) profiles when compared to the standardized audio-stimuli. These lower responses obtained by the audio modality could be attributable to the male voice that was narrating all audio stimuli, therefore possibly generating an inhibitory effect in heterosexual males. It is also possible that, while all three modalities were story driven, only audio modality lacked visual sensory stimulation, which could account for the lower proportion of valid profiles. For instance, in the visual modality participants were still able to gather the context and the chronological order of events while being stimulated

visually, whereas in the audio modality participants were being described the context and had to rely on imagination for further stimulation.

Overall, the current study supports the new combined audiovisual modality for the assessment of sex offenders with child victims. Furthermore, it demonstrates that the new modalities (visual and audiovisual) were able to accurately classify low responders. These results support the lowering of the validity criteria of 2.5 mm when using these visual or audiovisual stimuli. Although we do not recommend completely eliminating the minimum change criteria, using the scores obtained on the neutral stimuli as the minimum circumference change to declare the profile valid could be a viable option as it has been previously used in past research studies (e.g., Seto & Lalumière, 2001).

Although this is not the first study to use a combination of audio and visual stimuli for the assessment of sex offenders with child victims, it is the first to have created computer-generated stimuli that accurately represent the events taking place in a validated set of audio stimuli. Furthermore, it is the first to combine synthetic characters with audio vignettes and establish their added value in terms of an increase in raw scores. We did not show a direct increase in the classification accuracy with the new audiovisual modality, but we did manage to show an indirect increase by lowering the number of invalid profiles. As mentioned previously, there are clinics that have successfully used a combination of audio and visual stimuli for many years. However, these visual stimuli usually constitute cues (e.g., the participant is asked to imagine that the child seen is the one in the story; Blanchard et al., 2007; Blanchard et al., 2012; Freund et al., 1991). Our method allows us to depict what is actually happening in the story in a first-person perspective, further insuring that the participant can actually see himself committing the sexual acts. This method, in essence,

should increase the ecological validity of the PPG procedure by tackling both the nature of the stimuli and the nature of the response (i.e., stronger penile responses; Schmuckler, 2001).

Although we cannot know how our audiovisual stimuli compare to those used in other laboratories without testing them, it is safe to say that laboratories that present stimuli concurrently can save a great deal of time on their evaluations, thus reducing their costs.

It is worth mentioning that although some participants did find the characters disturbing, they also found the audio stimuli just as disturbing. The greatest discomfort reported by participants was not attributable to a modality in particular, but to a category of stimuli (i.e., the sexual assault with violence stimuli). Since no distress was reported by any of the participants at the end of the study, it is unlikely that the audiovisual modality was seen as being more upsetting than the audio modality. Furthermore, 26 of the 30 participants reported finding the audiovisual modality more sexually appealing than both the audio and visual modalities shown separately.

The new stimuli in this study are subject to local laws on child pornography since they are stimuli modalities depicting or describing a sexual interaction between an adult and a child. The biggest advantage in this case is that all computer-generated stimuli are fictional, ensuring that some of the controversy surrounding the use of visual stimuli in PPG is eluded. Moreover, it would be possible to use the set of audiovisual stimuli in this study and modify them to respect the local laws of the more restrictive countries. For instance, it would be possible to show a naked character exhibiting the emotional responses described in the audio stories without the physical and sexual interaction between the child and the offender. This method would also allow researchers and clinicians to adjust certain physical characteristics of the children being presented in the stories (e.g., skin color, hair color) in order to more

accurately match the client's preferences or ethnic background without having to create an entirely new stimulus set.

Alternatively, the virtual environment described in the story could be presented without having any of the computer-generated characters. This additive method could easily be exported to other stimuli sets (e.g., offenders against adults). Finally, it would be possible to create fully animated stimuli that could generate even greater penile responses. These new or modified sets would have to be validated separately, but the method in itself is now easily attainable and could insure that clinicians use stimuli that are more standardized across laboratories.

Limitations and Conclusion.

There were some limitations to the present study. The first and most prominent is the small sample size. Although we obtained medium to strong effect sizes for most of the pertinent associations ($d \geq .50$), the small sample did hinder the analyses in terms of the power needed to obtain significant differences for the AUC's. The low availability of sexual offenders is also the reason a second comparison group comprised of adult rapists was not added to the study (i.e., participation had to be voluntary).

Secondly, the ATSA guidelines suggest the use of at least two stimuli per category of presentation (ATSA, 2001). However, in order to keep the length of the evaluation reasonable this study only used one stimulus per category (e.g., one sexual assault of a boy with violence stimulus). Future studies could easily compensate for this limitation by using a randomized controlled trial design comparing only the audio modalities to the audiovisual one.

Thirdly, all participants were presented the three modalities sequentially, therefore all stories were heard twice and all synthetic characters were seen twice. Although stimuli were well counterbalanced, this repeated measure could have contributed to the lower overall penile responses. This method was used due to low participant availability, but should be addressed in future studies by having only one modality presented to each participant.

Lastly, all stimuli were story-oriented, including the ones in the visual modality. This orientation makes comparing the present results on the visual stimuli to the ones generally obtained in the literature more complicated since it is difficult to separate the contribution of the story aspect from that of the naked characters. Therefore, it could be interesting to compare nude characters (with no interactions or environmental context) to story based stimuli such as the ones used in this study. Similarly, it would be interesting to compare our audiovisual stimuli to the ones used by the other laboratories. To conclude, the current results support the use of audiovisual stimuli in PPG. These new stimuli engendered greater penile responses, and they reduced the number of invalid profiles, therefore adding to the already existing advantages of both types of stimuli.

Acknowledgements

We wish to acknowledge the financial support of the Social Sciences and Humanities Research Council of Canada and the Fonds de Recherche Société et Culture (Research Grant awarded to the first author).

References

- Abel, G., Barlow, D., Blanchard, E. & Mavissakalian, M. (1975). Measurement of sexual arousal in male homosexuals: Effects of instructions and stimulus modality. *Archives of Sexual Behavior*, 4, 623-629. doi:10.1007/BF01544270
- Abel, G., Blanchard, E. B. & Barlow, D. H. (1981). Measurement of sexual arousal in several paraphilias: The effects of stimulus modality, instructional set and stimulus content on the objective. *Behaviour Research and Therapy*, 19, 25-33.
doi:http://dx.doi.org/10.1016/0005-7967(81)90109-1
- Abel, G., Huffman, J., Warberg, B. & Holland, C. (1998). Visual Reaction Time and Plethysmography as Measures of Sexual Interest in Child Molesters. *Sexual Abuse: A Journal of Research and Treatment*, 10(2), 81-95. doi:10.1023/a:1022063214826
- Akerman, G. & Beech, A. R. (2011). A systematic review of measures of deviant sexual interest and arousal. *Psychiatry, Psychology and Law*, 19, 118-143.
doi:10.1080/13218719.2010.547161
- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.
- Association for the Treatment of Sexual Abusers. (2001). *Practice standards and guidelines for the members of the Association for the Treatment of Sexual Abusers*. Beaverton, OR: Author.

- Barbaree, H. E. & Marshall, W. L. (1988). Deviant sexual arousal, offense history, and demographic variables as predictors of reoffense among child molesters. *Behavioral Sciences & the Law*, 6, 267-280. doi:10.1002/bsl.2370060209
- Barker, J. G. & Howell, R. J. (1992). The plethysmograph: A review of recent literature. *Journal of the American Academy of Psychiatry and the Law Online*, 20(1), 13-25.
- Barsetti, I., Earls, C. M., Lalumière, M. L. & Bélanger, N. (1998). The differentiation of intrafamilial and extrafamilial heterosexual child molesters. *Journal of Interpersonal Violence*, 13, 275-286. doi:10.1177/088626098013002007
- Blanchard, R., Klassen, P., Dickey, R., Kuban, M. E. & Blak, T. (2001). Sensitivity and specificity of the phallometric test for pedophilia in nonadmitting sex offenders. *Psychological Assessment*, 13(1), 118-126. doi:10.1037/1040-3590.13.1.118
- Blanchard, R., Kolla, N., Cantor, J., Klassen, P., Dickey, R., Kuban, M. & Blak, T. (2007). IQ, handedness, and pedophilia in adult male patients stratified by referral source. *Sexual Abuse: A Journal of Research and Treatment*, 19, 285-309. doi:10.1007/s11194-007-9049-0
- Blanchard, R., Kuban, M. E., Blak, T., Cantor, J. M., Klassen, P. E. & Dickey, R. (2009). Absolute versus relative ascertainment of pedophilia in men. *Sexual Abuse: A Journal of Research and Treatment*, 21, 431-441. doi:10.1177/1079063209347906
- Blanchard, R., Kuban, M., Blak, T., Klassen, P., Dickey, R. & Cantor, J. (2012). Sexual attraction to others: A comparison of two models of alloerotic responding in men. *Archives of Sexual Behavior*, 41, 13-29. doi:10.1007/s10508-010-9675-3

- Blanchard, R., Lykins, A., Wherrett, D., Kuban, M., Cantor, J., Blak, T., ... Klassen, P. (2009). Pedophilia, hebephilia, and the DSM-V. *Archives Sexual Behavior*, 38, 335-350. doi:10.1007/s10508-008-9399-9
- Chaplin, T. C., Rice, M. E. & Harris, G. T. (1995). Salient victim suffering and the sexual responses of child molesters. *Journal of Consulting and Clinical Psychology*, 63, 249-255. doi:10.1037/0022-006X.63.2.249
- Chartier, S., Renaud, P., Bouchard, S., Proulx, J., Rouleau, J. L., Fedoroff, P. & Bradford, J. P. (2006). Sexual preference classification from gaze behavior data using a multilayer perceptron. *Annual Review of Cyber Therapy and Telemedicine*, 4, 149-157.
- Clift, R. J. W., Rajlic, G. & Gretton, H. M. (2009). Discriminative and predictive validity of the penile plethysmograph in adolescent sex offenders. *Sexual Abuse: A Journal of Research and Treatment*, 21, 335-362. doi:10.1177/1079063209338491
- Dennis, E., Rouleau, J.-L., Renaud, P., Nolet, K. & Saumur, C. (2014). A pilot development of virtual stimuli depicting affective dispositions for penile plethysmography assessment of sex offenders. *The Canadian Journal of Human Sexuality*, 23, 200-208. doi:10.3138/cjhs.2529
- Dombert, B., Mokros, A., Brückner, E., Schlegl, V., Antfolk, J., Bäckström, A., ... Santtila, P. (2013). The virtual people set: Developing computer-generated stimuli for the assessment of pedophilic sexual interest. *Sexual Abuse: A Journal of Research and Treatment*, 25, 557-582. doi:10.1177/1079063212469062
- Earls, C. M. (1983). Some issues in the assessment of sexual deviance. *International Journal of Law and Psychiatry*, 6, 431-441. doi:10.1016/0160-2527(83)90030-4

- Earls, C. M., Quinsey, V. L. & Castonguay, L. G. (1987). A comparison of three methods of scoring penile circumference changes. *Archives of Sexual Behavior*, *16*, 493-500.
doi:10.1007/BF01541713
- Farrall, W. R. & Card, R. D. (1988). Advancements in physiological evaluation of assessment and treatment of the sexual aggressor. *Annals of the New York Academy of Sciences*, *528*, 261-273. doi: 10.1111/j.1749-6632.1988.tb50869.x
- Fernandez, Y. (2002). *Phallometric testing with sexual offenders against female victims: Limits to its value*. In Forum on Corrections Research (Vol. 14, No. 1, pp. 7-12). Correctional Services of Canada.
- Freund, K. (1963). A laboratory method for diagnosing predominance of homo- or heterosexual interest in the male. *Behaviour Research and Therapy*, *1*, 85-93.
doi:http://dx.doi.org/10.1016/0005-7967(63)90012-3
- Freund, K. & Blanchard, R. (1989). Phallometric diagnosis of pedophilia. *Journal of Consulting and Clinical Psychology*, *57*, 100-105. doi:10.1037/0022-006X.57.1.100
- Freund, K., Watson, R., Dickey, R. & Rienzo, D. (1991). Erotic gender differentiation in pedophilia. *Archives of Sexual Behavior*, *20*, 555-566. doi:10.1007/BF01550954
- Gaither, G. A. & Plaud, J. J. (1997). The effects of secondary stimulus characteristics on men's sexual arousal. *The Journal of Sex Research*, *34*, 231-236.
doi:10.1080/00224499709551890
- Goyette, M. (2012). *Évaluation des intérêts sexuels auprès d'agresseurs sexuels d'enfants par la vidéooculographie et la présentation de stimuli générés par ordinateur* (doctoral dissertation). Retrieved from <http://hdl.handle.net/1866/8714>

- Hall, G. C. N., Hirschman, R. & Oliver, L. L. (1995). Sexual arousal and arousability to pedophilic stimuli in a community sample of normal men. *Behavior Therapy*, 26, 681-694. doi:[http://dx.doi.org/10.1016/S0005-7894\(05\)80039-5](http://dx.doi.org/10.1016/S0005-7894(05)80039-5)
- Hanley, J. A. & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143, 29-36. doi:10.1148/radiology.143.1.7063747
- Hanley, J. A. & McNeil, B. J. (1983). A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology*, 148, 839-843. doi:10.1148/radiology.148.3.6878708
- Harris, G., Lalumière, M. L., Seto, M., Rice, M. & Chaplin, T. (2012). Explaining the erectile responses of rapists to rape stories: The contributions of sexual activity, non-consent, and violence with injury. *Archives of Sexual Behavior*, 41, 221-29. doi:10.1007/s10508-012-9940-8
- Harris, G. T., Rice, M. E., Quinsey, V. L., Chaplin, T. C. & Earls, C. (1992). Maximizing the discriminant validity of phallometric assessment data. *Psychological Assessment*, 4, 502-511. doi:10.1037/1040-3590.4.4.502
- Hosmer Jr, D. W. & Lemeshow, S. (2004). *Applied logistic regression*: John Wiley & Sons.
- Jamison, P. L. & Gebhard, P. H. (1988). Penis size increase between flaccid and erect states: an analysis of the Kinsey data. *Journal of Sex Research*, 24(1), 177-183.
- Julien, E. & Over, R. (1988). Male sexual arousal across five modes of erotic stimulation. *Archives of Sexual Behavior*, 17(2), 131-143. doi:10.1007/BF01542663

- Kalmus, E. & Beech, A. R. (2005). Forensic assessment of sexual interest: A review. *Aggression and Violent Behavior, 10*(2), 193-217.
doi:<http://dx.doi.org/10.1016/j.avb.2003.12.002>
- Kuban, M., Barbaree, H. E. & Blanchard, R. (1999). A Comparison of Volume and Circumference Phallometry: Response Magnitude and Method Agreement. *Archives of Sexual Behavior, 28*(4), 345-359. doi:10.1023/a:1018700813140
- Lalumière, M. L. & Harris, G. T. (1998). Common questions regarding the use of phallometric testing with sexual offenders. *Sexual Abuse: A Journal of Research and Treatment, 10*, 227-237. doi:10.1177/107906329801000306
- Launay, G. (1999). The phallometric assessment of sex offenders: an update. *Criminal Behaviour and Mental Health, 9*, 254-274. doi:10.1002/cbm.317
- Laws, D. R. (1984). The assessment of dangerous sexual behavior in males. *Medicine and Law, 3*, 127-140. Retrieved from
<http://www.scopus.com/inward/record.url?eid=2-s2.0-0021281832&partnerID=40&md5=4006dc85318a3ffdd193186872c14ee>
- Laws, D. R., Hanson, R. K., Osborn, C. A. & Greenbaum, P. E. (2000). Classification of child molesters by plethysmographic assessment of sexual arousal and a self-report measure of sexual preference. *Journal of Interpersonal Violence, 15*, 1297-1312.
doi:10.1177/088626000015012004
- Malcolm, P. B., Andrews, D. A. & Quinsey, V. L. (1993). Discriminant and predictive validity of phallometrically measured sexual age and gender preference. *Journal of Interpersonal Violence, 8*, 486-501. doi:10.1177/088626093008004004

- Marshall, W. L., Barbaree, H. E. & Christophe, D. (1986). Sexual offenders against female children: Sexual preferences for age of victims and type of behaviour. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, 18, 424-439. doi:10.1037/h0079966
- Marshall, W. L. & Fernandez, Y. M. (2000). Phallometric testing with sexual offenders: Limits to its value. *Clinical Psychology Review*, 20, 807-822. doi:http://dx.doi.org/10.1016/S0272-7358(99)00013-6
- Marshall, W. L. & Fernandez, Y. M. (2003). Sexual preferences: Are they useful in the assessment and treatment of sexual offenders? *Aggression and Violent Behavior*, 8, 131-143. doi: http://dx.doi.org/10.1016/S1359-1789(01)00056-8
- McAnulty, R. & Adams, H. (1990). Patterns of sexual arousal of accused child molesters involved in custody disputes. *Archives of Sexual Behavior*, 19, 541-556. doi:10.1007/bf01542464
- Murphy, W., Haynes, M., Stalgaitis, S. & Flanagan, B. (1986). Differential sexual responding among four groups of sexual offenders against children. *Journal of Psychopathology and Behavioral Assessment*, 8, 339-353. doi:10.1007/BF00960731
- Perkins, N. J. & Schisterman, E. F. (2006). The inconsistency of “optimal” cutpoints obtained using two criteria based on the receiver operating characteristic curve. *American Journal of Epidemiology*, 163, 670-675. doi:10.1093/aje/kwj063
- Proulx, J. (1989). Sexual preference assessment of sexual aggressors. *International Journal of Law and Psychiatry*, 12, 275-280. doi:http://dx.doi.org/10.1016/0160-2527(89)90019-8

- Proulx, J., Pellerin, B., Paradis, Y., McKibben, A., Aubut, J. & Ouimet, M. (1997). Static and dynamic predictors of recidivism in sexual aggressors. *Sexual Abuse: A Journal of Research and Treatment*, 9, 7-27. doi: 10.1007/BF02674889
- Quinsey, V. L. & Chaplin, T. C. (1988). Penile responses of child molesters and normals to descriptions of encounters with children involving sex and violence. *Journal of Interpersonal Violence*, 3, 259-274. doi:10.1177/0886260888003003001
- Quinsey, V. L., Rice, M. E. & Harris, G. T. (1995). Actuarial prediction of sexual recidivism. *Journal of Interpersonal Violence*, 10(1), 85-105.
- Renaud, P., Chartier, S., Rouleau, J.-L., Proulx, J., Décarie, J., Trottier, D., ... Bouchard, S. (2009). Gaze behavior nonlinear dynamics assessed in virtual immersion as a diagnostic index of sexual deviancy: Preliminary results. *Journal of Virtual Reality and Broadcasting*, 6. doi: 10.20385/1860-2037/6.2009.3
- Renaud, P., Rouleau, J.-L., Proulx, J., Trottier, D., Goyette, M., Bradford, J. P., ... Bouchard, S. (2010). Virtual characters designed for forensic assessment and rehabilitation of sex offenders: standardized and made-to-measure. *Journal of Virtual Reality and Broadcasting*, 7, 1860-2037. doi:10.20385/1860-2037/7.2010.5
- Rice, M. E., Quinsey, V. L. & Harris, G. T. (1991). Sexual recidivism among child molesters released from a maximum security psychiatric institution. *Journal of Consulting and Clinical Psychology*, 59, 381-386. doi:10.1037/0022-006X.59.3.381
- Rosen, R. C. & Keefe, F. J. (1978). The measurement of human penile tumescence. *Psychophysiology*, 15, 366-376. doi:10.1111/j.1469-8986.1978.tb01393.x
- Schmuckler, M. A. (2001). What is ecological validity? A dimensional analysis. *Infancy*, 2, 419-436. doi:10.1207/S15327078IN0204_02

- Seto, M. C. (2007). Psychophysiological assessment of paraphilic sexual interests. In E. Janssen (Ed.), *The psychophysiology of sex* (pp. 475-491). Indianapolis: University of Indiana Press.
- Seto, M. C. & Lalumière, M. L. (2001). A brief screening scale to identify pedophilic interests among child molesters. *Sexual Abuse: A Journal of Research and Treatment*, *13*, 15-25. doi:10.1023/A:1009510328588
- Stinson, J. D., Becker, J. V. & Sales, B. D. (2008). Self-regulation and the etiology of sexual deviance: Evaluating causal theory. *Violence and Victims*, *23*, 35-51. doi:10.1891/0886-6708.23.1.35
- Tanner, J. M. & Whitehouse, R. H. (1976). Clinical longitudinal standards for height, weight, height velocity, weight velocity, and stages of puberty. *Archives of Disease in Childhood*, *51*, 170-179. doi:10.1136/adc.51.3.170
- Thornton, D. & Laws, D. R. (Eds.). (2009). *Cognitive Approaches to the Assessment of Sexual Interest in Sexual Offenders*. United-Kingdom: John Wiley & Sons Ltd.
- Trottier, D. (2013). *Les apports de la réalité virtuelle et de la vidéo-oculographie à l'évaluation des intérêts sexuels* (Doctoral dissertation). Retrieved from <http://hdl.handle.net/1866/9191>
- Trottier, D., Renaud, P., Rouleau, J.-L., Goyette, M., Saumur, C., Boukhalfi, T. & Bouchard, S. (2015). Erratum to: Using immersive virtual reality and anatomically correct computer-generated characters in the forensic assessment of deviant sexual preferences. *Virtual Reality*, *19*, 303-303. doi:10.1007/s10055-015-0277-1

Trottier, D., Rouleau, J.-L., Renaud, P. & Goyette, M. (2014). Using eye tracking to identify faking attempts during penile plethysmography assessment. *The Journal of Sex Research, 51*, 946-955. doi:10.1080/00224499.2013.832133

Zappalà, A., Antfolk, J., Bäckström, A., Dombert, B., Mokros, A. & Santtila, P. (2013). Differentiating sexual preference in men: Using dual task rapid serial visual presentation task. *Scandinavian Journal of Psychology, 54*, 320-327.
doi:10.1111/sjop.12050

Zuckerman, M. (1971). Physiological measures of sexual arousal in the human. *Psychological Bulletin, 75*, 297-329. doi :10.1037/h0030923

Chapitre III - Deuxième article

Pupillometry using computer-generated characters: how does it compare to PPG when identifying sex-offenders with child victims?

Article en préparation pour soumission

Shawn Marschall-Lévesque, candidat au Ph. D.

Rouleau Joanne-Lucine, Ph. D.

Alexa Larouche Wilson, candidat au Ph. D.

Patrice Renaud, Ph. D.

Abstract

Penile plethysmography has been criticized for the invasive nature of the procedure and for its use of images of real children. These critiques highlight the need for an alternative physiological measure of sexual interests. The current study attempts to address these criticisms by measuring the aesthetic phase of sexual interests while using pupillometry with computer-generated characters. Two groups of participants, sexual offenders with child victims and non-offenders, were presented with audiovisual and visual stimuli, utilizing computer generated characters, from which phallometric and pupillometric data were collected. Results demonstrated that both measures were able to distinguish between the two groups $F(4, 112) = 7.50, p < .001$, with the non-offender group reacting more to the adult stimulus than the offender group and the offender group reacting more to the stimuli containing children. Furthermore, ROC analysis showed that phallometric (AUC = .83, $p = .001$), pupillometric (AUC = .74, $p = .026$) and the combined measure (AUC= .89 $p < .001$) accurately classified participants from both groups. Differences in modality of presentation are also discussed, as well as the future use of pupillometry and considerations for future studies.

Key words: Pupillometry, Phallometry, Sex-offenders with child victims, Sexual Interests

Introduction

Penile plethysmography (PPG) as a measure of sexual interests has been actively studied and used since the 1950's and 60's (Freund, 1963, 1965) and is an integral part of a standardized assessment of sexual interests (Gaither & Plaud, 1997; Kalmus & Beech, 2005; Lalumière & Harris, 1998). Among its many strengths are its ability to discriminate between sexual interests towards children, violence, and its objectivity as a recidivism risk assessment tool (Laws, 1984; Malcolm, Andrews & Quinsey, 1993; Marshall & Fernandez, 2000). Nevertheless, PPG has fostered criticism in the scientific community. One of these concerns the invasive nature of the procedure which requires participants to either place a rubber gauge on their penis (circumferential PPG) or an air tight cylinder over their penis (volumetric PPG) while having their sexual responses monitored (Kalmus & Beech, 2005; Kuban, Barbaree & Blanchard, 1999). Another concern has been the PPG's tendency to generate a high number of invalid profiles, which can reach up to 34% (Akerman & Beech, 2012; Barker & Howell, 1992; Farrall & Card, 1988; Kalmus & Beech, 2005; Marshall, Barbaree & Christophe, 1986; Murphy, Haynes, Stalgaitis & Flanagan, 1986; Quinsey & Chaplin, 1988; Stinson, Becker & Sales, 2008). Finally, the use of real images (*i.e.*, actual images of children) is generally discouraged in PPG procedures because it can be seen as revictimization of the children being presented (Abel, Huffman, Warberg & Holland, 1998; Kalmus & Beech, 2005; Launay, 1994). For these reasons, PPG is not currently used in many countries, greatly limiting the measure's overall usefulness. This highlights the necessity for a complementary measure or, in some cases, a replacement measure for the assessment of sexual interests in children.

Measuring Sexual Interests.

Sexual interest, regardless of its nature, can be conceptualized as a three phase system (Singer, 1984). The first phase corresponds to the aesthetic response (*i.e.*, increased attention and interest towards an aesthetically pleasing stimuli). The second phase is described as the approach response, which consists of a physical response from the subject with the purpose of getting closer to the stimuli (*e.g.*, moving closer to the sexually appealing stimuli). Finally, the third phase is considered the genital response, and it includes, but is not limited to, genital tumescence, increased heart rate and breathing, and muscle tension. It is believed that by adequately measuring at least one of these phases we can ascertain participants' sexual interests. The PPG can be conceptualized as a measure of the third and final phase of this model (Gaither & Plaud, 1997; Kalmus & Beech, 2005; Lalumière & Harris, 1998). The use of aesthetic response or approach response phase has been proposed to measure sexual interests, and several attempts have been made (Barker & Howell, 1992; Kalmus & Beech, 2005; Rosenzweig, 1942; Wright & Adams, 1994).

Viewing Time.

First introduced to the field of sexual research in the in the 1940's, viewing time relies on the premise that individuals will look at images presenting sexually appealing material (*e.g.*, adult females for heterosexual males) for longer periods of time than at comparable images that do not (adult males for heterosexual males; Rosenzweig, 1942). These finding were adapted to create an indirect measure of the aesthetic response (Kalmus & Beech, 2005) by asking participants to complete a task requiring them to rate a potentially sexually appealing image or locate a specific object on the said image, while being distracted by the image itself. In essence, this measure is akin to a Stroop test which requires the participants to

forgo an automatic response in order to complete a selected task (MacLeod, 1991). In doing so, it is hypothesized that the distracter images presenting participants' preferred sexual interests (i.e., adult male, adult female, young boy or young girl) will generate longer delay in the completion of the task, resulting in longer viewing times (Israel & Strassberg, 2009). By comparing these viewing times to different sets of images presenting adults and children, it becomes possible to ascertain preferential sexual interest (Abel, Jordan, Hand, Holland, & Phipps, 2001; Flak, Beech, & Fisher, 2007; Gress, 2005; Wright & Adams, 1994). Although studied for a long time, this measure was popularized by Abel and colleagues with the *Abel Assessment for Sexual Interest* (AASI; 2001), a standardized test sold for the assessments of sexual interests towards children (Kalmus & Beech, 2005). The AASI is by no means the only viewing time measure that has been developed, the other example being the Affinity (Glasgow, Osborne, et Croxen, 2003), but it is the most well known. The AASI has produced discriminative results between sexual offenders and non-offenders comparable to those obtained with PPG (Abel, Huffman, Warberg, & Holland, 1998; Harris, Rice, Quinsey, & Chaplin, 1996; Letourneau, 2002). This measure has also shown good internal consistency (Krueger, Bradford, et Glancy, 1998). Furthermore, the AASI is described by its creator as far less invasive than PPG (Abel, Huffman, Warberg, & Holland, 1998).

However, viewing time has received its fair share of criticism over the years. For instance, the theoretical understanding of how the measure functions on a cognitive level is still quite unclear (Imhoff et al., 2010). Furthermore, the measure is transparent, making it susceptible to faking attempts (Fischer & Smith, 1999; Kalmus & Beech, 2005). Finally, viewing time cannot easily be combined with other measures which usually rely on the premise that stimuli are presented for the same amounts of time (PPG, EEF, FMRI). This

makes triangulation of data difficult. Therefore, another more direct measure of the aesthetic response of sexual interest would be useful. Pupillometry (PUP) has been described as one such measure (Kalmus & Beech, 2005).

Pupillometry.

Pupillometry was first introduced to the field of sexual research by Hess and colleagues in the 1960's and 70's, triggering a large amount of interest in it as a stand-alone measure of sexual interests (Hess, Beaver & Shrouf, 1975; Hess, Seltzer & Shlien, 1965). Similar to PPG procedure, PUP consists of the physiological measure of pupil size during the presentation of various stimuli, with larger pupils generally suggesting a greater interest. Researchers demonstrated that PUP could be used to distinguish homosexual from heterosexual males (Hess & al., 1965), heterosexual males from females (Hess & Polt, 1960; Simms, 1967), sexually oriented from non-sexually oriented stimuli (Aboyoun & Dabbs, 1998), and finally neutral stimuli from various female stimuli (Nunnally, Knott, Duchnowski, & Parker, 1967; Peavler & McLaughlin, 1967).

Initially, Hess hypothesized that constriction of the pupil was associated with negative emotional valence towards a stimulus while dilatation of the pupil was associated with positive emotional valence towards said stimulus (Janisse, 1973; Zuckerman, 1971). This hypothesis has evolved over the years, and researchers have shown that pupil changes are associated with the intensity of an emotional response to a stimulus rather than to the valence of the emotion being generated by this stimulus (*i.e.*, positive or negative; Mathôt & Van der Stigchel, 2015). All other things being equal, this suggests that greater pupil dilatation occurs when presented with stimuli generating strong emotional reactions (positive or negative), and that weaker dilatation occurs when presented with neutral stimuli (Bradley, Miccoli, Escrig &

Lang, 2008). For reviews on pupil dilatation, see the following references (Janisse, 1973; Kalmus & Beech, 2005; Laeng, Sirois & Gredebäck, 2012; Mathôt & Van der Stigchel, 2015). The question arises on whether this pattern would emerge when individuals are presented with stimuli that generate strong aesthetic responses (first phase of sexual interest). If so, PUP could be used to ascertain sexual preferences towards children or adults.

Many studies have been conducted using PUP as a discriminative measure of sexual interests, though not with sexual offenders as participants (Attard-Johnson, Bindemann & Ciardha, 2016; Atwood & Howell, 1971; Hamel, 1974; O'Neill & Hinton, 1977; Rieger, Cash & al., 2015; Rieger & Savin-Williams, 2012; Schnelle, Kennedy, Rutledge & Golden, 1974). A more recent study presented images (adult and children) to a normative sample of adults (men and women) and found that results supported the current hypothesis, with heterosexual men obtaining greater pupil dilatation towards adult women than towards adult men or children. Similarly, heterosexual women showed greater pupil dilatation to both adult men and women than to children (Attard-Johnson & al., 2016). In addition, Rieger's research found greater pupil dilatation to images corresponding to participants' sexual orientation (*i.e.*, heterosexual or homosexual; Rieger & al., 2015; Rieger & Savin-Williams, 2012; Rieger, Savin-Williams, Chivers & Bailey, 2015). These studies demonstrated that pupil dilatation varies significantly in accordance to one's sexual preferences, supporting PUP as an effective measure of the aesthetic responses of sexual interests. These results also support the idea of PUP being a discriminative measure for sex-offenders with child victims. If true, it would allow for a less invasive measure of sexual interests. However, the use of images of real children would warrant the same ethical considerations afflicting PPG. A visual substitute for real images would therefore be needed.

Computer-Generated Characters.

Recent studies have shown that computer generated characters representing children and adults could be a viable alternative to the use of real images to assess sexual interests towards children. These characters are modeled to represent either stage I or II children or stage V adults on the Tanner and Whitehouse (1976) pubertal development scale. Characters modeled on these three stages are the ones which have been shown to discriminate between sexual offenders with child victims and non-offenders with results comparable to those obtained with audio stimuli (Chapter II; Dombert & al., 2013; Goyette, Rouleau & Renaud., 2012; Renaud & al., 2010; Trottier, Rouleau & Renaud., 2013; Trottier & al., 2015). Moreover, these stimuli have been used successfully with eye tracking technology which could also be used to assess pupil dilatation (Goyette & al., 2012; Trottier, Rouleau, Renaud & Goyette, 2014).

The current study will examine if PUP using computer generated characters could be used as a discriminative measure of sexual interests between sexual-offenders with child victims and non-offenders. In addition, we wanted to evaluate if PUP results were comparable to those obtained with PPG. We hypothesized that participants from the sexual offenders group would react more strongly to the child stimuli on both measures (PPG and PUP) than participants in the comparison group, while participants in the comparison group would react more strongly to the adult stimuli than participants in the child molester group. Furthermore, both measure should allow for significant and comparable discriminant validity. Finally, if both measures allowed for significant classification accuracy, then it is expected that both measures (PUP and PPG) could be combined to create an overall better index of sexual interests.

Method

Participants.

Participants for this study are distributed into two groups of 15 adult males each: sexual-offenders with child victims and non-offenders (comparison group). This sample is the same that was used in Chapter II and data for both studies were collected at the same time (*i.e.*, all participants accepted to participate in both studies). To be included in the offender group participants had to have been convicted or had to have admitted to having sexual contact with a child 13-years-old or younger while being at least 16 years old and at least five years older than their victim (American Psychiatric Association, 2000, 2013). Participants were recruited following their court ordered clinical assessment at the Institut Philippe-Pinel de Montréal (IPPM) or from hand distributed fliers during treatment maintenance sessions for sexual offenders. Participants in the non-offender group were recruited using an advertisement in a local newspaper. None of the participants in the comparison group reported having had sexual contact with a child that was 13 years old or younger while being at least 16 years old or reported ever having had pedophilic interests or fantasies. No history of sexual assault, either as the perpetrator or victim, was reported in the comparison group. Participants from both groups were comparable in terms of age, sexual orientation and relationship status. Overall, participants from the clinical group were less educated and less wealthy than their counterparts in the comparison group. See Chapter II for more information on the participants sociodemographic characteristics. Although more information pertaining to the sexual offences perpetrated by participants in the clinical group was assessed, the low sample size of the group did not allow for these characteristics to be used efficiently in statistical analyses. For instance, out of the 15 participants, 11 had a male victim, nine had more than one victim,

12 had a victim aged 11 or less, three participants had victims that were 12 years old, and five participants had an unrelated victim.

Measure and Data Processing.

Penile plethysmography. Phallometric data were obtained using the *prefest dataPac* and computer software *PrefTest Pro Suite* with a mercury in rubber strain gauge (Rosen & Keefe, 1978). Tumescence was assessed by subtracting baseline measures (*i.e.*, flaccid penis) from greatest point of penile tumescence for all seven categories of stimuli: Neutral (N), Adult female (AF), Adult male (AM), sexual assault of a girl without violence (GNV), sexual assault of a boy without violence (BNV), sexual assault of a girl with violence (GV) and sexual assault of a boy with violence (BV). Scores that did not reach the minimum 2.5 mm validity scores (6/30) were replaced by their group mean. PPG results were assessed for 120 seconds (90 second stimuli presentation and 30 seconds of no stimuli presentation). Millimeter results were then transformed into ipsative *Z* scores across stimuli categories within each participant's scores (Earls, Quinsey & Castonguay, 1987). In order to control for sexual orientation, heterosexual and homosexual stimuli were collapsed for the adult stimulus by using the score corresponding to the participants self-reported sexual orientation, resulting in a six-stimuli set (N, Adult, GNV, BNV, GV and BV). Although the neutral stimulus was used for the *Z* scores, they were not included in further analyses in order to obtain five categories of scores, which paralleled the number of categories in PUP (discussed below).

Pupillometry. Pupil dilatation data were recorded via the *SMI-ETG* (eye tracking glasses) and the *iViewETG 2.0* program. The data were then analyzed using the *BeGaze* (version 3.4) program. These glasses use two infrared cameras directed towards the eye and record pupil size at a frequency of 60 Hz. Raw data for both the right and left eye were

averaged together to obtain mean pupil dilatation. Missing pupil data (4%) were replaced by the participants' group mean (offender or non-offender group). Data were missing because of technical problems (the glasses were non-functional for the assessment of one participant and, for two participants, the program did not record the pupil data on two vignettes).

Several methodological considerations were taken into account when processing PUP data. The aesthetic responses we were trying to measure should correspond to a short autonomic window in which the pupil reacts proportionally to the intensity of the emotion felt by the participant when presented with the stimulus. In order to try and capture this window, only the first five seconds of pupillary responses were used for each image presented in each category of stimulus. That is, for each category of stimulus six images were presented, from these images only the first five seconds of pupillary responses were used. Furthermore, it has been shown that automatic light responses to the changing of an image occur after 200 ms from image onset and reach their peak at between 500 and 1 000 ms (Beatty & Lucero-Wagoner, 2000; Bernhardt, Dabbs & Riad, 1996). In order to compensate for this potential confound, the first 1 200 ms of pupil data for every image were not used to compute mean pupil dilatation. A similar control was suggested by Goldwater (1972) and used by Naber, Alvarez & Nakayama (2013). This resulted in 3 800 ms of PUP data for each image being analyzed at a rate of 60 Hz (*i.e.*, approximately one measure per second). Data for each of these six images were then averaged to obtain a mean pupil dilatation score for each category stimulus (*i.e.*, N, AdultF, AdultM, GNV, BNV, GV, BV).

Raw scores on the neutral (N) stimuli were subtracted from the raw scores on the sexual stimuli (AdultF, AdultM, GNV, BNV, GV, BV) for all participants to obtain a dilatation or contraction score from the neutral stimulus (*i.e.*, how much the pupil dilated or

constricted by comparison to a neutral stimuli). These data were then transformed into Z scores in order to more easily compare scores across participants and measures. This method of processing data was used for two reasons. First, contrary to phallometric assessment, where it is easier to determine participants' base rate (*i.e.*, a flaccid penis), it is much harder to determine a base rate for a participant's pupil since it can be affected by a multitude of factors. Second, by using a sexually neutral stimulus (*i.e.*, a man visiting a zoo), which had comparable luminosity to the other sexual stimuli, we created a point of comparison from which we hoped to evaluate sexual interests (*i.e.*, aesthetic response). Similar methods have been used in other pupillometric studies that either used the overall mean to all stimuli (Attard-Johnson & al., 2016; Dabbs, 1997) or the means of neutral stimuli (Chapman, Chapman & Brelje, 1969; Libby, Lacey & Lacey, 1973; Rieger & al., 2015; Rieger & Savin-Williams, 2012; Rieger, Savin-Williams & al., 2015). Finally, in order to control for sexual orientation, the two adult stimuli were collapsed according to participants' self-reported sexual orientation (*i.e.*, results on the adult female stimulus were used for the heterosexual participants while the results to the adult male stimulus were used for the homosexual participants). These procedures resulted in five categories of scores for the PUP measure (*i.e.*, Adult, GNV, BNV, GV and BV).

SSPI. All participants in the sexual-offender group were also assessed on the Screening Scale for Pedophilic Interests (SSPI; Seto & Lalumière, 2001). This brief, four items, scale is completed by the assessor, based on the characteristics of the offender's victims. Items include: a) the presence of a male victim, b) the presence of more than one victim, c) the presence of a prepubescent victim (*i.e.*, 11 years old or younger) and d) the

presence of an extra familial victim. Scores on the SSPI have been shown to be significantly correlated with PPG results and recidivism rates (Seto, Harris, Rice, & Barbaree, 2004).

Stimuli. The stimuli used for both the PPG and PUP assessment had three modalities, (*i.e.*, audio, visual and audiovisual). Psychometric information on these stimuli is available in Chapter II. For the current study, only data on the visual and audiovisual modality were presented. The results for the audio stimuli were excluded from this study because data for the PUP measure showed more than 50% missing data. This missing data were caused by participants closing their eyes while listening to the audio stimuli. Although participants were instructed to keep their eyes open and look at the blank screen, they did not comply with the request. This will be further discussed in the limitations section. Overall, the visual and audiovisual modality used for the PPG procedure showed good specificity ($Sp = .87$, $Sp = .93$) and sensitivity ($Se = .63$, $Se = .73$). In addition, they accurately classified participants in their appropriate group (offenders and non-offenders) as shown by the measures' classification accuracy for the visual ($AUC = .76$) and audiovisual ($AUC = .85$) stimuli.

Overall, both modalities were composed of eight stimuli, including two neutral (N), one adult female (AdultF), one adult male stimulus (AdultM), one sexual assault without violence against a female child (GNV), one sexual assault without violence against a male child (BNV), one sexual assault of a female child with violence (GV) and one sexual assault of a male child with violence (BV). The GNV stimulus contained a passive ten-year-old girl, whereas the BNV stimulus contained a passive eight-year-old boy. The GV stimulus contained a ten-year-old girl who was a victim of severe physical violence and coercion. Lastly, the BV stimulus contained an eight-year-old boy who was also a victim of severe physical violence and coercion. The vignettes were all narrated in the second person singular by an adult male

in order to allow the participant to imagine himself as being the protagonist of the story. The images of children created for the visual elements of the presentations were modeled to respect Tanner and Whitehouse (1976) levels of pubertal development. Children were modeled to correspond to Tanner stage II characteristics. Hence, both female and male children were generated to show little to no pubic hair, the female children show very low to no thelarche of the breast and finally, the male characters' penis size follow Tanner stage II descriptions (testicular volume between 1.6 and 6 ml; thinning skin on scrotum that reddens and enlarges). For more information on these stimuli, see chapter II.

Each of the eight visual stimuli are composed of six images presented in the first person showing the sexual and non-sexual interactions taking place in the stories. For the audiovisual stimuli, while these images were presented, an audio track was played narrating in the second person singular the events taking place in the story. Each stimulus (visual or audiovisual) was 90 seconds long followed by a 30 second span where no stimuli (audio or visual) were presented. The luminosity of the images was adjusted to be within 30% of overall mean lumens (131 lumens) for all images (*i.e.*, 20 lumens above or below the 131 lumens mean). Table 1 shows mean luminance values and standard deviations for each of the seven stimuli, including maximum and minimum lumens values for images within each stimulus. As mentioned previously, the original stimulus set in Chapter II included a second neutral stimulus which was not included in the analysis in the current study because the luminance could not be adjusted while still maintaining the images' coherence with the scenario as well as the images' sharpness and contrast. The scenario in question consisted of a man ordering a meal in a dark and low-lit restaurant. Although luminance was controlled in our study, a recent study utilizing PUP to ascertain sexual interests found no relevant differences in results

between their luminance controlled set of images and their original uncontrolled set (Attard-Johnson & al., 2016).

Table 1

Mean luminance values (with standard deviations) for the seven stimuli with maximum and minimum lumens values for images within each stimulus.

Stimuli	Mean	SD	Max	Min
N2	137	68	158	103
AdultF	141	49	152	121
AdultM	127	66	147	100
GNV	147	65	182	116
BNV	130	78	160	117
GV	124	64	144	96
BV	112	49	126	102

Note. N2: neutral stimuli number 2; Adult F: consenting adult female; Adult M: consenting adult male; GNV: sexual assault of a girl without violence; BNV: sexual assault of a boy without violence; GV: sexual assault of a girl with violence; BV: sexual assault of a boy with violence.

Research Design and Procedure.

The procedure lasted approximately two and a half hours. At arrival of participants at the lab, the equipment and the procedure were explained to them and their signed consents were obtained. Participants completed a sociodemographic questionnaire and were then led to the cubicle where the procedure was to take place. The experimenter helped the participants install the *SMI-ETG* (eye tracking glasses) and left them to install the penile stress gauge by themselves. Stimuli were presented in a 3X2 block design, with two independent groups (clinical and comparison) and three repeated measures (audio, visual and audiovisual stimuli). As mentioned previously, in the current study only the visual and audiovisual results were used. Two counter-balancing procedures were used (within and between blocs) to account for the presentation order for all participants as they saw and heard all three blocks (*i. e.*, audio, visual and audiovisual). As compensation for their participation, participants were awarded 75 Canadian dollars. This project was approved by both, the Institut Philippe-Pinel and the Université du Québec en Outaouais ethics boards.

Results

Discriminant validity.

Because of the small sample size, we will only report and interpret data for analyses showing medium to large effect sizes (*i. e.*, $d \geq .50$ or $\eta_p^2 \geq .06$). To ascertain whether PUP data presents similar patterns to PPG data in terms of distinguishing between the two groups, a mixed method repeated measure MANOVA (5X4X2) was used. Five categories of stimuli were employed (Adult, GNV, BNV, GV and BV) for the four measures (VisualPUP, AudioVisualPUP, VisualPPG, AudioVisualPPG), with both independent groups (offenders and non-offenders). Two large and significant effect size interactions were found: one within subject interaction (Stimuli X Measure) and one between subject interaction (Stimuli X Group).

The within subject interaction (Measure X Stimuli), $F(12, 112) = 4.36, p < .001, \eta_p^2 = .14$, was decomposed using a paired sample t-test (Table 2). Overall, the results suggest that participants obtained higher scores on both the PPG measures for the Adult, GNV and BNV stimuli than they did on both the PUP measures. Furthermore, the results indicate that there were no differences between the four measures (*i.e.*, VisualPUP, AudiovisualPUP, VisualPPG and AudiovisualPPG) for the GV stimuli. Participants did, however, react more strongly to both the PUP measures for the BV stimuli than they did to both the PPG measures. This is an interesting result considering that PPG scores on the sexual assault with violence category are the lowest scores obtained by the participants on both PPG measures, while the PUP scores for the sexual assault with violence of a boy stimulus represent some of the highest scores obtained for both these measures (for a graphical representation of results, see Figure 1).

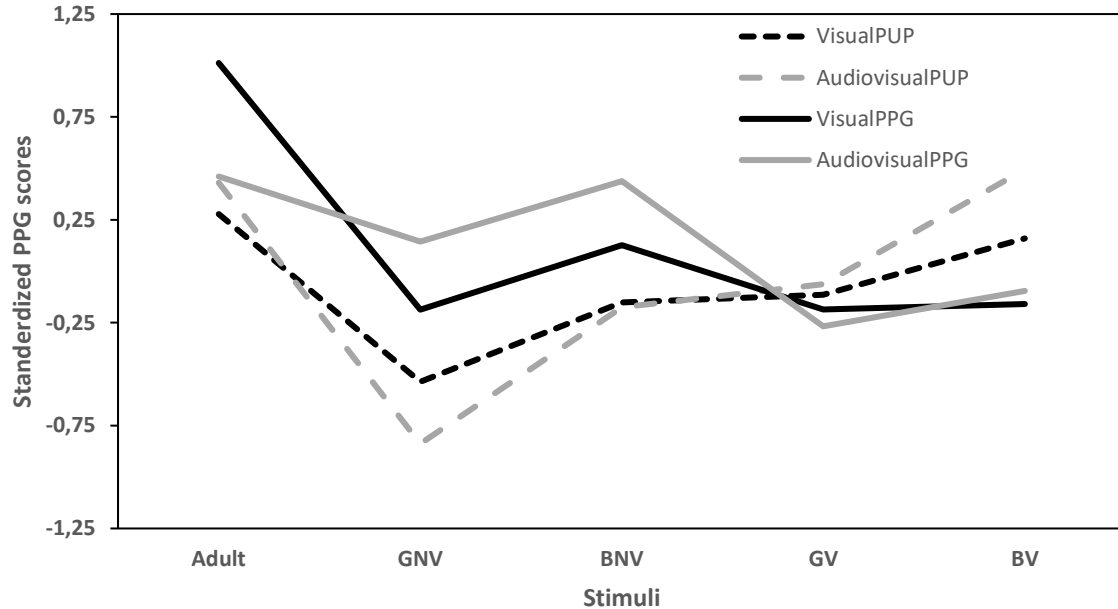


Figure 1
Participants overall standardized results by stimuli and group.

These results could be explained by the notion that PPG measures sexual interests while PUP measures general interests and strong emotional responses; most participants show no sexual interests in the use of violence against children but do seem to show general interest and strong emotional reactions towards the use of violence against children (see Table 2). However, results from this analysis do not allow us to establish if the general and emotional reactions seen with the PUP results overshadow possible discriminative differences in terms of group membership (*i.e.*, offenders and non-offenders).

Table 2

Decomposed interaction for stimuli X measure (with means and standard deviations).

Interactions	Adult		GNV		BNV		GV		BV	
	M (SD)	t (d)	M (SD)	t (d)	M (SD)	t (d)	M (SD)	t (d)	M (SD)	t (d)
VPUP	.28 (.71)	-1.06 (.39)	-.54 (.87)	1.55 (.57)	-.15 (.80)	.11 (.04)	-.11 (.77)	-.30 (.11)	.16 (.94)	-2.00 (.74)
X AVPUP	.43 (.83)		-.84 (.86)		-.17 (.79)		-.06 (.72)		.50 (.59)	
VPUP	.28 (.71)	-3.95 (1.47)	-.54 (.87)	-1.60 (.60)	-.15 (.80)	-1.43 (.53)	-.11 (.77)	.36 (.14)	.16 (.94)	1.44 (.54)
X VPPG	1.01(1.02)		-.19 (.86)		.13 (.89)		-.19 (.69)		-.16 (.86)	
VPUP	.28 (.71)	-.88 (.33)	-.54 (.87)	-2.72 (1.01)	-.15 (.80)	-2.74 (1.02)	-.11 (.77)	.81 (.30)	.16 (.94)	1.41 (.52)
X AVPPG	.46(1.20)		.14 (.86)		.44 (.95)		-.27 (.64)		-.10 (.85)	
AVPUP	.43 (.83)	-2.58 (.96)	-.84 (.86)	-2.96 (1.10)	-.17 (.79)	-1.49 (.55)	-.06 (.72)	.74 (.27)	.50 (.59)	3.09 (1.15)
X VPPG	1.01(1.02)		-.19 (.86)		.13 (.89)		-.19 (.69)		-.16 (.86)	
AVPUP	.43 (.83)	-.15 (.06)	-.84 (.86)	-4.32 (1.60)	-.17 (.79)	-2.73 (1.02)	-.06 (.72)	1.11 (.41)	.50 (.59)	3.39 (1.26)
X AVPPG	.46(1.20)		.14 (.86)		.44 (.95)		-.27 (.64)		-.10 (.85)	
VPPG	1.01(1.02)	2.53 (.94)	-.19 (.86)	-1.59 (.59)	.13 (.89)	-1.49 (.55)	-.19 (.69)	.55 (.21)	-.16 (.86)	-.31 (.11)
X AVPPG	.46(1.20)		.14 (.86)		.44 (.95)		-.27 (.64)		-.10 (.85)	

Note. VPUP: visual Pupilometry; AVPUP: audiovisual pupilometry; VPPG: visual plethysmography; AVPPG: audiovisual plethysmography; Adult: consenting adult; GNV: sexual assault of a girl without violence; BNV: sexual assault of a boy without violence; GV: sexual assault of a girl with violence; BV: sexual assault of a boy with violence; Paired sample t-tests were used for the within subject interaction; effect size differences of $d \geq .50$ are reported in bold.

In order to answer this question, we had to examine the between-subject interaction (Stimuli X Group) $F(4, 112) = 7.50, p < .001, \eta_p^2 = .21$, which was also decomposed using a simple t-test with Bonferroni correction ($\alpha = .01$; see Table 3). Results showed large and significant effect size differences between groups for both the Adult and BNV stimuli, with the non-offender group reacting more strongly to the Adult stimulus than the offender group and the offender group reacting more strongly to the BNV stimulus than the non-offender group. Similarly, there was a moderate effect size difference between both groups for the GV and BV stimulus, with the offender group reacting more strongly than the non-offender group in both cases.

Table 3

Decomposed interaction for stimuli X group (with means and standard deviations).

Variables		M	SD	t	d
Stimuli X Group					
Adult	Offenders	.16	.58	3.63**	1.37
	Non-Offenders	.93	.59		
GNV	Offenders	-.33	.46	-.33	.13
	Non-Offenders	-.38	.44		
BNV	Offenders	.34	.44	-3.58**	1.35
	Non-Offenders	-.22	.41		
GV	Offenders	-.04	.37	-1.76	.67
	Non-Offenders	-.28	.37		
BV	Offenders	.27	.56	-1.96	.74
	Non-Offenders	-.07	.36		

Note. Adult: consenting adult; GNV: sexual assault of a girl without violence; BNV: sexual assault of a boy without violence; GV: sexual assault of a girl with violence; BV: sexual assault of a boy with violence; Paired sample t-test were used for the within subject interaction while paired sample t-tests were used for the between subject interaction; degrees of freedom are constant at ($Df = 28$, simple t-test; $Df = 29$, paired sample t-test) * = $p < .01$; ** = $p < .001$.

Overall, these results are in concordance with the literature suggesting that PUP is a measure of intensity of interest (*e.g.*, Janisse, 1973; Mathôt & Van der Stigchel, 2015), while also suggesting that this interest varies in terms of group affiliation. In addition, our results support the literature on the possible usefulness of PUP as a classification measure (*i.e.*, criterion validity).

Classification Accuracy.

In order to more accurately evaluate criterion validity of the PUP and compare it to the PPG, we used Receiver Operating Characteristics (ROC) curves to calculate both their specificity (Sp) and sensitivity (Se). For both the PPG and PUP measures, Sp refers to a measure's capacity to correctly identify a non-offender as being from the comparison group, while Se refers to the measure's capacity to correctly identify a child molester as belonging to the clinical group (Marshall & Fernandez, 2000). This was done by calculating sexual interest indices for each of the four measures (VisualPUP, AudiovisualPUP, VisualPPG and

AudiovisualPP) and by combining both visual indices (PUP and PPG) and both Audiovisual indices (PUP and PPG) to create combined indices, one visual and one audiovisual. Indices were obtained by subtracting participants scores for the adult stimulus for each measure from their respective highest score on a child stimulus for the corresponding measure (e.g., highest VisualPPG score on the BNV stimulus minus highest VisualPPG score on the adult stimulus). The two combined indices were obtained by calculating the mean of both visual indices and both audiovisual indices. Results showed that all but one of the sexual interests' indices (VisualPUP) were able to accurately classify participants in their appropriate group (with significant AUC's varying between .71 and .89). Comparing AUC's using the Hanley and McNeil (1982, 1983) showed only one significant difference between the visual PUP indices (AUC = .63) and the combined audiovisual indices (AUC = .89; $z = -2.16$, $p = .031$). See Table 4 for AUC results, including the appropriate clinical cut-off score corresponding to Youden's index and the maximized specificity scores.

Table 4
AUC for all four measures and for the combined Indices with corresponding cut-off scores using either Youdens' index or maximized Sp.

<u>Indices</u>	<u>AUC</u>	<u>95% CI</u>	<u>Youden's index</u>			<u>Maximized Sp</u>		
			<u>Se</u>	<u>Sp</u>	<u>Cut-off Score</u>	<u>Se</u>	<u>Sp</u>	<u>Cut-off Score</u>
<u>PUP</u>								
Visual	.63	.41-.85	.60	.80	.24	.33	.93	1,47
Audiovisual	.74*	.56-.92	.93	.47	-.26	.40	.93	1,24
<u>PPG</u>								
Visual	.76*	.67-.99	.73	.80	-.18	.20	.93	2.13
Audiovisual	.83***	.57-.94	.87	.73	.51	.33	.93	2.31
<u>Combined</u>								
Visual	.71*	.51-.90	.67	.80	.26	.27	.93	1.10
Audiovisual	.89***	.77-1.00	.87	.80	.39	.67	.93	1.08

Note. CI = Confidence Interval; Youden's index corresponds to the best compromise between Se and Sp, with the corresponding sexual interest indices cut-off scores; Maximized Sp corresponds to Sp between .90 and .95 with corresponding Se and sexual interest indices cut-off scores; * = $p < .05$. ** = $p < .01$. *** = $p < .01$.

These 6 indices were then examined using Spearman correlation to see if they were associated with one another (Table 5). As expected, most indices were significantly correlated to one another and those that were not, still had correlations going in the expected direction.

Table 5
Spearman correlations between pupillometric, plethysmographic and combined indices.

	Visual PUP	Audiovisual PUP	Visual PPG	Audiovisual PPG	Visual Comb	Audiovisual Comb
VisualPUP	-					
AudivisualPUP	.48**	-				
VisualPPG	.47*	-0.02	-			
AudiovisualPPG	.34	0.22	.60**	-		
VisualComb	.71**	0.18	.90**	.57**	-	
AudiovisualComb	.49**	.68**	.45*	.84**	.52**	-

Note. Comb: combined; * = p<.05; ** = p<.01

Because previous results showed that participants reacted differently to the sexual assault with violence stimuli on both measures, it was decided to investigate if PUP results for violent scenarios could still be used to accurately classify participants. In order to evaluate this, three different sexual interest indices were calculated for both audiovisual measures (*i.e.*, non-violent indices, violence indices and overall indices). Indices were obtained by subtracting participants scores for the adult stimulus from their highest score on a non-violent stimuli (GNV or BNV), on a violent stimuli (GV or BV), or their highest score towards a child stimulus (any of the four child stimuli). Because the AUC results for the visual PUP indices did not reach levels of significance, further analyses were only conducted for the audiovisual stimuli.

The results showed that all six sexual interest indices were able to accurately classify sexual offenders with child victims, with significant AUC's varying between .73 and .84. Comparing AUC's using the Hanley and McNeil (1982, 1983) method did not show any

significant differences in classification accuracy neither between any of the six sexual interest indices nor between the two overall measures (PPG or PUP). This is somewhat surprising because we were not sure that stimuli consisting of sexual assault with violence would allow for accurate classification using the PUP measure, because these results differed from the ones obtained from PPG. This suggests that although the sexual assault with violence stimuli seemed to generate strong emotional reactions, as shown by pupil dilatation, the amplitude of this reaction still varied in accordance to group membership. See Table 6 for AUC results, including the appropriate clinical cut-off score corresponding to Youden's index and the maximized specificity scores. Overall, these consistent results show that not only were all three PUP sexual interest indices able to accurately identify and classify participants into their appropriate groups, but they did so with similar efficiency to the phallometric sexual interest indices.

Table 6
AUC for both PPG and PUP measures with corresponding cut-off Scores using either Youdens' index or maximized Sp.

Indicies	AUC	95% CI	Youden's index			Maximized Sp		
			Se	Sp	Cut-off Score	Se	Sp	Cut-off Score
PPG								
NV	.84***	.70 - .99	.80	.80	.51	.33	.93	2.08
Violence	.82**	.66 - .98	.73	.80	.14	.53	.93	.99
Overall	.83**	.67 - .99	1.00	.67	-.36	.33	.93	2.31
PUP								
NV	.73*	.55 - .91	.60	.87	-.34	.27	.93	.22
Violence	.78*	.59 - .95	.67	.87	-.27	.53	.93	.22
Overall	.74*	.56 - .92	.93	.47	-.26	.40	.93	1.24

Note. NV; no violence; CI = Confidence Interval; Youden's index corresponds to the best compromise between Se and Sp, with the corresponding sexual interest indices cut-off scores; Maximized Sp corresponds to Sp between .90 and .95 with corresponding Se and sexual interest indices cut-off scores; * = $p < .05$. ** = $p < .01$.

Criterion validity.

Criterion validity was assessed by correlating results on the SSPI with the standardized scores for the visual and audiovisual PPG and PUP measures. Moreover, both PPG and PUP results were correlated to one another to further assesses concurrent validity. Because of the high number of variables involved, results are presented in three distinct tables (Table 7, Table 8 and Table 9).

Table 7

Spearman correlations between SSPI variables and visual and audiovisual (standardized) PPG results.

	SSPI	Male Vic	More1 Vic	11or Less	Noaff Vic	VAd	VGNV	VBNV	VGv	VBV	MAd	MGNV	MBNV	MGV	MBV
SSPI	1.00														
Male Vic	.73**	1.00													
More1 Vic	.42	.04	1.00												
11or Less	.16	-.42	.23	1.00											
NoaffVic	.45^T	.21	-.22	.00	1.00										
VAd	-.05	-.21	.17	.26	-.25	1.00									
VGNV	-.13	-.31	-.40	.20	.44^T	-.01	1.00								
VBNV	-.25	-.24	.02	-.13	.00	-.18	-.17	1.00							
VGv	.26	.21	-.25	.00	.43	-.22	.17	-.22	1.00						
VBV	.17	.14	.40	-.12	-.11	-.50**	-.30^T	.17	-.03	1.00					
MAd	.18	-.10	.10	.18	.40	.52**	-.19	-.14	-.10	-.23	1.00				
MGNV	-.02	-.17	-.10	.48^T	-.22	-.03	.08	.02	.20	.05	-.34^T	1.00			
MBNV	.28	.10	.54*	.23	-.25	-.21	.05	.25	-.25	.48**	-.47**	.17	1.00		
MGV	-.03	.07	-.15	-.32	.24	-.37*	.21	.01	.20	-.16	-.09	-.41*	-.25	1.00	
MBV	-.27	-.07	-.23	-.22	-.07	-.26	.28	-.06	.12	.11	-.56**	.06	-.03	.29	1.00

Note. SSPI = total score on SSPI measure; MaleVic= presence of a male victim; More1Vic = presence of more than one victim; 11orLess = presence of a victim aged 11 years old or less; NoaffVic = presence of a non affiliated victim; Ad = adult stimulus; GNV: sexual assault of a girl without violence; BNV: sexual assault of a boy without violence; GV: sexual assault of a girl with violence; BV: sexual assault of a boy with violence; T = $p < .10$. * = $p < .05$. ** = $p < .01$; significant results and tendencies are presented in bold.

Results for the few correlations that did reach significance or tendencies show that individual SSPI items were positively associated with PPG results for both the visual and audiovisual measure. It is not surprising that few correlations reached significance when considering that only 15 participants were included in the analyses. Nevertheless, these results are coherent with the literature in showing that higher scores on the items for the SSPI should be associated with greater results on the PPG measure.

Table 8

Spearman correlations between SSPI variables and visual and audiovisual (standardized) PUP results.

	SSPI	Male Vic	More1 Vic	11or Less	Noaff Vic	VAd	VGNV	VBNV	VGV	VBV	MAd	MGNV	MBNV	MGV	MBV
SSPI	1.00														
Male Vic	.73**	1.00													
More1 Vic	.42	.04	1.00												
11or Less	.16	-.42	.23	1.00											
Noaff Vic	.45^T	.21	-.22	.00	1.00										
VAd	.15	-.10	.34	.02	.22	1.00									
VGNV	.12	-.03	.27	-.06	.25	-.06	1.00								
VBNV	.29	.21	-.12	.19	.17	-.06	-.28	1.00							
VGV	-.67**	-.45^T	-.64**	-.04	-.09	-.14	-.31^T	-.16	1.00						
VBV	-.22	.10	-.26	-.14	-.37	-.55**	-.19	.02	-.02	1.00					
MAd	.28	.21	.05	-.03	.34	.34^T	.20	-.22	-.13	-.22	1.00				
MGNV	-.11	-.07	-.13	-.19	.14	-.18	.30	.15	-.20	-.13	-.13	1.00			
MBNV	-.21	.07	-.02	-.44^T	-.14	-.05	-.18	.01	.25	.25	-.07	-.06	1.00		
MGV	-.03	-.38	-.21	.78**	.02	-.11	-.23	.02	.27	.10	-.33^T	-.38*	-.20	1.00	
MBV	-.07	.00	-.09	.10	-.09	-.38*	-.25	.02	.20	.42*	-.40*	-.22	.22	.34^T	1.00

Note. SSPI = total score on SSPI measure; MaleVic = presence of a male victim; More1Vic = presence of more than one victim; 11orLess = presence of a victim aged 11 years old or less; NoaffVic = presence of a non affiliated victim; Ad = adult stimulus; GNV: sexual assault of a girl without violence; BNV: sexual assault of a boy without violence; GV: sexual assault of a girl with violence; BV: sexual assault of a boy with violence; T = $p < .10$. * = $p < .05$. ** = $p < .01$; significant results and tendencies are presented in bold.

Although not as intuitive, results for the PUP measure are coherent with what is known about this sample of participants. For instance, the negative correlations between the visual sexual assault of a girl with violence stimulus (VGV) and the presence of a male victim is not surprising. Moreover, the two other correlations between the VGV stimulus and the overall SSPI score and the presence of more than one victim can be further explained by the fact that the SSPI score is highly correlated ($r = .73$, $p = .01$) with the presence of a male victim and the fact that those participants that had more than one victim usually also had a male victim. Finally, the correlations between the presence of a victim aged 11 years old or younger (11orLess) and the audiovisual sexual assault of a boy without violence stimulus (MBNV; $r = -.44$, $p = .100$) as well as the audiovisual sexual assault of a girl with violence stimulus (MGV; $r = .78$, $p = .001$) is explained by the fact that participants that had younger victims generally had female victims. Although these results should be interpreted with caution because of small sample size ($n = 15$) and low power, they are coherent with what was expected.

Table 9
Spearman correlations between standardized PPG and PUP results.

	PPG										PUP										
	VAd	VGNV	VBNV	VGv	VBV	MAd	MGNV	MBNV	MGV	MBV	VAd	VGNV	VBNV	VGv	VBV	MAd	MGN	MBNV	MGV	MBV	
<u>PPG</u>																					
VAd	1.00																				
VGNV	-.01	1.00																			
VBNV	-.18	-.17	1.00																		
VGv	-.22	.17	-.22	1.00																	
VBV	-.50**	-.30^T	.17	-.03	1.00																
MAd	.52**	-.19	-.14	-.10	-.23	1.00															
MGNV	-.03	.08	.02	.20	.05	-.34^T	1.00														
MBNV	-.21	.05	.25	-.25	.48**	-.47**	.17	1.00													
MGV	-.37*	.21	.01	.20	-.16	-.09	-.41*	-.25	1.00												
MBV	-.26	.28	-.06	.12	.11	-.56**	.06	-.03	.29	1.00											
<u>PUP</u>																					
VAd	.37*	-.27	.14	.03	-.14	.32 ^T	-.04	.02	.01	-.44*	1.00										
VGNV	.02	-.04	-.09	-.26	.18	.42*	-.25	-.08	.02	-.17	-.06	1.00									
VBNV	-.08	-.19	.27	.29	.30	-.10	.19	.09	-.16	.18	-.06	-.28	1.00								
VGv	-.28	.17	.25	-.13	-.21	-.32^T	-.07	.11	-.01	.02	-.14	-.31^T	-.16	1.00							
VBV	-.12	.27	-.13	-.01	-.14	-.21	-.21	-.05	.17	.32^T	-.55**	-.19	.02	-.02	1.00						
MAd	.11	-.18	-.03	-.23	.09	.45*	-.10	.09	-.22	-.68**	.34^T	.20	-.22	-.13	-.22	1.00					
MGNV	.02	.03	-.09	-.04	.33^T	.07	-.06	-.09	.14	.24	-.18	.30	.15	-.20	-.13	1.00					
MBNV	-.16	-.36*	.03	-.22	.15	-.12	-.08	-.01	-.22	-.05	-.05	-.18	.01	.25	-.07	-.06	1.00				
MGV	.01	.18	-.06	.19	-.31^T	.00	.10	-.01	-.03	-.01	-.11	-.23	.02	.27	.10	-.33^T	-.38*	1.00			
MBV	-.07	.07	-.02	.16	-.29	-.14	-.13	-.21	.00	.13	-.38*	-.25	.02	.20	.42*	-.40*	-.22	.22	1.00		

Note. Ad = adult stimulus; GNV: sexual assault of a girl without violence; BNV: sexual assault of a boy without violence; GV: sexual assault of a girl with violence; BV: sexual assault of a boy with violence; T = $p < .10$. * = $p < .05$. ** = $p < .01$, significant results and tendencies are presented in bold.

Lastly, results in Table 9 show acceptable criterion validity for both PUP and PPG measures with non-deviant adult stimuli being positively correlated with one another and negatively correlated with deviant child stimuli. Furthermore, all other significant or tangential correlations are in the expected directions with nonviolent stimuli being positively correlated with other nonviolent stimuli and negatively correlated with violent stimuli. Similarly, stimuli were also correlated in accordance to the sex of the victims presented. Overall, results for the three tables show acceptable criterion validity between SSPI scores and both PPG and PUP scores. Moreover, both PPG and PUP measure also show acceptable criterion validity between one another.

Discussion

This study attempted to show that PUP using computer generated characters, could be used as a discriminative measure of sexual-offenders with child victims and non-offenders as effectively as PPG. The results showed mirroring patterns of responses for the non-violent stimuli between the phallometric and the pupillometric assessments. However, participants' PUP results to the stimuli consisting of the sexual assault with violence were contrary to what was obtained on the PPG stimuli. On average participants showed low penile responses to the stimuli with sexual assault with violence and strong pupil responses to the stimuli with sexual assault with violence. These results seem coherent with available literature (Bradley & al., 2008; Janisse, 1973) that showed that strong emotional responses generate strong pupil responses and not necessarily strong sexual responses. Although this emotional pupil response seemed to have been present in our study (*i.e.*, pupil response to the sexual assault with violence stimuli), it did not invalidate or overshadow the discriminative, between group responses, as was suggested by Janisse (1973). Hence, all sexual interest indices created with the audiovisual PUP data could accurately distinguish between sexual offenders with child victims and the non-offenders. It is important to note that these results do raise questions about

pupillometry's capacity to assess preferential sexual interest towards sexual assault with violence or violence, as is normally possible with phallometric assessment. With further studies, pupillometry could be used to ascertain the presence of sexual interests towards children as opposed to adults, but not towards certain sexual acts (*i.e.*, humiliation or violence) because these will most likely generate strong negative emotional responses in all participants. For instance, 26 of the participants verbally mentioned having not enjoyed watching or hearing the stimuli incorporating violence. From these participants four mentioned having felt anger while listening to or viewing the stimuli that included violence. Further considerations must also be taken into account with regards to PUP's ability to accurately discriminate sexual interest because the visual PUP indices did not allow for accurate classification accuracy. It is possible that the addition of the audio vignettes in the audiovisual stimuli is what allowed the PUP indices for this modality to be accurate. Without the PUP results for the audio modality, however, this hypothesis is harder to support. Ways in which to address missing data with audio modality in future studies are discussed further in the limitations section.

Overall, the results showed that, as is the case with PPG indices, audiovisual PUP indices could accurately classify participants in the appropriate groups. Moreover, findings show that combining PPG and PUP sexual interests' indices generated an overall sexual interest index that showed the best classification accuracy out of all the indices examined. This suggests that both measures do assess the same concept (*i.e.*, sexual interests) in different ways and are in line with Singer's (1984) model, with PPG results assessing the genital response phase and PUP assessing the aesthetic response phase.

Pupillometry has its fair share of advantages compared to PPG. As mentioned by Rieger and Savin-Williams (2012), PUP is less invasive. Furthermore, it can be easily added

to any PPG procedure at a relatively low cost while not adding to the procedures' overall length. Using PUP with computer-generated characters could also allow physiological assessments of sexual interests in countries where PPG is not accessible by providing a non-invasive measure (*i.e.*, glasses) using stimuli which are ethically more viable than images of real children. If our results are corroborated by future studies, PUP would most likely be used as a complementary measure with PPG. Although at this point in time we do not recommend PUP as a stand-alone measure to replace PPG, it is possible that with further research PUP could be added to other physiological measures (*e.g.*, eye tracking, EEG) in an attempt to obtain a triangulated response of sexual interests.

Limitations, Future Studies and Conclusion.

There are a few limitations to this study with the first and foremost being its small sample size. Although large and medium effect size differences were found, the results would need to be replicated with a larger sample. This low sample size is mostly attributable to the low availability of sexual offenders. It is also the reason why a second group comprised of sexual offenders against adult women was not included. Second, while there were six images per category of stimulus, it is suggested that PPG assessments include at least two stimuli per category (*e. g.*, two adult female stimuli; ATSA). While we did use mean pupil dilatation for the six images per stimulus, it would be recommended that future studies use at least two sets of six images. Third, although we did seem to find the effect of emotional response on pupil dilatation, we did not have a sufficient variety of stimuli showing different levels of violence to accurately measure and describe this effect. Fourth, and as mentioned in the methods section, the missing data made it impossible to include the audio PUP results in the analyses. Asking participants to keep their eyes open and examine the blank screen was not sufficient to

generate valid results with the audio stimuli. Therefore, we recommend that future studies comparing pupillometric results with audio and visual stimuli use scrambled images while presenting their audio modality. Although not used in conjunction with audio stimuli, Attard-Johnson, Bindemann and Ciardha (2016) used a similar method to generate neutral stimuli which had the added advantage of allowing for control of image luminescence without presenting an image that was recognizable by a participant. Finally, although we did account for the luminosity of the images, the fact that the images had to be created to be coherent with the descriptions of the audio stimuli did limit the amount of control we could apply to them. This limitation could be addressed by creating stimuli with the intention of using it in with PUP.

As recommendation, future studies could expand on the present results by using two large banks of stimuli (sexual and non-sexual) presenting a continuum of violence, ranging from none to severe violence. By comparing results between sexual offenders with child victims and non-offenders it could be possible to quantify the emotional impact or influence that the violence has on pupil dilatation and account for it. This could allow for a more specific measure of sexual preference. Furthermore, the present study attempted to assess two phases of Singer's (1984) model (i.e., the aesthetic response and the genital response), it did not, however, attempt to assess the approach response. The most ecological and costly way of assessing this second phase would be to immerse the participant in a virtual environment where he would have control over his movements and at which point they would be analyzed by the researcher. A more indirect and affordable way to assess the approach response could be to measure the participants' eye movement patterns, that is, the way in which the

participant explores the visual stimuli that are in front of him (*e.g.*, actively seeking them out or avoiding them).

In conclusion, current results support PUP as a complementary measure to the phallometric assessment of child molesters. They also support the use of computer-generated characters in the field of sexual research as a possible replacement for less ethical images of real children. However, as of now, the results do not support pupillometry as a measure of sexual interests towards violence. Although more research is needed to replicate our results, the use of PUP seems to have a promising future.

Acknowledgments

We wish to acknowledge the financial support of the Social Sciences and Humanities Research Council of Canada and the Fonds de Recherche Société et Culture (Research Grant awarded to the first author).

References

- Abel, G., Huffman, J., Warberg, B. & Holland, C. (1998). Visual Reaction Time and Plethysmography as Measures of Sexual Interest in Child Molesters. *Sexual Abuse: A Journal of Research and Treatment*, 10(2), 81-95. doi:10.1023/a:1022063214826
- Abel, G. G., Jordan, A., Hand, C. G., Holland, L. A., & Phipps, A. (2001). Classification models of child molesters utilizing the Abel Assessment for sexual interest TM. *Child Abuse & Neglect*, 25(5), 703-718.
- Aboyoun, D. C. & Dabbs, J. M. (1998). The Hess pupil dilation findings: Sex or novelty? *Social Behavior and Personality: an international journal*, 26(4), 415-419. doi:10.2224/sbp.1998.26.4.415
- Akerman, G. & Beech, A. R. (2012). A systematic review of measures of deviant sexual interest and arousal. *Psychiatry, Psychology and Law*, 19(1), 118-143. doi:10.1080/13218719.2010.547161
- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.
- Association for the Treatment of Sexual Abusers. (2001). *Practice standards and guidelines for the members of the Association for the Treatment of Sexual Abusers*. Beaverton, OR: Author.
- Attard-Johnson, J., Bindemann, M. & Ciardha, C. Ó. (2016). Pupillary Response as an Age-Specific Measure of Sexual Interest. *Archives of Sexual Behavior*, 45(4), 855-870.

doi:10.1007/s10508-015-0681-3

Atwood, R. W. & Howell, R. J. (1971). Pupillometric and personality test score differences of female aggressing pedophiliacs and normals. *Psychonomic Science*, 22(2), 115-116.

doi:10.3758/bf03332522

Barker, J. G. & Howell, R. J. (1992). The plethysmograph: A review of recent literature.

Journal of the American Academy of Psychiatry and the Law Online, 20(1), 13-25.

Beatty, J. & Lucero-Wagoner, B. (2000). The pupillary system. *Handbook of*

psychophysiology (vol. 2, p. 142-162). New York, États-Unis: Cambridge University Press.

Bernhardt, P. C., Dabbs, J. M. & Riad, J. K. (1996). Pupillometry system for use in social psychology. *Behavior Research Methods, Instruments & Computers*, 28(1), 61-66.

doi:10.3758/BF03203637

Bradley, M. M., Miccoli, L., Escrig, M. A. & Lang, P. J. (2008). The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology*, 45(4), 602-607.

doi:10.1111/j.1469-8986.2008.00654.x

Chapman, L. J., Chapman, J. P. & Brelje, T. (1969). Influence of the experimenter on pupillary dilation to sexually provocative pictures. *Journal of Abnormal Psychology*,

74(3), 396-400. doi: doi: 10.1037/h0027493

Dabbs, J. M. (1997). Testosterone and pupillary response to auditory sexual stimuli.

Physiology & behavior, 62(4), 909-912. doi:10.1016/S0031-9384(97)00268-0

Dombert, B., Mokros, A., Brückner, E., Schlegl, V., Antfolk, J., Bäckström, A. & Santtila, P. (2013). The virtual people set: developing computer-generated stimuli for the

- assessment of pedophilic sexual interest. *Sexual Abuse: A Journal of Research and Treatment*, 25(6), 557-582. doi: 10.1177/1079063212469062
- Earls, C. M., Quinsey, V. L. & Castonguay, L. G. (1987). A comparison of three methods of scoring penile circumference changes. *Archives of Sexual Behavior*, 16(6), 493-500. doi:10.1007/BF01541713
- Farrall, W. R. & Card, R. D. (1988). Advancements in physiological evaluation of assessment and treatment of the sexual aggressor. *Annals of the New York Academy of Sciences*, 528(1), 261-273. doi: 10.1111/j.1749-6632.1988.tb50869.x
- Fischer, L., & Smith, G. (1999). Statistical Adequacy of the Abel Assessment for Interest in Paraphilias. *Sexual Abuse: A Journal of Research and Treatment*, 11(3), 195- 205. doi:10.1023/A:1021308224123
- Flak, V., Beech, A., & Fisher, D. (2007). Forensic assessment of deviant sexual interests: The current position. *Issues in Forensic Psychology*, 6, 70.
- Freund, K. (1963). A laboratory method for diagnosing predominance of homo-or heterosexual interest in the male. *Behaviour Research and Therapy*, 1(1), 85-93. doi:10.1016/0005-7967(63)90012-3
- Freund, K. (1965). Diagnosing heterosexual pedophilia by means of a test for sexual interest. *Behaviour Research and Therapy*, 3(4), 229-234. doi:10.1016/0005-7967(65)90031-8
- Gaither, G. A. & Plaud, J. J. (1997). The effects of secondary stimulus characteristics on men's sexual arousal. *The Journal of Sex Research*, 34(3), 231-236. doi:10.1080/00224499709551890

- Glasgow, D. V., Osborne, A., & Croxen, J. (2003). An assessment tool for investigating paedophile sexual interest using viewing time: An application of single case methodology. *British Journal of Learning Disabilities*, 31(2), 96-102.
- Goldwater, B. C. (1972). Psychological significance of pupillary movements. *Psychological Bulletin*, 77(5), 340-355. doi:10.1037/h0032456
- Goyette, M., Rouleau, J.-L. & Renaud, P. (2012). *Évaluation des intérêts sexuels auprès d'agresseurs sexuels d'enfants par la vidéooculographie et la présentation de stimuli générés par ordinateur* (doctoral dissertation).
Retrieved from <http://hdl.handle.net/1866/8714>
- Gress, C. L. (2005). Viewing time measures and sexual interest: Another piece of the puzzle. *Journal of Sexual Aggression*, 11(2), 117-125.
- Hamel, R. F. (1974). Female subjective and pupillary reaction to nude male and female figures. *The Journal of psychology*, 87(2), 171-175.
doi:10.1080/00223980.1974.9915687
- Hanley, J. A. & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143(1), 29-36.
doi: 10.1148/radiology.143.1.7063747
- Hanley, J. A. & McNeil, B. J. (1983). A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology*, 148(3), 839-843. doi:10.1148/radiology.148.3.6878708

- Harris, G. T., Rice, M. E., Quinsey, V. L., & Chaplin, T. C. (1996). Viewing time as a measure of sexual interest among child molesters and normal heterosexual men. *Behaviour Research and Therapy*, 34(4), 389-394.
- Hess, E. H., Beaver, P. W. & Shrouf, P. E. (1975). Brightness contrast effects in a pupillometric experiment. *Perception & Psychophysics*, 18(2), 125-127.
doi:10.3758/BF03204099
- Hess, E. H. & Polt, J. M. (1960). Pupil size as related to interest value of visual stimuli. *Science*, 132(3423), 349-350. doi:10.1126/science.132.3423.349
- Hess, E. H., Seltzer, A. L. & Shlien, J. M. (1965). Pupil response of hetero- and homosexual males to pictures of men and women: A pilot study. *Journal of Abnormal Psychology*, 70(3), 165. doi:10.1037/h0021978
- Imhoff, R., Schmidt, A. F., Nordsiek, U., Luzar, C., Young, A. W., & Banse, R. (2010). Viewing time effects revisited: Prolonged response latencies for sexually attractive targets under restricted task conditions. *Archives of Sexual Behavior*, 39(6), 1275-1288.
- Israel, E., & Strassberg, D. S. (2009). Viewing time as an objective measure of sexual interest in heterosexual men and women. *Archives of Sexual Behavior*, 38(4), 551-558.
- Janisse, M. P. (1973). Pupil size and affect: A critical review of the literature since 1960. *Canadian Psychologist/Psychologie canadienne*, 14(4), 311-329.
doi:10.1037/h0082230
- Kalmus, E. & Beech, A. R. (2005). Forensic assessment of sexual interest: A review. *Aggression and Violent Behavior*, 10(2), 193-217. doi:10.1016/j.avb.2003.12.002

- Krueger, R. B., Bradford, J. M., & Glancy, G. D. (1998). Interest—A Brief Description. *J Am Acad Psychiatry Law*, 26(2).
- Kuban, M., Barbaree, H. E. & Blanchard, R. (1999). A comparison of Volume and Circumference Phallometry: Response Magnitude and Method Agreement. *Archives of Sexual Behavior*, 28(4), 345-359. doi:10.1023/a:1018700813140
- Laeng, B., Sirois, S. & Gredebäck, G. (2012). Pupillometry a window to the preconscious? *Perspectives on psychological science*, 7(1), 18-27. doi: 10.1177/1745691611427305
- Lalumière, M. L. & Harris, G. T. (1998). Common questions regarding the use of phallometric testing with sexual offenders. *Sexual Abuse: A Journal of Research and Treatment*, 10(3), 227-237. doi: 10.1023/A:1021369904984
- Launay, G. (1994). The phallometric assessment of sex offenders: Some professional and research issues. *Criminal Behaviour and Mental Health*, 4(1), 48-70.
doi:10.1002/cbm.1994.4.1.48
- Laws, D. R. (1984). The assessment of dangerous sexual behavior in males. *Medicine and Law*, 3(2), 127-140. Retrieved from <http://www.scopus.com/inward/record.url?eid=2-s2.0-0021281832&partnerID=40&md5=4006dc85318a3ffdd193186872c14ee>
- Letourneau, E. J. (2002). A comparison of objective measures of sexual arousal and interest: Visual reaction time and penile plethysmography. *Sexual Abuse*, 14(3), 207-223.
- Libby, W. L., Lacey, B. C. & Lacey, J. I. (1973). Pupillary and cardiac activity during visual attention. *Psychophysiology*, 10(3), 270-294. doi:10.1111/j.1469-8986.1973.tb00526.x

- MacLeod, C. M. (1991). Half a century of research on the Stroop effect: an integrative review. *Psychological Bulletin*, 109(2), 163.
- Malcolm, P. B., Andrews, D. & Quinsey, V. L. (1993). Discriminant and predictive validity of phallometrically measured sexual age and gender preference. *Journal of Interpersonal Violence*, 8(4), 486-501. doi:10.1177/088626093008004004
- Marshall, W. L., Barbaree, H. E. & Christophe, D. (1986). Sexual offenders against female children: Sexual preferences for age of victims and type of behaviour. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, 18(4), 424-439. doi: 10.1037/h0079966
- Marshall, W. L. & Fernandez, Y. M. (2000). Phallometric testing with sexual offenders: Limits to its value. *Clinical Psychology Review*, 20(7), 807-822. doi:10.1016/S0272-7358(99)00013-6
- Mathôt, S. & Van der Stigchel, S. (2015). New Light on the Mind's Eye: The Pupillary Light Response as Active Vision. *Current Directions in Psychological Science*, 24(5), 374-378. doi:10.1177/09637214155593725
- Murphy, W., Haynes, M., Stalgaitis, S. & Flanagan, B. (1986). Differential sexual responding among four groups of sexual offenders against children. *Journal of Psychopathology and Behavioral Assessment*, 8(4), 339-353. doi:10.1007/BF00960731
- Naber, M., Alvarez, G. A. & Nakayama, K. (2013). Tracking the allocation of attention using human pupillary oscillations. *Frontiers in psychology*, 4, 919. doi:10.3389/fpsyg.2013.00919

- Nunnally, J. C., Knott, P. D., Duchnowski, A. & Parker, R. (1967). Pupillary response as a general measure of activation. *Perception & Psychophysics*, 2(4), 149-155.
doi:10.3758/BF03210310
- O'Neill, M. T. & Hinton, J. W. (1977). Pupillographic assessment of sexual interest and sexual arousal. *Perceptual and Motor Skills*, 44(3 suppl), 1278-1278.
- Peavler, W. S. & McLaughlin, J. P. (1967). The question of stimulus content and pupil size. *Psychonomic Science*, 8(12), 505-506. doi:10.3758/BF03331723
- Quinsey, V. L. & Chaplin, T. C. (1988). Penile responses of child molesters and normals to descriptions of encounters with children involving sex and violence. *Journal of Interpersonal Violence*, 3(3), 259-274. doi:10.1177/088626088003003001
- Renaud, P., Rouleau, J.-L., Proulx, J., Trottier, D., Goyette, M., Bradford, J. P. & Bouchard, S. (2010). Virtual characters designed for forensic assessment and rehabilitation of sex offenders: standardized and made-to-measure. *Journal of Virtual Reality and Broadcasting*, 7, 1860-2037. doi:10.20385/1860-2037/7.2010.5
- Rieger, G., Cash, B. M., Merrill, S. M., Jones-Rounds, J., Dharmavaram, S. M. & Savin-Williams, R. C. (2015). Sexual arousal: The correspondence of eyes and genitals. *Biological psychology*, 104, 56-64. doi:10.1016/j.biopsycho.2014.11.009
- Rieger, G. & Savin-Williams, R. C. (2012). The eyes have it: Sex and sexual orientation differences in pupil dilation patterns. *PloS one*, 7(8), e40256.
doi:10.1371/journal.pone.0040256

- Rieger, G., Savin-Williams, R. C., Chivers, M. L. & Bailey, J. M. (2015). Sexual arousal and masculinity-femininity of women. *Journal of Personality and Social Psychology*, *111*(2), 265-283. doi: 10.1037/pspp0000077
- Rosen, R. C. & Keefe, F. J. (1978). The Measurement of Human Penile Tumescence. *Psychophysiology*, *15*(4), 366-376. doi:10.1111/j.1469-8986.1978.tb01393.x
- Rosenzweig, S. (1942). The photoscope as an objective device for evaluating sexual interest. *Psychosomatic Medicine*, *4*(2), 150-158.
- Schnelle, J. F., Kennedy, M., Rutledge, A. W. & Golden, S. B. (1974). Pupillary response as indication of sexual preference in a juvenile correctional institution. *Journal of Clinical Psychology*, *30*(2), 146-150.
- Simms, T. M. (1967). Pupillary response of male and female subjects to pupillary difference in male and female picture stimuli. *Perception & Psychophysics*, *2*(11), 553-555.
doi:10.3758/BF03210265
- Singer, B. (1984). Conceptualizing sexual arousal and attraction. *Journal of Sex Research*, *20*(3), 230-240. doi:10.1080/00224498409551222
- Seto, M. C., Harris, G. T., Rice, M. E., & Barbaree, H. E. (2004). The screening scale for pedophilic interests predicts recidivism among adult sex offenders with child victims. *Archives of Sexual Behavior*, *33*(5), 455-466.
- Seto, M. C., & Lalumière, M. L. (2001). A brief screening scale to identify pedophilic interests among child molesters. *Sexual Abuse*, *13*(1), 15-25.
- Stinson, J. D., Becker, J. V. & Sales, B. D. (2008). Self-regulation and the etiology of sexual deviance: Evaluating causal theory. *Violence and Victims*, *23*(1), 35-51.

doi:10.1891/0886-6708.23.1.35

Trottier, D., Rouleau, J.-L. & Renaud, P. (2013). *Les apports de la réalité virtuelle et de la vidéo-oculographie à l'évaluation des intérêts sexuels* (Doctoral dissertation).

Retrieved from <http://hdl.handle.net/1866/9191>

Trottier, D., Renaud, P., Rouleau, J.-L., Goyette, M., Saumur, C., Boukhalfi, T. & Bouchard, S. (2015). Erratum to: Using immersive virtual reality and anatomically correct computer-generated characters in the forensic assessment of deviant sexual preferences. *Virtual Reality*, 19(3), 303-303. doi:10.1007/s10055-015-0277-1

Trottier, D., Rouleau, J.-L., Renaud, P. & Goyette, M. (2014). Using eye tracking to identify faking attempts during penile plethysmography assessment. *The Journal of Sex Research*, 51(8), 946-955. doi:10.1080/00224499.2013.832133

Wright, L. W., & Adams, H. E. (1994). Assessment of sexual preference using a choice reaction time task. *Journal of Psychopathology and Behavioral Assessment*, 16(3), 221-231.

Zuckerman, M. (1971). Physiological measures of sexual arousal in the human. *Psychological Bulletin*, 75(5), 297. doi :10.1037/h0030923

Chapitre IV - Discussion et conclusion

Retour sur la problématique

La présence d'intérêts sexuels envers des enfants demeure, à ce jour, lorsque pris individuellement, le meilleur prédicteur du risque de récidive sexuelle chez les individus ayant déjà commis une agression sexuelle envers un enfant (Hanson et Bussiere, 1998) et fait partie intégrante d'une évaluation plus large du risque de récidive sexuelle. De plus, la présence de ce type d'intérêt est l'un des critères principaux du trouble pédophilique (critère A : présence de fantasmes et d'intérêts sexuels envers les enfants; American Psychiatric Association, 2000, 2013). Il s'avère donc nécessaire, voire essentiel, d'élaborer des méthodes rigoureuses et efficaces pour évaluer la présence de ces intérêts. À cette fin, Singer (1984) propose que l'intérêt ou l'attirance sexuelle est un processus composé de trois phases ou étapes distinctes (*c.-à-d.* la réponse esthétique, la réponse d'approche et la réponse génitale). Ces trois phases constituent des cibles ou patrons de réponses physiologiques et cognitives distinctes pour sonder la présence des intérêts sexuels envers les enfants.

Toutefois, seulement l'une de ces cibles (la phase génitale) est présentement mesurée durant l'évaluation des intérêts sexuels, et ce, par l'entremise d'un outil validé : la pléthysmographie pénienne (Akerman et Beech, 2011; Freund, 1965; Hanson et Bussiere, 1998; Hanson et Morton-Bourgon, 2005; Launay, 1999; Laws, Hanson, Osborn et Greenbaum, 2000; Malcolm, Andrews et Quinsey, 1993; Proulx., 1993). Malgré une utilisation soutenue en Amérique du Nord depuis plus de 60 ans, la PPG a peu évolué dans son fonctionnement ou dans les stimuli employés pour générer les réponses sexuelles. Les limitations de cette mesure demeurent donc toujours présentes.

La principale de ces faiblesses concerne la portion de profils invalides (jusqu'à 34 %) qui entraîne en une impossibilité de classer plusieurs individus pouvant nécessiter des

services spécialisés ou des conditions de libération plus restreintes (Akerman et Beech, 2011; Barker et Howell, 1992; Farrall et Card, 1988; Kalmus et Beech, 2005; Marshall, Barbaree et Christophe, 1986; Murphy, Haynes, Stalgaitis et Flanagan, 1986; Quinsey et Chaplin, 1988; Stinson, Becker et Sales, 2008). Plusieurs raisons ont été mises de l'avant pour expliquer ce nombre de profils invalides. L'une des principales raisons de cette faiblesse est que certains stimuli couramment utilisés en PPG ne sont tout simplement pas assez excitants pour susciter une réponse génitale valide. Ainsi, plusieurs sujets ne présenteront pas une réponse érectile de 2,5 mm ou plus, seuil de validité le plus communément accepté pour la PPG circonférentielle. Une deuxième limite de la PPG concerne l'aspect éthique de la procédure et des stimuli employés. En effet, certains chercheurs considèrent que l'utilisation d'images réelles d'enfants contribue à victimiser les enfants qui y sont présentés et d'autres chercheurs considèrent même que la procédure PPG en elle-même est trop invasive pour être pleinement éthique (Abel, Huffman, Warberg et Holland, 1998; Kalmus et Beech, 2005; Launay, 1994). Heureusement, ces deux problématiques peuvent être, en partie, corrigées en travaillant sur le type de stimuli employés et en trouvant de nouvelles mesures physiologiques de l'intérêt sexuel.

La littérature sur les types de stimuli disponibles en PPG est féconde et démontre clairement que les stimuli faisant usage de plus d'une modalité de stimulation (visuelle et auditive) sont plus efficaces à générer des réponses érectiles (Abel, Blanchard et Barlow, 1981; Gaither et Plaud, 1997; Julien et Over, 1988). L'objectif étant alors d'utiliser des stimuli audiovisuels respectant des notions éthiques plus rigoureuses, mais permettant l'évaluation des intérêts sexuels. À cette fin, les nouvelles technologies permettent d'utiliser des logiciels pour créer des personnages virtuels ou synthétiques (PGO) respectant les caractéristiques

morphologiques humaines des enfants aux adultes (Goyette, Rouleau et Renaud, 2012; Trottier, Rouleau, Renaud et Goyette, 2014).

De plus, ces PGO permettent de remettre de l'avant des stimuli visuels et ouvrent la porte à l'inclusion d'une mesure complémentaire à la PPG (*p. ex.*, la pupillométrie). La PUP peut faire usage des mêmes stimuli audiovisuels que la PPG tout en permettant l'évaluation d'une autre phase de l'intérêt sexuel, soit la phase esthétique (Kalmus et Beech, 2005). De plus, des études récentes montrent que les mesures de dilatation de la pupille obtenues avec la PUP peuvent être utilisées pour différencier les intérêts sexuels de participants provenant de la population générale (Attard-Johnson, Bindemann et Ciardha, 2016; Rieger, Cash et al., 2015; Rieger et Savin-Williams, 2012; Rieger, Savin-Williams, Chivers et Bailey, 2015).

Objectifs et principaux résultats

La présente thèse avait deux objectifs principaux. Le premier consistait à agencer des PGO aux stimuli auditifs standards afin de produire des stimuli éthiques générant de plus fortes réponses pénitentes et conséquemment réduire le nombre de profils invalides avec la PPG. Le deuxième objectif, quant à lui, consistait à maximiser l'utilisation de cette nouvelle modalité audiovisuelle en démontrant qu'elle pouvait être utilisée en complémentarité avec la pupillométrie et la PPG lors de l'évaluation des intérêts sexuels envers les enfants. Ces objectifs furent abordés lors de deux études distinctes (Chapitre II et III).

Ainsi, la première étude avait pour objectif d'améliorer les stimuli auditifs déjà utilisés et éthiquement acceptables en les agencant à des personnages générés par ordinateur, aussi éthiquement acceptables, pour engendrer de plus fortes réponses pénitentes que celles observables en utilisant seulement l'une de ces deux modalités. Il était ainsi souhaité de

réduire le nombre de profils invalides (pouvant atteindre jusqu'à 34 %), tout en conservant une bonne validité discriminante et sans devoir avoir recours à des stimuli faisant usage d'images réelles d'enfants (Akerman et Beech, 2011; Marshall, Barbaree et Christophe, 1986). En revanche, la deuxième étude visait l'évaluation d'une seconde mesure aussi capable de mesurer la présence d'intérêts sexuels envers les enfants. Pour ce faire, il a été proposé d'utiliser la pupillométrie pour évaluer la première phase de l'intérêt sexuel, soit la phase esthétique. Il était attendu que cette seconde mesure permette une discrimination efficace et complémentaire des intérêts sexuels tout en palliant l'aspect invasif de la PPG. De plus, il était souhaité que cette mesure obtienne une validité discriminante comparable à la PPG.

Résultats.

Les résultats du premier article suggèrent que l'agencement de PGO aux stimuli auditifs permet effectivement d'augmenter les réponses péniennes des participants. En ce sens, l'ajout de PGO à des stimuli auditifs standardisés a permis une augmentation de 33 % de la réponse génitale des participants en comparaison à l'utilisation des stimuli auditifs présentés seuls, et une augmentation de 40 % de la réponse génitale en comparaison aux PGO présentés seuls. De plus, cette nouvelle modalité a permis de réduire significativement le nombre de profils invalides, qui est passé de 33 % pour les stimuli auditifs à 20 % pour les stimuli visuels et audiovisuels.

Parallèlement, les profils de réponses des participants ont montré des effets de groupe plus précis permettant de distinguer les participants en fonction de leurs résultats. Ainsi, les participants du groupe de comparaison ont réagi plus fortement aux stimuli adultes que les participants du groupe clinique. À l'opposé, les participants du groupe clinique ont réagi en moyenne plus fortement aux stimuli d'enfants (*c.-à-d.* GNV, BNV, BV) que les participants

du groupe de comparaison. Ces résultats étaient consistants pour les trois modalités de présentation suggérant que les stimulations auditives, visuelles et audiovisuelles opèrent de la même manière et génèrent des patrons de réponses similaires. Finalement, l'analyse des courbes de ROC a permis de démontrer que les trois modalités de stimuli discriminent efficacement les participants avec des aires sous la courbe comparables, variant entre $AUC = 0,81$ et $AUC = 0,84$. En ce sens, ces trois modalités de présentation, éthiquement responsables, permettaient de classer efficacement les participants dans leur groupe respectif sur la base de leurs patrons de résultats PPG. Plus largement, ces résultats suggèrent que l'utilisation de stimuli audiovisuels ayant recours à des personnages générés par ordinateur peut significativement augmenter la capacité de la PPG à obtenir de fortes réponses génitales, réduire le nombre de profils invalides et catégoriser efficacement des individus en fonction de leurs intérêts sexuels, et ce, tout en respectant des normes éthiques plus restrictives concernant l'utilisation des stimuli visuels employés.

Les résultats du second article suggèrent, quant à eux, que les résultats pupillométriques obtenus lors de la présentation de stimuli visuels ou audiovisuels faisant usage de PGO ne se distinguent pas des résultats PPG quant à leur capacité à discriminer entre les participants ayant déjà commis une agression sexuelle envers des enfants et ceux n'ayant jamais commis de telles agressions. En ce sens, les participants du groupe clinique ont présenté, en moyenne, de plus grandes dilatations de la pupille lors de la présentation des stimuli d'agression sexuelle de garçon sans violence, de viol de garçons et de viol de fillettes, que les participants du groupe de comparaison. À l'inverse, les participants du groupe de comparaison ont obtenu, quant à eux, de plus grandes dilatations pupillaires lors de la présentation de stimuli adultes que les participants du groupe d'agresseurs sexuels d'enfants.

De plus, cette deuxième étude suggère que les résultats PUP audiovisuels peuvent être utilisés pour générer des indices d'intérêts sexuels similaires à ceux utilisés en PPG. À cet effet, les résultats obtenus aux analyses de courbes de ROC ont montré que les indices PUP audiovisuels classifiaient efficacement les participants en fonction de leur groupe d'appartenance avec des aires sous la courbe variant entre $AUC = 0,73$ et $AUC = 0,78$, et ce, avec des résultats comparables à ceux obtenus pour les résultats PPG ($AUC = 0,82$ à $AUC = 0,84$). De plus, en combinant les indices de déviations audiovisuels PUP et PPG, les résultats démontrent que nous obtenons une plus forte capacité discriminante ($AUC = 0,89$). Néanmoins, les résultats suggèrent aussi que l'indice de déviance obtenu en utilisant les données visuelles PUP seules ne permet pas de classifier les participants dans leur groupe d'appartenance.

Plus généralement, les résultats de la seconde étude indiquent que la pupillométrie avec des stimuli audiovisuels pourrait être utilisée comme mesure complémentaire à la pléthysmographie pénienne et avec de plus amples recherches pourrait être un substitut plus éthique à la PPG dans les pays où cette mesure n'est pas acceptée. À court terme, les données pupillométriques pourraient être combinées aux données pléthysmographies pour obtenir une mesure plus fidèle et donc plus valide. Cette étude est à notre connaissance la première à utiliser la pupillométrie avec des PGO et à démontrer une discrimination efficace avec une population clinique.

Intégration des résultats

Il s'avère donc que l'agencement de PGO à des stimuli auditifs validés pourrait être un moyen judicieux d'améliorer l'efficacité de l'évaluation des agresseurs sexuels d'enfants en

réduisant le nombre de profils invalides et en permettant d'utiliser une seconde mesure (*c.-à-d.* PUP). Ainsi, les résultats obtenus au premier article sont cohérents avec la littérature suggérant que la stimulation de plus d'une modalité sensorielle est associée à de plus fortes réponses pénienues (Abel et al., 1981; Gaither et Plaud, 1997; Julien et Over, 1988) et en retour, ces réponses ont un impact positif sur le nombre de profils valides lors de l'évaluation PPG. Néanmoins, il est important de noter que malgré l'amélioration des réponses pénienues aux stimuli audiovisuels, le nombre de profils valides associés à cette modalité était identique à celui observé pour les stimuli visuels présentés seuls, ce qui supporte l'apport des stimuli visuels à la PPG et à la sexualité (Goyette et al., 2012; Trottier et al., 2014). Finalement, il ne semble pas que l'augmentation des résultats pénienus aux stimuli audiovisuels ait eu un apport direct sur la validité discriminante de la pléthysmographie lorsque le seuil minimal de validité (2,5 mm) est respecté (AUC variant entre 0.81 et 0.84). L'apport est plutôt indirect et passe par le fait que les résultats d'une plus grande partie des participants (amélioration de 13 %) peuvent être analysés.

Dans un autre ordre d'idée, l'utilisation des PGO dans cette étude a permis de prendre en compte la littérature suggérant que l'utilisation d'images réelles d'enfants constitue un problème éthique enfreignant les droits fondamentaux des enfants qui y sont présentés (Abel et al., 1998; Kalmus et Beech, 2005; Launay, 1994). Ceci est d'autant plus important étant donné que les stimuli visuels utilisés dans ce projet ont été entièrement générés par ordinateur. Cette tendance à réagir physiologiquement au PGO est similaire à ce qui a été observé dans les autres études ayant eu recours à ce type de personnages et supporte leur utilisation dans les domaines d'études sur la sexualité humaine (Goyette et al., 2012; Renaud et al., 2010; Trottier et al., 2015; Trottier et al., 2014). Les réactions des participants semblent même avoir été

cohérentes avec leur historique d'agression sexuelle puisqu'aucun des participants n'a rapporté avoir commis d'agression sexuelle avec violence (viol) envers un enfant ou avoir trouvé les scénarios de viol sexuellement attirants. Cette information est corroborée par les résultats démontrant que l'ensemble des participants (groupe clinique et de comparaison) ont moins réagi aux stimuli neutres et aux stimuli de viol qu'aux stimuli adultes.

Ce sont parallèlement ces PGO, c'est-à-dire des stimuli visuels, qui ont permis d'introduire la pupillométrie de manière éthique dans le cadre de l'évaluation des intérêts sexuels envers les enfants. Par exemple, nos résultats corroborent ceux de la littérature scientifique suggérant que la pupillométrie puisse être utilisée pour différencier des participants en fonction de leurs intérêts sexuels (Attard-Johnson et al., 2016; Rieger, Cash et al., 2015; Rieger et Savin-Williams, 2012; Rieger, Savin-Williams et al., 2015). Ils y contribuent aussi en montrant que les résultats PUP diffèrent en fonction du groupe d'appartenance des participants. Ainsi, les résultats suggèrent que les données pupillométriques audiovisuelles peuvent être interprétées et traitées comme des données phallométriques pour calculer des indices d'intérêts sexuels qui permettent de classifier efficacement les participants en fonction de leur groupe d'appartenance. De plus, les résultats suggèrent que les indices de déviance obtenus à l'aide des stimuli audiovisuels (PPG et PUP) peuvent être combinés pour obtenir un nouvel indice présentant un haut niveau de discrimination ($AUC = 0,89$). Il est cependant important de noter que malgré l'efficacité des indices de déviance visuels en PPG à catégoriser les participants, les indices PUP obtenus avec les mêmes stimuli n'ont pas été efficaces. Il est possible que l'incorporation des stimuli auditifs pour générer la modalité audiovisuelle ait été le facteur ayant accordé à cette modalité sa capacité discriminante supérieure. Néanmoins, puisque les résultats visuels PUP présentent

des données manquantes trop importantes, il n'est présentement pas possible de confirmer ou d'infirmier cette hypothèse.

Parallèlement, nos résultats ont montré que les participants ont réagi de manière plus importante aux stimuli de viol qu'aux autres stimuli d'enfants. Ceci semble cohérent avec la littérature suggérant que la dilatation de la pupille varie en fonction de la charge émotionnelle générée par le stimulus (Janisse, 1973). Par ailleurs, les résultats suggèrent que malgré ces réactions émotionnelles, une part substantielle de la dilatation de la pupille pourrait être attribuable à un effet de préférence, puisque les résultats semblent avoir varié en fonction du groupe d'appartenance (clinique et de comparaison). Il semblerait donc que les données pupillométriques audiovisuelles offrent une manière efficace d'évaluer la phase esthétique de l'intérêt sexuelle. En ce sens, en utilisant à la fois les données PPG et PUP obtenues lors de la présentation du même ensemble de stimuli (audiovisuel), il serait possible d'évaluer la phase esthétique et la phase génitale des intérêts sexuels, et ce faisant, avoir une évaluation plus complète des intérêts sexuels sans avoir à compromettre les notions d'éthiques.

Implications théoriques.

Même si les principaux objectifs de cette thèse avaient des aspirations cliniques, plusieurs implications théoriques découlent des résultats obtenus. La première consiste en la manière de traiter les données pupillométriques. Contrairement à la pléthysmographie où un niveau de base peut être facilement obtenu (détumescence), le niveau de base en pupillométrie est moins évident puisqu'il peut être influencé par plusieurs facteurs. C'est pour cette raison que les études dans ce domaine font généralement usage d'un stimulus neutre dont les résultats sont soustraits aux stimuli sexuels (Rieger et Savin-Williams, 2012). Ce stimulus neutre agit

comme niveau de base pour les stimuli sexuels. Le deuxième article vient donc supporter l'importance d'avoir recours à un stimulus neutre lors de l'évaluation des données PUP.

De plus, les résultats suggèrent qu'en récoltant et en traitant les données pupillométriques, il devient alors possible d'avoir accès à une mesure de la phase esthétique de l'intérêt sexuel. Plusieurs chercheurs avaient suggéré que la PUP pourrait être utilisée de la sorte (Kalmus et Beech, 2005), mais peu de données empiriques sur une population clinique étaient disponibles pour supporter cette notion. La littérature montrait qu'il existait des différences significatives quant à la dilatation de la pupille lors de la présentation de stimuli sexuels entre les hommes et les femmes ou entre des participants hétérosexuels et homosexuels (Attard-Johnson et al., 2016; Atwood et Howell, 1971; Hamel, 1974; O'Neill et Hinton, 1977; Rieger, Cash et al., 2015; Rieger et Savin-Williams, 2012; Schnelle, Kennedy, Rutledge et Golden, 1974). Nos résultats corroborent et enrichissent les données précédentes en montrant que le même processus s'applique aussi aux intérêts sexuels envers les enfants. Dans un même ordre d'idée, les résultats obtenus soutiennent que les deux mesures utilisées peuvent agir comme indice des facteurs neuropsychologiques qui sont une partie centrale de la théorie intégrée de l'agression sexuelle (Ward et Beech, 2006).

Une autre contribution théorique, celle-ci associée au premier article, concerne la méthode utilisée pour générer les stimuli audiovisuels. Les études précédentes, ayant démontré que le stimulus multimodal générait des réponses péniennes supérieures aux stimuli simples avaient fait usage d'une méthode soustractive pour créer leurs stimuli. Par exemple, Gaither et Plaud (1997), de même que Julien et Over (1988) ont tous les deux débuté par l'utilisation d'une vidéo sexuelle (audiovisuelle) qu'ils ont ensuite modifiée pour obtenir leurs autres stimuli (visuels, auditifs et autres). Ils ont donc réduit la qualité de leur stimulus initial et ils

ont démontré que cela était associé à une diminution des réponses sexuelles. À l’opposé, nous avons fait usage d’une méthode additive au cours de laquelle nous avons créé des stimuli complémentaires aux stimuli auditifs. Nous avons ensuite démontré que ces stimuli étaient associés à de plus fortes réponses sexuelles, supportant ainsi l’importance des stimuli multimodaux et suggérant l’utilisation d’une méthode additive lors de la création de stimuli destinés à évaluer les intérêts sexuels.

Parallèlement, cette méthode additive a eu pour conséquence d’augmenter l’amplitude des réponses pénienues aux stimuli, mais elle ne semble pas avoir eu d’impact majeur sur la distinction ou la discrimination entre ces stimuli. En ce sens, l’augmentation de la réponse en données brutes fut importante et significative, mais fut présente pour l’ensemble des stimuli, faisant en sorte que la différence relative entre les résultats d’un même participant est demeurée constante. Cet effet est aussi présent pour les participants du groupe de comparaison qui ne rapportent, de prime abord, aucun intérêt sexuel envers les stimuli incluant des enfants, mais dont leurs résultats se sont vus amplifiés par ce type de stimuli. Ces résultats semblent corroborer la crainte rapportée par certains chercheurs en ce qui a trait à l’effet plafond possible lorsque les stimuli utilisés consistent en des vidéos (Kalmus et Beech, 2005; Proulx, 1989). Ces chercheurs stipulent que la modalité vidéo pourrait mener à des érections complètes aux différents types de stimuli (adultes et enfants), résultant en une incapacité à discriminer les intérêts sexuels préférentiels des participants. Ainsi, une fois la pleine érection atteinte à l’une des catégories de stimulus, toute augmentation de tumescence sur les autres catégories mènerait nécessairement à une diminution de la capacité discriminative de l’évaluation.

Finalement, les deux articles contribuent à la littérature sur la réalité virtuelle en montrant que les participants ont réagi aux PGO de manière attendue sur deux mesures physiologiques distinctes (pléthysmographie et pupillométrie). Ces résultats supportent l'utilisation des PGO et suggèrent que l'aspect essentiel du traitement de l'information n'est pas la véracité de l'image (synthétique ou réelle), mais plutôt son contenu qui suscite des réponses physiologiques similaires.

Implications cliniques et perspective d'avenir.

Plusieurs implications cliniques émanent de cette thèse. La première porte sur le développement d'une méthode efficace pour améliorer les stimuli auditifs déjà utilisés dans la majorité des laboratoires de pléthysmographie en Amérique du Nord. Cette méthode consiste en la juxtaposition de personnages créés par ordinateur représentant visuellement ce qui est décrit dans les vignettes auditives. Cette procédure représente une manière peu coûteuse d'améliorer divers stimuli déjà validés et déjà en cours d'utilisation, sans avoir recours à des images réelles d'enfants. Ces PGO peuvent être facilement manipulés pour correspondre aux exigences de tous scénarios auditifs, incluant les gestes sexuels commis, les caractéristiques physiques de personnages (couleur des cheveux, des yeux, ethnie), et ce, sans devoir s'inquiéter de la langue du scénario auditif utilisé. Les personnages peuvent donc être modélés pour représenter des enfants et des adultes de différentes ethnies ayant différentes caractéristiques morphologiques.

De plus, les deux articles contribuent à mettre de l'avant l'utilisation des PGO comme substitut éthique aux images réelles dans l'évaluation des intérêts sexuels envers les enfants et supportent la littérature déjà présente à cet effet (Chartier et al., 2006; Dennis, Rouleau, Renaud, Nolet et Saumur, 2014; Goyette et al., 2012; Renaud et al., 2010; Trottier, Rouleau et

Renaud, 2013; Trottier et al., 2015). Ces PGO remédient donc aux critiques éthiques associées aux stimuli visuels et remettent de l'avant cette modalité considérée par plusieurs comme étant essentielle à la stimulation de l'intérêt sexuel (Singer, 1984). De plus, ce sont ces PGO qui permettraient à la pupillométrie d'être utilisée cliniquement au Canada. En ce sens, l'article deux supporte la littérature récente montrant le potentiel clinique de la PUP auprès d'échantillons normatifs (Attard-Johnson et al., 2016; Rieger, Cash et al., 2015; Rieger et Savin-Williams, 2012). Si cette tendance se maintient, la PUP pourrait agir comme mesure complémentaire à la PPG d'ici quelques années et pourrait potentiellement agir comme une mesure substitut à la PPG dans les pays où celle-ci n'est pas permise. Ainsi, la PUP combinée aux outils actuellement validés pourrait faire partie intégrante du processus d'évaluation légale et être un atout important pour prendre des décisions concernant les besoins d'interventions cliniques des patients ou des conditions légales à imposer afin de protéger la population générale.

Une des implications cliniques majeures non abordées jusqu'à présent concerne l'aspect financier associé aux PGO et à la PUP. Contrairement à une image photographique qui est difficilement changeable, une fois créés, les PGO peuvent être utilisés dans différents scénarios pour générer un large éventail de stimuli (adulte ou enfant, avec ou sans violence, hétérosexuel ou homosexuel), et ce, à faible coût. De plus, l'ajout de ces PGO aux bandes auditives n'allonge pas la durée du protocole d'évaluation, mais elle augmente l'amplitude des réponses et le nombre de profils valides. Le même avantage peut aussi être mis de l'avant quant à l'utilisation de la pupillométrie, puisque son coût d'achat est relativement faible et que les stimuli utilisés sont les mêmes que pour l'évaluation pléthysmographique. De plus, puisque la pupillométrie est une mesure peu invasive, elle a l'avantage de ne pas être encombrante

lorsqu'ajoutée à l'évaluation pléthysmographique standard. En somme, le matériel de base nécessaire pour compléter ces évaluations est relativement peu coûteux, peu encombrant et facile d'utilisation.

Finalement, une implication clinique non négligeable de la PUP est que contrairement à la PPG qui peut seulement être effectuée auprès des hommes, la PUP pourrait servir comme mesure physiologique des intérêts sexuels auprès des femmes qui ont commis des agressions sexuelles envers des enfants. Il serait alors pertinent pour de futures études de tenter de répliquer nos résultats à la fois auprès d'une population clinique d'hommes présentant des intérêts sexuels envers des enfants, mais aussi auprès d'une population clinique de femmes.

Implications légales et éthiques. L'évaluation des intérêts sexuels envers les enfants par l'utilisation de matériel pornographique présentant des enfants soulève de manière intrinsèque des notions d'éthique et de droits (*c.-à-d.* les droits des personnes présentés par les stimuli et les droits des cliniciens qui les utilisent). Par exemple, selon l'article 163.1.(1).a du *Code Criminel* (1985), la pornographie juvénile est définie comme étant :

« [t]oute représentation photographique, filmée, vidéo ou autre, réalisée ou non par des moyens mécaniques ou électroniques : soit où figure une personne âgée de moins de dix-huit ans ou présentée comme telle et se livrant ou présentée comme se livrant à une activité sexuelle explicite, soit dont la caractéristique dominante est la représentation, dans un but sexuel, d'organes sexuels ou de la région anale d'une personne âgée de moins de dix-huit ans ».

Cette définition s'élargit ensuite dans les sections subséquentes b), c) et d) pour inclure tout document écrit ou enregistrement sonore qui préconisent l'activité sexuelle avec une personne âgée de moins de 18 ans ou dont la caractéristique principale est la description sexuelle. La loi canadienne est donc sans équivoque et statue clairement que les bandes sonores standardisées, les images générées par ordinateur, ainsi que la combinaison des deux

types de stimuli, constituent de la pornographie juvénile. La particularité du code canadien est qu'il permet une exception légale dans certaines circonstances, tel que stipulé par l'article 163.1.6.a du *Code Criminel* (1985) « [n]ul ne peut être déclaré coupable d'une infraction au présent article si les actes qui constitueraient l'infraction : a) ont un but légitime lié à l'administration de la justice, à la science, à la médecine, à l'éducation ou aux arts ».

Dans cet ordre d'idée, les nouveaux stimuli créés pour ce projet doctoral constituent légalement de la pornographie juvénile dont le but est lié à l'administration de la justice (évaluation pré-sentence et prédiction du risque de récidive), l'avancement de la science (recherche) et des offres de services cliniques. Cette définition permettrait donc techniquement l'usage d'images réelles d'enfants lors d'évaluation légales et cliniques d'agresseurs sexuels d'enfants. Néanmoins, cette utilisation est proscrite au Canada, pour d'autres raisons éthiques incluant la notion de victimisation des enfants. Les enfants présentés dans les images n'ont jamais donné leur consentement et ne peuvent donc pas le donner pour être présentés de la sorte lors d'une évaluation. Or, les stimuli générés par ordinateur ainsi que les bandes auditives sont, en tous points, fictifs et ne génèrent donc pas de victime. La nouvelle modalité audiovisuelle serait donc acceptable en recherche et en clinique et représenterait un moyen éthique d'utiliser des stimuli visuels sans compromettre les droits fondamentaux d'individus canadiens.

Limites.

Les deux articles de la thèse présentent des limites similaires, inhérentes à la méthodologie utilisée et à la population clinique ciblée. La première de ces limites concerne le faible nombre de participants (15 par groupe). Cette limite est principalement attribuable à l'accessibilité de la population cible (les agresseurs sexuels d'enfants) et au fait que la

participation aux deux projets devait être volontaire. En ce sens, la participation à ce projet a été offerte à plus de 100 hommes adultes ayant déjà commis une agression sexuelle sur un mineur. De ce nombre, seulement 15 ont accepté. Plusieurs participants potentiels ont mentionné ne pas voir l'intérêt de participer à ce type de projet. De plus, malgré le fait que la confidentialité des résultats était garantie et ce fait, clairement communiqué aux participants potentiels, plusieurs d'entre eux ont avoué toutefois craindre que les résultats obtenus puissent les désavantager au sur le plan légal. Cette crainte était d'ailleurs la raison principale pour laquelle plusieurs participants potentiels ont refusé de participer à la recherche. De plus, des participants recrutés au Service Correctionnel du Canada ont révélé des appréhensions à revoir ou à découvrir l'Institut sécuritaire Philippe-Pinel où avait lieu l'expérimentation.

La fermeture de l'unité de traitement des agresseurs sexuels de l'Institut Phillippe-Pinel de Montréal a aussi contribué à diminuer le bassin de population susceptible de participer au projet de recherche. Ce manque d'accessibilité est la principale raison pour laquelle un groupe d'agresseurs sexuels de femmes adultes n'a pas été inclus dans l'étude comme deuxième groupe de comparaison. La majorité des patients vue à l'IPPM sont référés pour une évaluation des intérêts sexuels envers des enfants et une proportion significative de ces patients n'était pas éligible à participer au projet puisqu'ils n'avaient pas eu de contact physique avec un enfant : exhibitionnisme ou consommation de pornographie juvénile. Il est néanmoins important de noter que malgré ce faible nombre de participants, les deux études ont été en mesure d'obtenir des tailles d'effets moyennes et larges quant aux variables clés et répondre aux objectifs.

Une seconde limite de cette thèse concerne les caractéristiques sociodémographiques des participants. L'échantillon utilisé est composé de deux groupes d'hommes présentant de

faibles niveaux socio-économiques en comparaison avec la population générale. Certaines études suggèrent qu'il existe peu de différences à ce sujet entre les agresseurs sexuels d'enfants et les individus de la population générale (Curtin et Niveau, 1998). Il est donc possible que la généralisation des résultats soit négativement influencée par le faible niveau socio-économique des participants. Ce revenu est cependant similaire à celui d'autres études ayant recours à la pléthysmographie pénienne (Goyette, Rouleau et Renaud, 2012; Trottier, Rouleau, Renaud et Goyette, 2014).

Une troisième limite concerne l'utilisation d'un seul stimulus par catégorie de présentation (*p. ex.*, un stimulus de viol de garçon). *L'Association for the Treatment of Sexual Offenders* recommande l'utilisation d'au moins deux stimuli par catégorie (ATSA, 2001). Ce choix a été effectué pour assurer une durée raisonnable du protocole de recherche (*c.-à-d.* deux heures et demie) tout en permettant la comparaison des trois modalités de présentation. L'utilisation de deux stimuli par catégorie aurait impliqué un protocole de trois à quatre heures, augmentant de manière importante le risque d'un effet de fatigue. Il est probable que cette limite aurait de manière similaire diminué l'exactitude ou la portée des résultats de la seconde étude.

La méthodologie utilisée dans cette thèse (mesure répétée) constitue une quatrième faiblesse potentielle puisque tous les participants ont vu chacune des trois modalités de présentation. Malgré le fait que l'ordre des présentations ait été contrôlé à la fois entre les modalités et à l'intérieur des modalités, il est possible que cette exposition répétée aux mêmes histoires ait provoqué un effet d'habituation qui aurait pu diminuer l'amplitude des réponses des participants. Si tel est le cas, cet effet aurait été réparti à l'ensemble des stimuli. Ceci

implique qu'une étude tentant de répliquer nos résultats avec des groupes distincts obtiendrait probablement des résultats d'une plus forte intensité.

Finalement, les résultats obtenus aux stimuli de viol suggèrent une influence de la charge émotionnelle ressentie par les participants face à la violence sur leur dilatation de la pupille. Néanmoins, les stimuli utilisés présentaient peu de variabilité dans la violence avec seulement deux des stimuli présentant un haut niveau de violence et les autres n'en présentant aucun. Cette faible variabilité quant au contenu des stimuli représente la quatrième limite de cette thèse. Encore une fois, cette limite a été introduite pour éviter un effet de fatigue. Néanmoins, l'utilisation de stimuli présentant différents niveaux de violence aurait permis d'évaluer plus efficacement l'effet de la charge émotionnelle sur la dilatation de la pupille. Par exemple, elle aurait permis d'observer si la dilatation de la pupille est proportionnelle au niveau de violence observée et même d'évaluer si un effet plafond de cette réaction à la violence est possible.

Futures recherches.

Afin de valider les stimuli audiovisuels et l'utilisation clinique de la PUP à des fins d'évaluation, il apparaît essentiel que nos résultats soient répliqués dans une étude comportant un plus grand échantillon. De plus, étant donné les résultats positifs de la présente thèse, de futures recherches pourraient pallier certaines des limites mentionnées précédemment en ayant recours à un plan quasi expérimental (groupe clinique et de comparaison) et en présentant seulement la modalité audiovisuelle. Les gains de temps associés à l'utilisation d'une seule modalité permettraient l'utilisation d'au moins deux stimuli par catégorie et l'ajout de stimuli présentant différents niveaux de violence et de coercition. Il serait alors possible de pallier l'effet d'habituation inhérent à la mesure répétée tout en contrôlant l'effet de fatigue. Cette étude permettrait de corroborer la capacité discriminante de la modalité audiovisuelle pour la

PPG et la PUP tout en définissant davantage l'effet de la charge émotionnelle d'un stimulus sur les résultats PUP.

Finalement, il nous apparaît important d'aborder l'importance d'évaluer une troisième mesure physiologique pour l'évaluation des intérêts sexuels afin d'obtenir une réelle triangulation de l'information, ce qui n'était pas possible avec les mesures et la méthodologie utilisées dans cette thèse. Une solution possible serait de tenter de mesurer la réponse d'approche du modèle conceptuel de l'intérêt sexuel de Singer (1984). Différentes méthodes et mesures seraient à considérer, incluant le recours à l'immersion des participants dans un monde virtuel dans lequel il serait alors possible de mesurer et d'analyser leurs déplacements. Cette méthode s'avère dispendieuse, mais le développement et la mise en marché de nouvelles technologies devrait rendre l'immersion virtuelle plus accessible à l'ensemble des chercheurs et cliniciens. Une méthode alternative consisterait à utiliser l'oculométrie pour évaluer les patrons de regard des participants. Trottier et ses collaborateurs (2014) suggèrent que les patrons oculaires pourraient être un moyen d'évaluer la phase esthétique de l'intérêt sexuel. Néanmoins, la phase esthétique est fondamentalement constituée de réponses automatiques de courtes durées qui demeurent, hors du contrôle du participant. Or, il nous apparaît plutôt que les patrons oculaires sur lesquels les participants ont un contrôle, alors qu'ils explorent les stimuli, pourraient être utilisés comme substitut de la deuxième phase : la phase d'approche. Il est clair que dans un contexte plus naturaliste, la phase d'approche serait mesurée par la diminution de la distance entre les participants et le stimulus. Néanmoins, le contexte d'une évaluation PPG et PUP nécessite que le participant limite de tels mouvements par peur d'accrocher la jauge ou de faire varier la distance du participant de la source de stimulation.

Les mouvements oculaires pourraient donc agir comme substitut aux mouvements plus naturalistes pour l'évaluation de la phase d'approche.

Conclusion

Les résultats obtenus dans cette thèse constituent un avancement important des connaissances au plan clinique et théorique. Ils influencent à la fois les connaissances sur l'évaluation des préférences sexuelles, sur la pléthysmographie pénienne, sur la pupillométrie et sur l'utilisation de personnages générés par ordinateur. Dans un premier temps, les résultats obtenus supportent l'hypothèse qu'il est possible d'utiliser des PGO en PPG et contribuent à l'avancement des connaissances en démontrant que ces PGO permettent d'augmenter l'amplitude des réponses des participants tout en diminuant le nombre de profils invalides. Dans un deuxième temps, les résultats des travaux des deux articles contribuent à remettre de l'avant les stimuli visuels tout en respectant des standards éthiques plus rigoureux. Ce retour aux stimuli visuels est ce qui a permis l'utilisation et la validation partielle de la pupillométrie comme mesure d'évaluation des intérêts sexuels. La contribution clinique et théorique à ce niveau est la plus importante, car cette recherche est à notre connaissance la première à avoir démontré que l'usage de la pupillométrie pouvait être pertinente pour distinguer la présence d'intérêts sexuels cliniques avec des personnages générés par ordinateur. De plus, nos stimuli présentant des scénarios avec et sans coercition ont permis de démontrer à la fois la capacité discriminante de la PUP, et la présence de variations pupillaires associées à l'intensité de l'émotion générée par les stimuli.

Il est souhaité que ces résultats encouragent la recherche et le développement des nouvelles technologies dans le domaine de l'évaluation des intérêts sexuels. Ces technologies

nous permettent d'améliorer des outils déjà en utilisation depuis des décennies tout en encourageant l'ajout de nouvelles mesures opérationnelles. De plus, les nouvelles technologies nous permettent des avancements cliniques tout en respectant de hauts niveaux d'éthique et de rigueur. L'affermissement de ces nouvelles approches est donc espéré, attendu et nécessaire afin de poursuivre les progrès et revigorer le domaine de la recherche sur l'évaluation des intérêts sexuels.

Références

- Abel, G., Blanchard, E. B. et Barlow, D. H. (1981). Measurement of sexual arousal in several paraphilias: The effects of stimulus modality, instructional set and stimulus content on the objective. *Behaviour Research and Therapy*, 19(1), 25-33.
doi:[http://dx.doi.org/10.1016/0005-7967\(81\)90109-1](http://dx.doi.org/10.1016/0005-7967(81)90109-1)
- Abel, G., Huffman, J., Warberg, B. et Holland, C. (1998). Visual Reaction Time and Plethysmography as Measures of Sexual Interest in Child Molesters. *Sexual Abuse: A Journal of Research and Treatment*, 10(2), 81-95. doi:10.1023/a:1022063214826
- Akerman, G. et Beech, A. R. (2012). A systematic review of measures of deviant sexual interest and arousal. *Psychiatry, Psychology and Law*, 19(1), 118-143.
doi:10.1080/13218719.2010.547161
- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.
- Association for the Treatment of Sexual Abusers. (2001). *Practice standards and guidelines for the members of the Association for the Treatment of Sexual Abusers*. Beaverton, OR: Author.
- Attard-Johnson, J., Bindemann, M. et Ciardha, C. Ó. (2016). Pupillary Response as an Age-Specific Measure of Sexual Interest. *Archives of Sexual Behavior*, 45(4), 855-870.
doi:10.1007/s10508-015-0681-3

- Atwood, R. W. et Howell, R. J. (1971). Pupillometric and personality test score differences of female aggressing pedophiliacs and normals. *Psychonomic Science*, 22(2), 115-116.
doi:10.3758/bf03332522
- Barker, J. G. et Howell, R. J. (1992). The plethysmograph: A review of recent literature. *Journal of the American Academy of Psychiatry and the Law Online*, 20(1), 13-25.
- Chartier, S., Renaud, P., Bouchard, S., Proulx, J., Rouleau, J. L., Fedoroff, P. et Bradford, J. P. (2006). Sexual preference classification from gaze behavior data using a multilayer perceptron. *Annual Review of CyberTherapy and Telemedecine*, 4, 149-157.
- Curtin, F., & Niveau, G. (1998). Psychosocial profile of Swiss sexual offenders. *Journal of Forensic Science*, 43(4), 755-759.
- Dennis, E., Rouleau, J.-L., Renaud, P., Nolet, K. et Saumur, C. (2014). A pilot development of virtual stimuli depicting affective dispositions for penile plethysmography assessment of sex offenders. *The Canadian Journal of Human Sexuality*, 23(3), 200-208.
doi:10.3138/cjhs.2529
- Farrall, W. R. et Card, R. D. (1988). Advancements in physiological evaluation of assessment and treatment of the sexual aggressor. *Annals of the New York Academy of Sciences*, 528(1), 261-273. doi: 10.1111/j.1749-6632.1988.tb50869.x
- Freund, K. (1965). Diagnosing heterosexual pedophilia by means of a test for sexual interest. *Behaviour Research and Therapy*, 3(4), 229-234. doi:10.1016/0005-7967(65)90031-8
- Gaither, G. A. et Plaud, J. J. (1997). The effects of secondary stimulus characteristics on men's sexual arousal. *The Journal of Sex Research*, 34(3), 231-236.
doi:10.1080/00224499709551890

Goyette, M., Rouleau, J.-L. et Renaud, P. (2012). *Évaluation des intérêts sexuels auprès d'agresseurs sexuels d'enfants par la vidéooculographie et la présentation de stimuli générés par ordinateur* (doctoral dissertation).

Repéré à <http://hdl.handle.net/1866/8714>

Hamel, R. F. (1974). Female subjective and pupillary reaction to nude male and female figures. *The Journal of psychology*, 87(2), 171-175.

doi:10.1080/00223980.1974.9915687

Hanson, R. K. et Bussière, M. T. (1998). Predicting relapse: A meta-analysis of sexual offender recidivism studies. *Journal of Consulting and Clinical Psychology*, 66(2), 348-362. doi:10.1037/0022-006X.66.2.348

Hanson, R. K. et Morton-Bourgon, K. E. (2005). The Characteristics of Persistent Sexual Offenders: A Meta-Analysis of Recidivism Studies. *Journal of Consulting and Clinical Psychology*, 73(6), 1154-1163. doi:10.1037/0022-006X.73.6.1154

Janisse, M. P. (1973). Pupil size and affect: A critical review of the literature since 1960. *Canadian Psychologist/Psychologie canadienne*, 14(4), 311-329.

doi:10.1037/h0082230

Julien, E. et Over, R. (1988). Male sexual arousal across five modes of erotic stimulation. *Archives of Sexual Behavior*, 17(2), 131-143. doi:10.1007/BF01542663

Kalmus, E. et Beech, A. R. (2005). Forensic assessment of sexual interest: A review.

Aggression and Violent Behavior, 10(2), 193-217. doi:10.1016/j.avb.2003.12.002

- Launay, G. (1994). The phallometric assessment of sex offenders: Some professional and research issues. *Criminal Behaviour and Mental Health*, 4(1), 48-70.
doi:10.1002/cbm.1994.4.1.48
- Launay, G. (1999). The phallometric assessment of sex offenders: an update. *Criminal Behaviour and Mental Health*, 9(3), 254-274. doi:10.1002/cbm.317
- Laws, D. R., Hanson, R. K., Osborn, C. A. et Greenbaum, P. E. (2000). Classification of Child Molesters by Plethysmographic Assessment of Sexual Arousal and a Self-Report Measure of Sexual Preference. *Journal of Interpersonal Violence*, 15(12), 1297-1312.
doi:10.1177/088626000015012004
- Loi de 2016 sur la définition de pornographie juvénile*, LRC 1985, c 46, art 163.1.
- Malcolm, P. B., Andrews, D. et Quinsey, V. L. (1993). Discriminant and predictive validity of phallometrically measured sexual age and gender preference. *Journal of Interpersonal Violence*, 8(4), 486-501. doi:10.1177/088626093008004004
- Marshall, W. L., Barbaree, H. E. et Christophe, D. (1986). Sexual offenders against female children: Sexual preferences for age of victims and type of behaviour. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, 18(4), 424-439. doi: 10.1037/h0079966
- Murphy, W., Haynes, M., Stalgaitis, S. et Flanagan, B. (1986). Differential sexual responding among four groups of sexual offenders against children. *Journal of Psychopathology and Behavioral Assessment*, 8(4), 339-353. doi:10.1007/BF00960731
- O'Neill, M. T. et Hinton, J. W. (1977). Pupillographic assessment of sexual interest and sexual arousal. *Perceptual and Motor Skills*, 44(3 suppl), 1278-1278.

- Proulx, J. (1989). Sexual preference assessment of sexual aggressors. *International Journal of Law and Psychiatry*, 12(4), 275-280. doi:10.1016/0160-2527(89)90019-8
- Proulx, J. (1993). L'évaluation des préférences sexuelles. Dans J. Aubut (dir.), *Les agresseurs sexuels : Théorie, évaluation et traitement* (p. 98-106). Montréal, Canada: De la Chenelière.
- Quinsey, V. L. et Chaplin, T. C. (1988). Penile responses of child molesters and normals to descriptions of encounters with children involving sex and violence. *Journal of Interpersonal Violence*, 3(3), 259-274. doi:10.1177/088626088003003001
- Renaud, P., Rouleau, J.-L., Proulx, J., Trottier, D., Goyette, M., Bradford, J. P. et Bouchard, S. (2010). Virtual characters designed for forensic assessment and rehabilitation of sex offenders: standardized and made-to-measure. *Journal of Virtual Reality and Broadcasting*, 7, 1860-2037. doi:10.20385/1860-2037/7.2010.5
- Rieger, G., Cash, B. M., Merrill, S. M., Jones-Rounds, J., Dharmavaram, S. M. et Savin-Williams, R. C. (2015). Sexual arousal: The correspondence of eyes and genitals. *Biological psychology*, 104, 56-64. doi:10.1016/j.biopsycho.2014.11.009
- Rieger, G. et Savin-Williams, R. C. (2012). The eyes have it: Sex and sexual orientation differences in pupil dilation patterns. *PloS one*, 7(8), e40256. doi:10.1371/journal.pone.0040256
- Rieger, G., Savin-Williams, R. C., Chivers, M. L. et Bailey, J. M. (2015). Sexual arousal and masculinity-femininity of women. *Journal of Personality and Social Psychology*, 111(2), 265-283. doi: 10.1037/pspp0000077

- Schnelle, J. F., Kennedy, M., Rutledge, A. W. et Golden, S. B. (1974). Pupillary response as indication of sexual preference in a juvenile correctional institution. *Journal of Clinical Psychology*, 30(2), 146-150.
- Singer, B. (1984). Conceptualizing sexual arousal and attraction. *Journal of Sex Research*, 20(3), 230-240. doi:10.1080/00224498409551222
- Stinson, J. D., Becker, J. V. et Sales, B. D. (2008). Self-regulation and the etiology of sexual deviance: Evaluating causal theory. *Violence and Victims*, 23(1), 35-51.
doi:10.1891/0886-6708.23.1.35
- Trottier, D., Rouleau, J.-L. et Renaud, P. (2013). *Les apports de la réalité virtuelle et de la vidéo-oculographie à l'évaluation des intérêts sexuels* (Doctoral dissertation). Repéré à <http://hdl.handle.net/1866/9191>
- Trottier, D., Renaud, P., Rouleau, J.-L., Goyette, M., Saumur, C., Boukhalfi, T. et Bouchard, S. (2015). Erratum to: Using immersive virtual reality and anatomically correct computer-generated characters in the forensic assessment of deviant sexual preferences. *Virtual Reality*, 19(3), 303-303. doi:10.1007/s10055-015-0277-1
- Trottier, D., Rouleau, J.-L., Renaud, P. et Goyette, M. (2014). Using eye tracking to identify faking attempts during penile plethysmography assessment. *The Journal of Sex Research*, 51(8), 946-955. doi:10.1080/00224499.2013.832133
- Ward, T., & Beech, A. (2006). An integrated theory of sexual offending. *Aggression and Violent Behavior*, 11(1), 44-63. doi:<http://dx.doi.org/10.1016/j.avb.2005.05.002>