Université de Montréal

Le rôle médiateur du biais d'attribution d'intention hostile dans la relation entre l'agressivité et la personnalité antisociale : une étude des potentiels reliés aux évènements

par

Adriana Ursulet

Département de psychologie

Faculté des arts et des sciences

Mémoire présenté en vue de l'obtention du grade de Maîtrise ès sciences en Psychologie

Août 2020

Université de Montréal

Département de psychologie, Faculté des arts et des sciences

Ce mémoire intitulé

Le rôle médiateur du biais d'attribution d'intention hostile dans la relation entre l'agressivité et la personnalité antisociale : une étude de potentiels reliés aux évènements

Présenté par

Adriana Ursulet

A été évalué par un jury composé des personnes suivantes

Arnaud Saj
Président-rapporteur

Jean Gagnon
Directeur de recherche

Jean Proulx
Membre du jury

1. RÉSUMÉ

Tous les jours, dans le monde, des comportements agressifs sont commis à l'égard d'individus, causant des préjudices physiques, psychologiques et financiers. En réponse à une provocation, ces agressions sont dites réactives et peuvent être alimentées par des biais cognitifs d'attribution d'intention hostile et des styles de personnalité antisociale. Comblant un trou dans la littérature scientifique, cette étude a pour but d'évaluer le biais d'attribution d'intention hostile ainsi que son rôle dans la relation entre la personnalité antisociale et l'agressivité réactive. Dans cette perspective, les participants étaient invités à répondre à des questionnaires évaluant la personnalité, les processus cognitifs et l'agressivité. Puis, pendant l'enregistrement de leur activité cérébrale, ils devaient lire des scénarios d'interactions sociales et attribuer une intention aux comportements décrits comme ambigus et provocateurs. Nous avons analysé la N400, une composante de potentiels reliés aux évènements, associée à la présentation d'intentions inattendues hostiles ou non hostiles après chaque scénario. Des analyses de corrélations de Pearson et de régressions linéaires multiples ont été réalisées pour examiner la validité de notre modèle de médiation. Les résultats montrent que la N400 est plus forte lors de la présentation d'intention non hostile inattendue que lors de la présentation d'intention hostile inattendue dans les régions centropariétales. La personnalité antisociale et la violation des attentes hostiles étaient reliées positivement à l'agressivité réactive. La personnalité antisociale prédisait l'agressivité réactive même à l'ajout de la violation des attentes hostile ($Z = .30$, $p = .76$) ou de la violation des attentes non hostiles ($Z = -.32$, $p = .75$) comme médiateur. En somme, le rôle médiateur du biais d'attribution d'intention n'est pas confirmé et d'autres études sont nécessaires pour mieux comprendre le lien entre la personnalité antisociale et l'agressivité réactive.

Mots-clés : trouble de la personnalité, trouble de la personnalité antisociale, comportement agressif, biais cognitif, biais d'attribution d'intention, potentiels évoqués, électroencéphalographie, N400, paradigme de la violation des attentes hostiles.

2. ABSTRACT

Every day, around the world, aggressive behaviors are committed against individuals, causing physical, psychological and financial harm. In response to provocation, these assaults are said to be reactive and can be fuelled by cognitive biases of attributing hostile intent and antisocial personality styles. Filling a gap in the scientific literature, the purpose of this study is to evaluate hostile intent bias and its role in the relationship between antisocial personality and reactive aggression. To this end, participants were asked to complete questionnaires assessing personality, cognitive processes and aggression. Then, while recording their brain activity, they were asked to read scenarios of social interactions and to attribute intent to behaviors described as ambiguous and provocative. We analyzed the N400, an event-related potential component associated with the presentation of unexpected hostile or non-hostile intentions after each scenario. Pearson correlation and multiple linear regression analyses were performed to examine the validity of our mediation model. The results show that the N400 is stronger in the presentation of unexpected non-hostile intent than in the presentation of unexpected hostile intent in the centro-parietal regions. Antisocial personality and violation of hostile expectations were positively related to reactive aggression. Antisocial personality predicted reactive aggression even with the addition of hostile expectation violation ($Z = .30$, $p = .76$) or non-hostile expectation violation ($Z = -.32$, $p = .75$) as a mediator. In sum, the mediating role of intention attribution bias is unconfirmed and further studies are needed to better understand the link between antisocial personality and reactive aggression.

Keywords: personality disorders, antisocial personality disorder, aggressive behavior, cognitive bias, interpretive bias, evoked potentials, electroencephalography, N400, hostile expectancy violation paradigm.

Table des matières

Liste des tableaux

Liste des figures

3. ABRÉVIATIONS

ANOVA : analyse de variance

ASPD/TPAS : trouble de la personnalité antisociale

BPD/TPL : trouble de la personnalité limite

BSI : Brief Symptom Inventory

HAB : biais d'attribution d'intention hostile

EEG : électroencéphalographie

GAM : modèle général de l'agression

Hma/HC : hostile concordant

Hmi/HD : hostile discordant

HN400RC : violation des attentes hostiles dans la région RC

Hz : hertz

ICA : analyse en composante indépendante

ISRS : inhibiteur de la recapture de la sérotonine

INRS : inhibiteur de la recapture de la sérotonine et de la noradrénaline

LA : antérieur gauche

LC : central gauche

LP : postérieur gauche

MA : antérieur médian

MC : central médian

MP : postérieur médian

ms : milliseconde

NHma/NHC : non-hostile concordant

NHmi/NHD : non-hostile discordant

NHN400RC : violation des attentes non-hostiles dans la région in RC

PEAG : agression proactive

REAG : agression réactive

PAI : Personality Assessment Questionnaire

ERP : potentiels reliés aux évènements

RA : antérieur droit

RC : central droit

RP : postérieur droit

RPQ : Reactive-Proactive Aggression Questionnaire

SIP-AEQ : Social Information Processing-Attribution and Emotional Response Questionnaire

SD : erreur standard

μV : microvolt

4. REMERCIEMENTS

Tout d'abord, je voudrais remercier mon directeur de recherche Dr. Jean Gagnon de m'avoir accompagnée tout au long de ce périple académique. Merci Jean d'avoir cru en mon potentiel et de m'avoir accueillie au sein de votre équipe. Merci pour votre supervision bienveillante, prévenante et flexible. Vous étiez un guide exceptionnel et resterez un modèle inspirant mes projets professionnels. Je remercie également Dr. Pierre Jolicoeur pour son aide considérable durant les analyses d'électroencéphalographie. Merci pour m'avoir transmis vos connaissances et votre expertise de recherche en neurosciences. Grâce à vous, j'ai su surmonter beaucoup d'obstacles informatiques et naviguer sur des logiciels primordiaux pour l'analyse des résultats de mon mémoire. Merci à Dr. Monique Bessette d'avoir supporté le recrutement des participants pour ce projet de recherche. J'aimerais aussi adresser un grand merci à toute l'équipe du Laboratoire d'Électrophysiologie en Neurosciences sociales pour le support technique apporté durant les phases d'expérimentation du projet. Je remercie particulièrement mes collègues Wan Seo, Gasser, Joséphine, Valérie, Rodrigo et Émilie pour les discussions affectueuses, stimulantes et édifiantes sur le plan académique. Merci pour les rires, votre compassion et vos encouragements.

Ensuite, je voudrais exprimer toute ma gratitude au corps professoral du département de psychologie pour les enseignements qui m'ont été transmis durant ma maîtrise. J'ai pu apprendre des pratiques de recherche élémentaires pour le façonnement de mon projet et la rédaction de mon mémoire. Je tiens à remercier spécialement Dr. Geneviève Mageau pour sa pédagogie et ses recommandations au niveau des analyses statistiques quantitatives. D'autre part, je suis extrêmement reconnaissante envers le personnel de soutien informatique de la bibliothèque ÉPC-Biologie. Merci pour l'inestimable support que j'ai reçu dans la résolution de mes soucis informatiques. Merci au bibliothécaire Pascal Martinolli de m'avoir parfaitement outillée sur le plan des recherches et de l'exploitation de la documentation. Vos fiches de cours sur les bases de données et les logiciels de référence m'ont été indispensables.

Enfin, je voudrais remercier ma mère Marie-Aline pour l'infaillible soutien moral qu'elle m'a apporté durant la totalité de mon cursus universitaire, et ce, malgré les milliers de kilomètres qui nous séparent. Je ne te remercierais jamais assez pour tout ce que tu m'as apporté intellectuellement et spirituellement. Merci d'avoir épongé mes sueurs et mes désespoirs. Merci pour les vidéoconférences et les appels téléphoniques requinquants aux heures tardives et aux périodes les

plus critiques de ma vie. Par tes élans de positivisme et ta force d'esprit, tu as su me relever dans l'adversité, bâtir ma résilience face aux épreuves et m'escorter jusqu'à l'obtention de mon troisième diplôme. Tu es mon héroïne. De même, je tiens à remercier mon compagnon, Thomas pour le réconfort qu'il m'a apporté depuis le début de notre vie commune. Merci de m'avoir supporté émotionnellement. Merci pour les débats de résolution de problèmes. Je te suis également reconnaissante pour m'avoir communiqué tes valeurs de combativité et de persévérance. Un dernier merci à ma cousine Soïna pour tous les beaux messages motivants qu'elle m'a envoyés depuis notre pays natal. Tu m'as aidée à garder espoir et combattre la solitude et les doutes.

5. CHAPITRE I. INTRODUCTION GÉNÉRALE

Selon l'Organisation mondiale de la santé (OMS; World Health Organization, 2017), 1,4 million de personnes dans le monde meurent chaque année des suites de violence reçue ou auto-infligée. Chez les individus ayant 15 à 44 ans, la violence est l'une des premières causes de décès recensées. 7% des femmes décédées et 14% des hommes décédés le sont à cause de violences subies (Krug, 2002). Mais l'ampleur des conséquences néfastes de la violence ne s'observe pas seulement à travers le nombre de décès. En effet, beaucoup plus de personnes sont confrontées à des blessures physiques ou mentales suite à un acte de violence ou d'agression (Krug, 2002). Selon les estimations de l'OMS, lorsqu'une personne est tuée, 20 à 40 personnes sont blessées et hospitalisées (WHO, 2017). Par ailleurs, les comportements d'agression donnent lieu à des conséquences désastreuses sur le plan économique, judiciaire et social, tant pour l'individu que pour les familles, les communautés et le pays. Chaque année, des milliards de dollars sont déboursés pour payer les frais de santé, de police et de justice. De plus, les souffrances qu'endurent les victimes à long terme peuvent conduire à des troubles mentaux (e.g. troubles dépressifs, troubles de dépendance), des tentatives de suicide, des comportements sexuels à risque, des grossesses non désirées, des maladies ou des infections sexuellement transmissibles (WHO, 2017). Au Canada, 423 767 crimes violents ont été rapportés par la police en 2018 (Moreau, 2019). Parmi ces crimes, 240 449 agressions physiques, 37 401 agressions sexuelles, 66 508 menaces, 22 450 vols et 37 218 harcèlements ou communications obscènes ont été perpétrés. Comparativement au siècle précédent, ces crimes violents recensés en 2018 sont bien plus nombreux. En effet, depuis 1962, le taux de crimes violents par habitants a quadruplé, passant de 221 à 884 incidents pour 100 000 habitants. Entre 2014 et 2018, l'indice de gravité des crimes violents aurait augmenté de 17% et le taux d'agression sexuelle aurait augmenté de 34%. Ces chiffres sont d'autant plus alarmants si nous considérons toutes les agressions non déclarées par les victimes qui sont envahies par la honte, la peur des représailles ou la peur d'être stigmatisés notamment dans le cas de violence familiale (Conroy et al., 2019; Krug, 2002). Les enfants et les personnes âgées violentés sont, par exemple, très souvent dépendants de leurs agresseurs et habitent avec eux. La situation est encore plus dramatique quand on prend en compte que la majorité des victimes de violence sont agressées par leurs proches et les membres de leur famille (Conroy et al., 2019). Chez les victimes de 17 ans et moins, 31% ont été agressés par un membre de leur famille et 32% ont été agressés par une

connaissance (e.g. ami) en 2018. Pour les victimes de 15 à 89 ans, 30% ont été violentées par leur partenaire intime, 33% ont été violentées par une connaissance et 11% ont été violentées par un autre membre de leur famille en 2018. Comprendre et prévenir les comportements d'agression au sein des familles et de l'entourage des victimes est donc impératif.

5.1. LES TYPES D'AGRESSIVITÉ ET SES MODÈLES THÉORIQUES

Dans la population générale, l'agression et la violence sont des termes employés de manière interchangeable. Cependant, dans la communauté scientifique, l'agression est un terme plus global que la violence et englobe tous les comportements d'opposition ou d'attaque à l'égard d'un individu. Les comportements agressifs peuvent comprendre des actes mineurs (e.g. pousser), des actes plus sérieux (e.g. frapper) ou des actes sévères (e.g. tuer) (Allen & Anderson, 2017). Sur ce continuum de sévérité, la violence représente la manifestation la plus extrême de l'agression et conduit à des blessures physiques graves (Anderson & Bushman, 2002; Bushman & Huesmann, 2010; Huesmann & Taylor, 2006). Selon Anderson & Bushman (2002), l'agression est définie comme un acte observable dirigé envers autrui dans le but d'affliger des dommages physiques ou psychologiques. De manière plus spécifique, l'auteur doit avoir l'intention de nuire à la victime de manière immédiate. De plus, la personne subissant l'acte d'agression doit être motivée à éviter le préjudice (Bushman & Huesmann, 2010; DeWall, Anderson, & Bushman, 2013). Il existe plusieurs taxonomies pour classifier les différents types d'agression, mais leurs applications sont encore débattues. La taxonomie de Krahé (2013) est l'une des plus récentes qui ait été développée (voir **Tableau 1**). Elle permet de décrire les comportements agressifs selon neuf modalités de réponse, ayant chacune 2 à 4 sous-types. Pour notre étude, la classification employée est celle qui décrit l'agression selon la présence ou l'absence d'une provocation (i.e. selon l'aspect réactif ou proactif de la réponse). Comme l'avait premièrement décrit Dodge (1991), l'agressivité de type réactif désigne des éclats de colère incontrôlée en réaction à une provocation tandis que l'agressivité de type proactif désigne des comportements agressifs planifiés de sang-froid dans un but instrumental. Ces concepts ont été testés sous différentes appellations dans plusieurs autres études (Merk et al., 2005; Murray-Close et al., 2010; Raine et al., 2006). Dans ce contexte, l'agression réactive porte le nom d'agression hostile, impulsive ou affective tandis que l'agression proactive est appelée agression instrumentale, préméditée ou prédatrice (Bushman & Anderson, 2001; Houston et al., 2003; Merk et al., 2005; Ramirez & Andreu, 2006).

Tableau 1. Classification des différents types d'agression selon la taxonomie de Krahé (2013).

Types	Sous-types	Exemples
Mode de réponse	Verbal	Crier ou jurer sur quelqu'un
	Physique	Frapper ou tirer sur quelqu'un
	Gestuel	Faire des gestes menaçants
	Relationnel	Donner à quelqu'un le « traitement du silence »
Instantanéité	Direct	Frapper quelqu'un au visage
	Indirect	Répandre des rumeurs sur quelqu'un derrière son dos
Qualité de la réponse	Action	Faire en sorte qu'une autre personne se livre à des actes sexuels non désirés
	Inaction	Ne pas transmettre des informations importantes à un collègue de travail
Visibilité	Visible	Humilier quelqu'un devant les autres
	Cachée	Envoyer des SMS de menace à un camarade de classe
Provocation	Proactive (/non provoquée)	Prendre le jouet d'un autre enfant.
	Réactive (/de représailles)	Crier sur quelqu'un après avoir été agressé physiquement
Direction du but	Hostile	Frapper quelqu'un par colère ou par frustration
	Instrumental	Prendre un otage pour obtenir une rançon
Blessure causée	Physique	Les os cassés
	Psychologique	Craintes et cauchemars
Durée des effets	Éphémère	Petites ecchymoses
	Durable	Incapacité à long terme de nouer des relations
Unités sociales impliquées	Individus	La violence entre partenaires intimes
	Groupes	Émeutes et guerres

À l'origine, ces comportements agressifs surviennent en raison d'une interaction complexe de facteurs individuels et sociaux (Krug, 2002). Ils ont d'ailleurs été décrits dans plusieurs modèles théoriques. Nous pouvons citer par exemple, la théorie de l'apprentissage social de l'agression de Bandura (1978, 2001), la théorie de l'acquisition des scripts agressifs de Huesmann (1988, 1998) ou encore la théorie du traitement de l'information sociale de Crick & Dodge (1996). Le modèle général de l'agression (GAM) est un modèle plus récent et qui unifie toutes ces théories (Allen & Anderson, 2017; Allen, Anderson, & Bushman, 2018; Anderson & Bushman, 2002; Anderson & Carnagey, 2004; DeWall, Anderson, & Bushman, 2011). Sur le plan social, le GAM dépeint des facteurs tels que la présence d'une provocation, d'une frustration, d'une douleur, d'un inconfort, de drogues, d'une incitation ou d'indices d'agressivité (e.g. antécédents violents, possession d'une arme ou exposition récente à de la violence). Sur le plan individuel, le GAM décrit des facteurs

biologiques, développementaux, cognitifs et affectifs tels que le sexe (être un homme), les traits de caractère (e.g. impulsivité ou promptitude à accomplir des biais de perception, d'attente ou d'attribution hostile), les croyances (e.g. croire que l'agression est normale), les attitudes (e.g. évaluer l'agression de manière positive), les valeurs, les objectifs à long terme (e.g. désirer être craint par tous les moyens), les scripts d'attente (e.g. s'attendre à ce que l'autre agisse de manière agressive), les scripts perceptuels (e.g. tendance à percevoir des situations ambiguës comme hostiles), les scripts comportementaux (e.g. croire qu'il faut répondre à la violence par la violence) ou les affects (comme la colère, la rage ou le désir de vengeance). Les causes diffèrent également selon le type réactif ou proactif de l'agression (Barratt, Stanford, Dowdy, Liebman, & Kent, 1999; Bowen, Levasseur, & Desbiens, 2014; Crick & Dodge, 1996; Ramirez & Andreu, 2006). L'agression réactive (REAG) est décrite comme une difficulté d'adaptation associée à de l'impulsivité, des affects de colère, un manque de contrôle comportemental et, des scripts d'attente et de perception hostile dans le contexte d'une provocation. L'agression proactive (PEAG), elle, est plutôt reliée à des attitudes agressives et, à un besoin de gain et de domination sociale. Au final, la REAg et la PEAG diffèrent tant au niveau conceptuel et qu'au niveau des causalités. Toutefois, elles demeurent fortement corrélées l'une à l'autre et semblables à un certain point (Poulin & Boivin, 2000; Ramirez & Andreu, 2006). Par conséquent, nous avons choisi d'étudier l'impact du biais d'attribution d'intention hostile, un script d'attente et de perception hostile, sur la REAG et ce, comparativement à la PEAG.

5.2. LE BIAIS D'ATTRIBUTION D'INTENTION HOSTILE ET SA MESURE

Le biais d'attribution d'intention hostile (HAB) désigne une tendance à interpréter les intentions des autres comme étant hostiles, bien que la situation soit ambiguë (De Castro, Veerman, Koops, Bosch, & Monshouwer, 2002). Dans ce genre de situation (e.g. une personne renverse son café sur vous), l'individu sujet au HAB va trouver le comportement inacceptable et hostile, ce qui déclenchera sa colère et son envie d'exercer des représailles (i.e. de commettre un acte agressif réactif). Cette relation positive entre la REAG et le HAB a été démontrée plus d'une centaine de fois et au sein d'échantillons cliniques ou normaux constitués d'enfants ou d'adultes d'ethnicités différentes (Bailey & Ostrov, 2008; Basquill et al., 2004; Camodeca & Goossens, 2005; De Castro et al., 2002; Dodge, 2006; Dodge et al., 2015; Gagnon & Rochat, 2017; Hubbard et al., 2001; MacBrayer et al., 2003; Matthews & Norris, 2002; Miller & Lynam, 2006). De plus, comme la

REAG est fortement corrélée à la PEAG, il est possible d'observer une relation faible entre le HAB et la PEAG (De Castro et al., 2002). Dans cette continuité, nous avons choisi de réévaluer ces relations en utilisant une méthodologie novatrice et récente.

Le HAB peut être mesuré par l'usage de questionnaires autorapportés (Coccaro et al., 2009), de vignettes écrites (K. Bowen et al., 2016; Lobbestael et al., 2013; Lösel et al., 2007), de vignettes vidéo (Coccaro, Fanning, Fisher, et al., 2017; Lansford et al., 2006), de tâche sur ordinateur (Smeijers et al., 2017) ou de mesures d'électrophysiologie (Gagnon et al., 2016). Dans l'étude de Lobbestael et al. (2013), 8 vignettes décrivant sous forme de phrases des situations ambiguës et provocantes a été employé pour mesurer le HAB. Face à ces vignettes, le participant devait décrire la situation et évaluer le caractère hostile, positif, négatif et neutre de la situation selon une échelle à 4 points, allant de 1 = plus plausible à 4 = moins plausible. Dans l'étude de Lansford et al. (2006), il s'agissait de 24 vignettes vidéo dépeignant un enfant qui tentait sans succès d'entrer dans des groupes de pairs ou était confronté à des provocations de pairs. L'enfant devait visualiser les scènes comme s'il en était le protagoniste. Par la suite, il devait décrire ce qu'il s'y était produit et expliquer pourquoi ses pairs s'étaient comportés ainsi. Les réponses d'attribution étaient alors codées hostiles ou non. Dans l'étude de Smeijers et al. (2017), le HAB a été mesuré à l'aide d'une tâche de classification sur un ordinateur. Dans cette tâche, il s'agissait de regarder des photos de visages exprimant des affects (de colère, de peur, de dégoût et de bonheur) et d'indiquer, pour chaque visage, s'il était hostile ou non, et ceci, le plus rapidement possible. Toutes ces mesures ont une faiblesse majeure qui est de ne pas pouvoir capter les inférences spontanées et non conscientes. Or le HAB est un processus cognitif rapide qui se manifeste à la deuxième étape du traitement de l'information sociale (Crick & Dodge, 1996). Les troisièmes et quatrièmes étapes sont consacrées à la clarification des buts et l'élaboration de la réponse. Au bout de la cinquième étape, l'individu a déjà eu le temps d'évaluer les résultats probables de l'interaction et les patrons de réponses auxquels il aspire. Ainsi, il est possible que le participant réponde aux vignettes ou au questionnaire selon ce qu'il pense être socialement acceptable au lieu de révéler ses premières perceptions et attentes hostiles. En fait, il a le temps d'explorer différentes interprétations avant d'émettre son jugement.

Afin de contrer cette limite méthodologique et de proposer une mesure plus objective qu'un questionnaire autorapporté, Gagnon et al. (2016) ont développé une méthodologie innovante

pouvant mesurer les processus d'intention en temps réel grâce à l'enregistrement par électroencéphalographie (EEG) des potentiels reliés aux évènements (ERP). Pendant l'enregistrement de leur activité cérébrale, les participants devaient lire des mises en situation de scène de vie quotidienne sur un écran et deviner l'intention cachée derrière les comportements mis en scène. Chacune des mises en situation comprenait 3 phrases (voir **Tableau 2**). La première phrase présentait un contexte hostile ou non hostile. La deuxième phrase décrivait des comportements ambigus et possiblement provocateurs, effectués par un ou de plusieurs personnages, envers le lecteur. La dernière phrase précisait à travers un dernier mot cible, l'intention hostile ou non hostile des personnages. Suivant ce format, les scénarios ont été créés selon quatre conditions : hostile concordante, hostile discordante, non-hostile concordante et non-hostile discordante (voir **Tableau 2**). Lorsque la condition était hostile, l'intention du personnage était hostile. À l'inverse, lorsque la condition était non hostile, l'intention était non-hostile. Quand la condition était concordante, le caractère hostile ou non hostile du contexte concordait avec le caractère hostile ou non hostile de l'intention. Enfin, quand la condition était discordante, le caractère hostile ou non hostile du contexte différait du caractère hostile ou non hostile de l'intention. Selon Gagnon et al. (2016), les scénarios non hostiles discordants suscitaient la violation des attentes hostiles du lecteur tandis que les scénarios hostiles discordants provoquaient la violation des attentes non hostiles du lecteur. En fait, la violation des attentes hostiles signifie que l'intention révélée à travers le mot cible est non hostile alors que le lecteur s'attend à une intention hostile. À l'inverse, la violation des attentes non hostiles signifie que l'intention est hostile alors que le lecteur s'attend à une intention non hostile. De plus, la présentation des scénarios de condition discordante déclenchait la composante cérébrale ERP N400, et ce, particulièrement lors de la violation des attentes hostiles. Cette composante servait ainsi de mesure objective et instantanée des inférences hostiles produites spontanément, sur les intentions motivant le comportement ambigu d'autrui.

Tableau 2. Exemple de scénarios selon les quatre conditions de la tâche de violation des attentes hostile de Gagnon et al. (2016).

Liste	Première phrase – contexte	Seconde phrase - comportement	Dernière phrase - intention	Condition
1	Vos parents n'aiment pas vous mêler à leurs disputes.			NHC

2	Vos parents sont fâchés contre vous.	Alors que vous rentrez, ils se déplacent dans une autre chambre	Vos parents veulent vous **épargner.**	NHD
1	Vous jouez au soccer contre une équipe qui a un style agressif,	Lors d'une échappée, le défenseur vous fait trébucher.	Le défenseur veut vous **blesser.**	HC
2	Vous avez une pratique de soccer avec votre équipe.			HD

NHC = non-hostile concordant ; NHD = non-hostile discordant ; HC = hostile concordant ; HD = hostile discordant. Ici, le mot cible est écrit en gras. Deux listes de 160 scénarios ont ainsi été créées.

Comme l'ont décrit beaucoup d'autres études, la N400 se traduit sur les signaux d'EEG, par une déflexion négative perçue aux alentours de 400 ms post-stimulus et une amplitude maximale aux régions centropariétales (Fitz & Chang, 2019; Frank et al., 2015; Gagnon et al., 2016; Kutas & Federmeier, 2011; Leuthold et al., 2012). Étant engendrée par un large panel de stimuli (e.g. photo de visages, mots écrits ou sons), la N400 a une fonction qui fait encore débat dans la communauté scientifique. Elle est décrite comme un recours à la mémoire sémantique (Kutas & Federmeier, 2011), une surprise ressentie face à la présentation d'un mot (Frank et al., 2015) ou une erreur de prédiction du stimulus dans un contexte d'apprentissage (Fitz & Chang, 2019). Selon Leuthold et al. (2012), la composante N400 surviendrait suite à la présentation d'un mot inattendu ou incohérent avec le contexte dans un scénario. Par exemple, dans le scénario hostile discordant du **Tableau 2**, l'intention hostile du défenseur (« Le défenseur veut vous blesser ») est en contradiction avec le contexte non hostile (« Vous avez une pratique de soccer avec votre équipe »). L'apparition du mot blesser peut paraître surprenante compte tenu du contexte. Il semble difficile de croire que le défenseur ait voulu nous blesser en nous faisant trébucher dans le cadre d'un entrainement. Au contraire, nous sommes plus portés à croire qu'il nous ait fait trébucher par inadvertance dans le feu de l'action. En fait, les informations données dans les scénarios qui ne coïncident pas avec nos scripts et connaissances générales sont difficiles à intégrer, ce qui déclenche la N400. C'est dans ce cadre théorique que Gagnon et al. (2016) ont développé leur méthode de mesure alternative du HAB.

Étant moins sujette au biais de réponse, la méthode de Gagnon et al. (2016) est celle qui sera testée dans notre étude. Il s'agira ensuite d'évaluer la relation entre la mesure neurophysiologique du HAB et l'agressivité de type réactif et proactif. Dans la littérature, des observations neurobiologiques ont déjà été faites selon le niveau et le type (réactif ou proactif) de l'agression

(Ramirez & Andreu, 2006). Chez les personnes violentes, la REAG a, par exemple, été associée à des fonctions préfrontales déficitaires (Raine et al., 1998) et des amplitudes ERPs de la P300 plus faibles (Barratt et al., 1997) que chez les personnes non violentes. En ce qui concerne la PEAG, les variables neuropsychologiques étaient les mêmes autant chez les personnes violentes que chez les personnes non violentes (Barratt et al., 1997; Raine et al., 1998). À des fins de contrôle, nous allons donc évaluer les amplitudes des ERPs, captées pour mesurer le HAB, selon le niveau de REAG et de PEAG. Selon d'autres études, la relation entre le HAB et la REAG physique aurait tendance à être plus forte chez les personnes ayant des styles de personnalité qui favorisent la dépendance, l'insécurité, la colère ou l'instabilité émotionnelle (Brent et al., 1994; Ross & Babcock, 2009; Tweed & Dutton, 1998). À ce propos, des troubles de la personnalité, tels que le trouble de la personnalité antisociale, ont fréquemment été associés à l'agressivité (Dunsieth et al., 2004; Warren et al., 2002).

5.3. LA PERSONNALITÉ ANTISOCIALE

Selon le DSM-5 (American Psychiatric Association, 2013), le trouble de la personnalité antisociale (TPAS) se caractérise par une tendance à violer et mépriser les droits d'autrui, depuis l'âge de 15 ans. Typiquement, l'individu antisocial est impulsif, éprouve une difficulté à se conformer aux normes sociales, use de tromperie pour son propre plaisir, n'éprouve aucun remords, et, se montre irritable, agressif et/ou irresponsable. Parallèlement à ces symptômes, le TPAS a été associé à des comportements agressifs, violents ou suicidaires (Brent et al., 1994; Duberstein & Conwell, 1997) et ce, quel que soit le sexe (Dunsieth et al., 2004; Warren et al., 2002). Par extension, plusieurs études récentes ont rapporté une association entre le TPAS et l'agression qu'elle soit réactive ou proactive (Lobbestael et al., 2013; Ostrov & Houston, 2008; Ross & Babcock, 2009; Tweed & Dutton, 1998; Walters, 2007). De plus, une méta-analyse réalisée par Gardner, Boccaccini, Bitting, & Edens (2015), a démontré que les caractéristiques antisociales prédisent modérément les mauvaises conduites, les comportements violents et la récidive criminelle. En Occident, 47% des hommes et 21% des femmes incarcérés pour crime violent ont reçu un diagnostic de trouble de la personnalité antisociale (Fazel & Danesh, 2002). Ces chiffres sont considérables et soulignent l'urgence de développer les connaissances actuelles sur le TPAS et ses axes de traitement. Un de ces axes de recherche consisterait par exemple à diminuer les cognitions agressives (tel que le HAB) auxquels sont sujettes les personnes antisociales. Très peu

d'études ont regardé le rôle du HAB dans la survenue de comportement agressif chez la personnalité antisociale (Lobbestael et al., 2013; Smeijers et al., 2017). Lobbestael et al. (2013) ont montré que les traits antisociaux et le HAB (mesuré à l'aide de vignettes) étaient de bons prédicteurs positifs de la REAG. Quant à Smeijers et al. (2017), ils ont démontré que les personnes ayant un TPAS réalisaient plus de HAB lors de l'interprétation d'expressions faciales. De manière à étendre nos connaissances et combler des lacunes méthodologiques, nous proposons alors d'examiner le rôle du HAB (mesuré par EEG et par questionnaire) dans la relation entre les traits de la personnalité antisociale et la REAG. L'âge, le sexe et le niveau d'éducation seront introduits comme variables contrôles dans nos analyses en raison de leurs influences sur la prévalence du TPAS et de l'agressivité (APA, 2013; Johnson, 2010; Moran, 1999). De même, nous contrôlerons les traits de la personnalité paranoïde et limite en raison de leur forte comorbidité avec le TPAS (APA, 2013).

5.4. OBJECTIFS ET HYPOTHÈSES

Le but principal de cette étude est d'examiner le rôle du HAB, mesuré par EEG, dans la relation entre les traits de la personnalité antisociale et la REAG. Pour y répondre, nous formulons plusieurs objectifs et hypothèses. **1)** Premièrement, nous voulons répliquer et valider la méthode de mesure du HAB de Gagnon et al. (2016). Tout comme observé dans l'étude de Gagnon et al. (2016), nous faisons la première hypothèse que la N400 sera plus prononcée lors de la violation des attentes hostiles dans les régions cérébrales postérieures droites. **2)** Deuxièmement, nous avons pour objectif d'évaluer le rôle prédictif des traits du TPAS et du HAB (mesuré par EEG et questionnaire) sur la survenue des comportements agressifs. **(a)** Comme l'ont déjà rapporté de nombreuses études (Lobbestael et al., 2013; Ostrov & Houston, 2008; Ross & Babcock, 2009; Tweed & Dutton, 1998; Walters, 2007), nous formulons l'hypothèse que les traits antisociaux prédiront positivement et significativement la REAG et la PEAG autorapportée. **(b)** Ensuite, puisque des études antérieures sont parvenues à prouver que le TPAS prédisait le HAB (Lobbestael et al., 2013; Smeijers et al., 2017), nous faisons l'hypothèse que les traits antisociaux prédiront positivement et significativement le HAB mesuré par EEG et questionnaire. **(c_1)** Toujours en nous appuyant sur les résultats d'autres études (Bailey & Ostrov, 2008; Basquill et al., 2004; Camodeca & Goossens, 2005; De Castro et al., 2002; Dodge, 2006; Dodge et al., 2015; Gagnon & Rochat, 2017; Hubbard et al., 2001; MacBrayer et al., 2003; Matthews & Norris, 2002; Miller & Lynam, 2006), nous

présupposons que les mesures neurophysiologiques et autorapportées du HAB prédiront de manière positive et significative la REAG. **(c_2)** Par ailleurs, comme il est parfois possible d'observer une relation faible entre le HAB et la PEAG (De Castro et al., 2002), nous supposons qu'une faible relation positive et significative existera entre la mesure du HAB (par EEG et questionnaire) et le PEAG. **(d)** Enfin, nous présumons que l'ajout du HAB (autorapporté et mesuré par EEG) agira comme médiateur et changera la qualité de relation entre les traits antisociaux et la REAG, comme l'ont précédemment supposé les auteurs Lobbestael et al. (2013) et Smeijers et al., (2017).

5.5. MISE EN CONTEXTE DE LA RÉALISATION DE L'ARTICLE

L'article scientifique est le fruit d'une collaboration de plusieurs entités. Il présente les résultats du projet de maîtrise de Adriana Ursulet (AU), précédemment décrit. Pour rendre l'article plus attrayant et ainsi augmenter ses chances de publication, des données sur les traits de la personnalité limite issue du projet de maîtrise d'Émilie de Repentigny (EdR) ont été incluses. AU a néanmoins rédigé tout l'article. AU a participé à la conception du projet, au recrutement, à la passation des tâches expérimentales, à la cueillette, à l'entrée, l'analyse, le traitement et l'interprétation des données. EdR a contribué au recrutement, la passation des tâches expérimentales et la cueillette. Monique Bessette a contribué au recrutement. Pierre Jolicoeur (PJ) est copropriétaire du laboratoire LENS au sein duquel a eu lieu l'expérimentation. PJ a également soutenu les analyses des données d'électroencéphalographie. Jean Gagnon (JG) est le propriétaire du LENS et le directeur du projet. JG a supervisé toutes les étapes de réalisation de l'étude et a révisé le mémoire et l'article.

6. CHAPITRE II. ARTICLE SCIENTIFIQUE

Title - **The mediating role of hostile attribution bias in the relationship between aggression and cluster B personality: An event-related potentials study.**

Adriana Ursulet, M.Sc., Émilie de Repentigny, M.Sc., Joyce Quansah, M.Sc., Monique Bessette, PhD, Pierre Jolicoeur, PhD, Jean Gagnon, PhD.

The aim of this study was to better understand the role of the hostile attribution bias (HAB) in the relationship between cluster B personality traits and reactive aggression. We also sought to replicate the electrophysiology method developed by Gagnon et al. (2016) to evaluate the HAB Seventy-two French-speaking adults were asked to complete online questionnaires assessing their personality traits, cognitive processes and aggressive behaviors. While brain activity was recorded, they were asked to read scenarios involving daily life interactions and to imagine why the characters (whose intentions were ambiguous) behaved in a provocative manner towards them. Following each scenario, we analyzed the N400 component of the event-related brain potential associated with the presentation of unexpected hostile or non-hostile intentions after each scenario. Results showed a stronger N400 amplitude during presentation of unexpected non-hostile intentions (hostile expectancy violations) in the centro-parietal regions. There was no mediating effect of hostile or non-hostile expectancy violation in the relationship between cluster B personality characteristics and reactive aggression. Finally, further studies are needed to better understand the cognitive processes underlying aggressive behaviors in cluster B personality disorders.

Keywords – personality disorders, antisocial personality disorder, borderline personality disorder, aggressive behavior, cognitive bias, interpretive bias, evoked potentials, electroencephalography, N400, hostile expectancy violation paradigm.

From department of Psychology, University of Montreal, Montreal, QC H3C 3J7, Canada; Centre for Interdisciplinary Research in Rehabilitation of Greater Montreal (CRIR), Montreal, QC H2H

2N8, Canada; Laboratoire d'Électrophysiologie en Neuroscience Sociale (LENS), Department of Psychology, University of Montreal, Montreal, QC H3C 3J7, Canada.
Preparation of this article was supported by a portion of Dr Gagnon's research funds.
Conflict of interest statement: none declared.
Address correspondence to Adriana Ursulet, Department of Psychology, University of Montreal, Montreal, QC H3C 3J7, Canada. Email: adriana.ursulet@umontreal.com

6.1. INTRODUCTION

Personality disorders are conditions that can have a destructive impact on an individual's quality of life and social interactions. Indeed, a person with a personality disorder will experience serious difficulties managing emotions, behaving according to culturally acceptable cognitions and interacting normally in everyday life (APA, 2013). In the United States, 9-15% of people develop a personality disorder and in most cases, this disorder is accompanied by numerous comorbid conditions (Lenzenweger et al., 2007). According to Grant et al. (2004), nearly 31 million people were affected between 2001 and 2002. The situation is even more worrisome considering that personality disorders are commonly associated with aggression (Dunsieth et al., 2004), violence (Ullrich et al., 2008), criminal behavior (Flórez et al., 2019; Ruiter & Greeven, 2000) and violent recidivism (Putkonen et al., 2003) worldwide. According to a systematic review, 65% of incarcerated men and 42% of incarcerated women have a personality disorder (Fazel & Danesh, 2002). These epidemiological data underline the importance of better understanding and treating personality disorders, particularly those in cluster B. According to the DSM-5, cluster B personality disorders are characterized by relational disorders and impulsive, emotional and/or unstable behavioral manifestations (APA, 2013). They include disorders such as antisocial personality disorder (ASPD) and borderline personality disorder (BPD) and tend to be strongly associated with a variety of maladaptive behaviors, including addictive, suicidal, or aggressive behavior (Duberstein & Conwell, 1997; Edens et al., 2007; Soloff et al., 2000).

BPD is characterized by a pervasive instability of affects, self-representations and interpersonal relationships (APA, 2013). It also includes the presence of impulsiveness, paranoia, feelings of emptiness and/or suicidal gestures. Affecting 1-2% of the general population (Lobbestael & McNally, 2016), BPD has a lifetime prevalence of approximately 6% among both sexes (Grant et

al., 2008). In a population of adolescents, BPD traits are associated with high levels of delinquency, antisocial behavior and all forms of aggression (e.g., sexual harassment, overt aggression and violence) (Chabrol et al., 2012; Penson et al., 2018). As such, the diagnosis of borderline personality is, according to some authors, a good predictor of violence and aggression (Soliman & Reza, 2001). Moreover, among the hospital settings, 65% of patients with BPD report having used physical, verbal or relational gestures that were aggressive (Zanarini et al., 2017). According to several authors, aggressive behaviors among borderline patients are guided by emotions (Haller & Kruk, 2006). In fact, BPD patients are prone to overreact, which leads to irritability, outbursts of anger, and subsequent physical aggression. Importantly, BPD is not the only personality disorder that has been shown to predict misconducts, violent behaviors and criminal recidivisms (Penson et al., 2018). According to a longitudinal study by Penson et al. (2018), ASPD is also a valid predictor.

The DSM-5 describes ASPD as pattern of violation of, and disregard for, the rights and interests of others (APA, 2013). It is expressed through a lack of social conformity, use of deception for personal gain, lack of remorse, and irresponsible, irritable or impulsive behavior. In the United States, the prevalence of ASPD is 3.63% in the general population and in the prison population, as high as 21-47% (Fazel & Danesh, 2002; Grant et al., 2004). In addition, adolescents with ASPD characteristics are likely to engage in future acts of violence and aggression (Frick et al., 2014). Moreover, in young adults, self-reports of two antisocial characteristics (i.e., sensation-seeking and egocentricity) have been associated with relational aggression (Werner & Crick, 1999). More generally, ASPD diagnosed in clinical populations has been shown to be a strong predictor of violence and aggression (Soliman & Reza, 2001). Further, high levels of aggression have been associated with ASPD regardless of gender (Dunsieth et al., 2004; Warren et al., 2002). Authors suggest that violent behavior by antisocial patients can be explained as being part of an instrumental goal, such as for the purpose of obtaining gratification (Haller & Kruk, 2006). According to the authors, antisocial individuals are said to be hypo-reactive (or hypo-aroused), which results in emotionally and physiologically disinhibiting them when performing instrumental aggressive acts.

In short, both BPD and ASPD are marked by impulsivity and aggressive behavior of all kinds (Chabrol et al., 2012; Lobbestael et al., 2015; Ostrov & Houston, 2008). As such, in the criminal population, individuals with antisocial and borderline disorders are more aggressive and impulsive than criminals without personality disorders altogether (Tiihonen et al., 1993). On the other hand, some distinctions exist in terms of the forms of aggression employed. Often, with BPD, aggression

is reactive (REAG) (or impulsive), relational and physical, while it is reactive and proactive (or instrumental) in ASPD (Gardner et al., 2012; Lobbestael et al., 2015; Newhill et al., 2009; Ross & Babcock, 2009).

Conceptually, aggression refers to intentional and observable action directed toward someone with the goal of physically or mentally harming them (Bushman & Huesmann, 2010). Aggression is said to be reactive when it occurs under provocation, threat or frustration (Gardner et al., 2012; Ramirez & Andreu, 2006). It is expressed through outbursts of uncontrolled anger and cognitive scripts involving distinct expectations and hostile perception. Conversely, aggression is said to be proactive (PEAG) when it is planned and carried out cold blood for the purpose of personal gain or social domination (Ramirez & Andreu, 2006). These two forms of aggression are highly correlated, indicating that most aggressive individuals engage in both forms (Ramirez & Andreu, 2006). Aggressive behaviors (REAG and PEAG) have disastrous economic, legal and social consequences (Conroy et al., 2019; Krug, 2002). Since the impacts are observable at the individual, family, community and national level, many programs have been developed to prevent and reduce aggression (WHO, 2017). One potential areas of intervention could consist of decreasing aggressive cognitions that cause the individual to perceive the world as a dangerous environment and to reconsider the use of aggression when a conflict occurs.

Relatedly, a meta-analysis of studies conducted with people without BPD or ASPD, showed a strong relationship between aggression and the hostile attribution bias (HAB) (Yeager et al., 2013). According to Crick & Dodge's (1996) theory of social information processing, HAB refers to a tendency to attribute hostile intentions to others despite the situation being ambiguous (De Castro, Veerman, Koops, Bosch, & Monshouwer, 2002). For example, if a person were to spill coffee on you and, even though there is no indication of such, you react by perceiving the behavior as deliberate and hostile, then you are said to be engaging in a HAB. This will likely trigger an emotional response such as anger, as well as a desire to retaliate (e.g. REAG). In more than 100 studies, the positive relationship between REAG and HAB has been demonstrated in clinical and normal samples of individuals of different ages and ethnicities (Bailey & Ostrov, 2008; Camodeca & Goossens, 2005; De Castro et al., 2002; Dodge et al., 2015; Gagnon & Rochat, 2017; Miller & Lynam, 2006).

In an ambiguous and provocative situation, people with BPD tend to interpret events (such as abuse or rejection) as threatening (Lobbestael & McNally, 2016). This leads them to be overly

sensitive to rejection to behave impulsively, and to feel negative emotions. According to several interpretations, dysregulation of affect and behavior, which is characteristic of BPD, is associated with various cognitive biases such as the HAB (Baer et al., 2012). In fact, studies by Barnow et al. (2009) and Arntz et al. (2011) found that people with BPD showed a readiness to perceive a person as negative, aggressive, malicious, abusive, and rejecting. Thus, it is quite possible that the HAB can explains why people with BPD act aggressively towards others. Smeijers et al. (2017) have shown that patients with BPD often produce a lot of hostile interpretation biases.

With regard to ASPD, few studies have tested the HAB as an explanatory variable for REAG (Lobbesteal, Cima et Arntz, 2013 ; Smeijers, Rinck, Bulten, Van den Heuvel et Verkes, 2017). According to Lobbestael et al. (2013), ASPD traits and HAB (measured using thumbnails and images) were good predictors of REAG. Further Smeijers et al. (2017), found that people with ASPD performed many HABs when looking at facial expressions.

The HAB can be measured using self-reports (Coccaro et al., 2009), written vignettes (Bowen et al., 2016; Lobbestael et al., 2013; Lösel et al., 2007), video vignettes (Coccaro, Fanning, Fisher, et al., 2017; Lansford et al., 2006), computer tasks (Smeijers et al., 2017) or electrophysiology methods (Gagnon et al., 2016). In the study by Lobbestael et al. (2013), HAB was measured using 8 images from the thematic apperception test and 8 text vignettes describing ambiguous and provocative scenes from daily life. Participants were asked to describe the scenes and rate the hostile, positive, negative and neutral character of each scene on a 4-point scale, ranging from most plausible to least plausible. In the Smeijers et al. (2017) study, HAB was measured using an image classification task. Patients were asked to look at pictures of affective facial expressions affect (anger, fear, disgust and happiness) and to indicate, as quickly as possible, whether each face was, or was not, hostile. Finally, the Lansford et al. (2006) study used 24 video vignettes depicting a child rejected or provoked by a peer group were used. Participants had to visualize scenes depicting rejection and provocation as though they themselves were the victim. They were then asked to describe and explain why the peers behaved as they did. Assignment responses were coded as either hostile or non-hostile. While all of the previously mentioned HAB measures provide interesting results, but they are not without flaws. Indeed, these methods do not allow for the measurement of spontaneous inferences and rapid intention attribution processes that are characteristic of the HAB. The latter occurs in the early stages of social information processing (Crick & Dodge, 1996).

Before providing their responses, participants have time to consider other, more socially acceptable interpretations.

In order to capture the first cognitive processes of real-time intention attribution, Gagnon et al. (2016) developed an innovative measurement method based on the recording of brain signals. The aim was to present different scenarios on a screen that, in written form, describe a character performing ambiguous behavior towards the reader in a context-specific manner (see **Table 1**). The context was either hostile or non-hostile and the reader was asked to attribute an intention to the character. Subsequently, the character's actual intention was revealed through a final target word and event-related potentials (ERPs) were recorded. The intention could be either hostile or non-hostile. In principle, when the hostile or non-hostile nature of the intention was at odds with the hostile or non-hostile nature of the context, expectations about the intention of the character being portrayed were violated. Conversely, when the hostile or non-hostile nature of the intention was consistent with the hostile or non-hostile nature of the context, expectations were not violated. According to Gagnon et al. (2016), the ERP component N400 was observable when expectations were violated. In the literature, N400 is described as a negative deflection occurring around 200 to 500 ms post stimulus presentation (Fitz & Chang, 2019; Frank et al., 2015; Gagnon et al., 2016; Kutas & Federmeier, 2011; Leuthold et al., 2012). Its amplitude is maximal in the centro-parietal regions of the brain and is triggered when the word presented is unexpected or inconsistent with the context in the scenario (Leuthold et al., 2012). In the study by Gagnon et al. (2016), the N400 directly measured expectation violation and its amplitude was stronger during the hostile expectations violation than during the non-hostile expectations violation.

The main goal of this study is to examine the mediating role of the HAB (measured by EEG and self-report) in the relationship between cluster B personality traits and REAG. To achieve this, we present several objectives and hypotheses. **1)** First, we want to replicate and validate the HAB measurement method developed by Gagnon et al. (2016). Our first hypothesis is that N400 will be more pronounced in the right posterior brain regions during the hostile expectations violation. **2)** Secondly, we aim to evaluate the predictive role of ASPD traits, BPD traits and HAB (as measured by EEG and self-report) on the occurrence of aggressive behaviors. **(a_1)** We hypothesize that ASPD traits will positively and significantly predict self-reported REAG. **(a_2)** We also hypothesize that BPD traits will predict REAG. **(b)** We hypothesize that ASPD and BPD traits will significantly predict self-reported HAB and hostile expectations violation. **(c)** We expect neurophysiological

and self-reported measure of HAB to significantly predict REAG. **(d)** Finally, we expect that self-reported and neurophysiological HAB will mediate the relationship between cluster B personality traits and REAG.

6.2. METHODS

PARTICIPANTS

Seventy-two French-speaking adults were recruited from university classes in two metropolitan universities, a list of former patients who consulted in a personality disorders clinic, and the general population through posters and announcements on Facebook and Kijiji. Interested individuals were then contacted by email to receive information about the study and to make an appointment for a laboratory visit. All participants were between 18 and 65 years of age, had normal vision with or without correction and no history of psychosis, neurological disorder or severe brain damage. Seventeen of them had been taking a central nervous system medication for at least 2 weeks prior to the day of the experiment. Of these, two were taking an anxiolytic, two a controlled-release methylphenidate stimulant, seven an SSRI (selective serotonin reuptake inhibitor) or SNRI (serotonin-norepinephrine reuptake inhibitor) antidepressant, of whom one in combination with a noradrenergic and specific serotonergic antidepressant and three with a thyroid regulator and/or an anxiolytic, and two used a stimulant combined with anxiolytics. Before the visit, participants were asked not to use other drugs for at least 1 week prior to the experiment. They were also advised not to fast, to have a normal night's sleep, to consume sugar and coffee in the same quantity as usual and not to drink alcohol during the last 24 hours before the meeting. Failure to comply with any of these instructions resulted in the postponement of the appointment. All participants received a financial compensation of $25 at the end of the appointment. Nine participants were excluded due to attrition, a significant amount of missing data, a mother tongue other than French, or excessive artifacts on the EEG signals caused by eye movements. The final sample consisted of 63 participants (46 females and 17 males) with an average age of 29 (SD = 1.44) and 15 years of education (SD = .40).

MEASURE

Before coming to the laboratory, participants received a consent form and a link by email, inviting them to answer 5 online questionnaires.

The first questionnaire was drawn from the French adaptation of the *Personality Assessment Questionnaire* (Morey, 1991, 2014). Only scales assessing borderline and antisocial personality traits were included. Each subscale consisted of 24 items. The *Antisocial Characteristics Scale* (ASPD features) consisted of 8 items measuring antisocial behavior, 8 items measuring egocentricity and 8 items measuring stimulus seeking. The *Borderline Characteristics Scale* (BPD features) consisted of 6 items assessing affective instability, 6 items assessing identity problems, 6 items assessing negative relationships, and 6 items assessing self-harm. Each item was scored on 4 Likert-type points, ranging from 0 = False, not at all true to 3 = Very true. For each subscale, the scores on the 24 items were added together to form a total score for antisocial traits and a total score for borderline traits. Higher scores reflected greater degree of personality traits. According to Morey, (2014), several studies have evaluated the reliability and validity of the PAI subscales in normal, clinical, and student populations (Karlin et al., 2005; Penson et al., 2018; Slavin-Mulford et al., 2012). Generally speaking, each of the clinical subscales had acceptable internal consistency ($\alpha \approx .70$), adequate test-retest reliability over a 4-week interval ($r = .86$, for the full subscales), and strong convergent validity (Morey, 2014). The ASPD features scale was highly correlated with the *Self-Report Psychopathy test* ($r = .54$ to $.80$) and the *Multiphasic Personality Inventory Antisocial Personality Disorder scale* ($r = .60$ to $.77$). The BPD features scale was highly correlated with the *Revised NEO Personality Inventory Neuroticism trait* ($r = .67$) and the *Minnesota Multiphasic Personality Inventory - Borderline Personality Disorder scale* ($r = .77$). In our study, the internal consistency was adequate for the ASPD features scale ($\alpha = .91$) and the BPD features scale ($\alpha = .90$).

Two-scales from the *Brief Symptom Inventory* (BSI; Derogatis, 1993) translated to French by Gosselin & Bergeron (1993), were used to measure level of depression and paranoia and used as control variables. The depression scale consisted of 6 items while the paranoid ideation scale consisted of 5 items. Each item was answered using a 5-point Likert-type scale, ranging from 0 = not at all to 4 = extremely. For each scale, a total score was calculated by adding up all item scores. The higher the score, the greater the level of traits. According to Derogatis (1993), the BSI dimensions had acceptable to moderate internal consistency ($\alpha = .71$ to $.85$), acceptable to adequate test-retest reliability ($r = .68$ to $.91$), and high convergent validity with the *Symptom Checklist-90-Revised* ($r = .92$ to $.99$). These results have been corroborated by several recent studies in normal

and clinical population samples (Adawi et al., 2019; Khalil et al., 2011). In our study, the Cronbach's alpha was moderate for the depression scale (α = .84) and paranoid ideation (α = .81).

A version of the *Reactive-Proactive Aggression Questionnaire* (RPQ; Raine et al., 2006), translated into French by Gagnon & Rochat (2017), was used to assess aggression behaviors. The questionnaire included an 11-item scale measuring REAG (e.g., getting angry at the provocation of others) and a 13-item scale measuring PEAG (e.g., hurting others to win a game). Each item was rated on a 3-point Likert-type scale, ranging from 0 = never to 2 = often. The REAG scores were calculated by adding the item scores. Higher scores indicated greater aggressive behaviors. Reliability and validity were tested in multiple samples of incarcerated and nonclinical individuals ages 6 to 64 years old (Cima et al., 2013; Fossati et al., 2009; Raine et al., 2006). According to Cima et al. (2013), the RPQ demonstrated moderate internal consistency for the reactive dimension (α = .83). The reactive scale was moderately to highly correlated with the *Aggression Questionnaire* (r = .26), the *Barratt Impulsiveness Scale-11* (r = .53), the *Psychopathic Personality Inventory* (r = .35), and the *Hostile-Aggression Scale* (r = .38 to .51) (Cima et al., 2013; Raine et al., 2006). In our sample, the internal consistency was moderate for reactive aggression scale (α = .82).

A French adaptation of the *Social Information Processing - Attribution and Emotional Response Questionnaire* (SIP-AEQ; Coccaro, Noblett, & McCloskey, 2009), translated by Gagnon, McDuff, Daelman, & Fournier (2015), was administered to measure the hostile attribution bias (HAB). The SIP-AEQ included 8 vignettes depicting scenes of everyday life where a character acts provocatively and has ambiguous intentions. For each vignette, participants were asked to rate the likelihood that the character's intention was directly hostile, indirectly hostile, neutral or instrumental (4 items per vignette). Each item was rated on a 4-point Likert-type scale, ranging from 0 = not at all likely to 3 = very likely. Hostile attribution biases (HAB - direct; HAB - indirect) were calculated by averaging the responses to the 8 vignettes for each intention type. The HAB score was determined by adding HAB - direct and HAB - indirect. The higher the score, the higher the HAB. According to Coccaro et al. (2009), internal consistency was moderate for HAB (α = .82). The SIP-AEQ had good convergent validity. The HAB was moderately correlated with the *Hostile Automatic Thought Questionnaire* (r = .27), the *Trait meta-mood scale* (r = -.23), and the variables assessing physical and verbal aggression in the *Buss-Perry Aggression Questionnaire* (r = .35). Similar psychometric values have been reported by several other studies (Chen, Coccaro,

& Jacobson, 2012; Coccaro, Fanning, & Lee, 2017) including a study with a French version (Gagnon et al., 2015). In our sample, the Cronbach's alpha was adequate for HAB ($\alpha = .93$).
The last questionnaire administered assessed age, gender, mother tongue and education status.

STIMULI

The stimuli constituted 320 scenarios depicting social interactions encountered in everyday life and was developed by Gagnon et al. (2016) to test hostile and non-hostile expectancy violations. Each scenario consisted of three sentences (see **Table 1**). The first sentence described a typically hostile or non-hostile context. The second sentence depicted a character whose intention was ambiguous, thus committing a potentially provocative behavior to the reader. The last sentence included a final target word that resolved the ambiguity by clarifying the intention behind the behavior. The scenarios were created under four conditions: hostile match (Hma), hostile mismatch (Hmi), non-hostile match (NHma) and non-hostile mismatch (NHmi). When the conditions were hostile, the target word indicated hostile intent on the part of the character's behavior. Conversely, when the conditions were non-hostile, the intention was depicted as non-hostile. Conditions were said to be a match when the hostile or non-hostile nature of the intention was consistent with the hostile or non-hostile nature of the context. Similarly, conditions were said to be mismatched when the hostile or non-hostile nature of the intention differed from the hostile and non-hostile nature of the context. Two lists of 160 scenarios (i.e., 2 X 40 scenarios for each of the four conditions) were used to balance the match and the mismatch conditions with the hostile and the non-hostile conditions across participants. The two lists shared the same behaviors and intentions but differed in the hostile or non-hostile nature of the context. The first two sentences were composed of a maximum of 25 words and the last sentence a maximum of 8 words. The third sentence was phrased negatively in almost 50% of the scenarios for each condition. The 2 lists were administered alternately and equally across participants. A list of 20 additional scenarios (i.e., 5 X 4 scenarios for each of the 4 conditions) was developed for the purpose of practice and comprehension trials.

Table 1. ***Examples of scenarios under the four conditions of the hostile expectancy violation paradigm.***

List	First sentence – context	Second sentence - behavior	Third sentence - intention	Condition
1	You're playing soccer against a team that has an aggressive style.	On a breakaway, the defender trips you up.	The defender wants to **hurt** you. [a]	Hma
2	You have a soccer practice with your team.			Hmi

1	You're having dinner with friends and Sylvie, who's obnoxious.	She doesn't mention that your shirt is stained.	Sylvie doesn't want to **embarrass** you. [b]	NHmi
2	You're having dinner with friends and Sylvie who's nice.			NHma

NHma = non-hostile match; NHmi = non-hostile mismatch; Hma = hostile match; Hmi= hostile mismatch. Here, the target word is in bold. Translation in English of [a] "Le défenseur veut vous **blesser**", [c] "Sylvie ne veut pas vous **embarrasser**".

PROCEDURE

After completing the online questionnaire and giving their written consent, participants were invited to the laboratory to perform the experimental task. While their brain activity was recorded, they were asked to read the daily life interaction scenarios and visualize them as though they were actually experiencing them. As they read the first two sentences, the reader had to imagine why the characters were behaving in such a way towards them (intention attribution process). Once ready, they could initiate the presentation of the third sentence. For each scenario, a trial consisted of presenting the first two sentences for at least 1500 ms. After pressing the space bar on the keyboard, a delay of 500 ms without stimuli was followed by a fixation cross appearing in the center of the screen for 1000 ms. A third sentence was then displayed, word by word, in the center of the screen and ended with the target word. Each word was presented for 300 ms, with a delay of 200 ms between words. Finally, a fixation cross was displayed in the center of the screen for 2000 ms. The participant had to keep his eyes focused on the center of the screen and refrain from blinking from the appearance of the first cross until the disappearance of the second cross. Each scenario presented in this way was a trial. In total, there were 4 practice trials followed by 10 blocks of 17 experimental trials (170 trials). Of these 17 trials, 16 were used to assess attributions of intention and 1 trial was used as a comprehension test. The 16 trials testing intention attribution included 4 scenarios for each of the 4 conditions (Hma, Hmi, NHma and NHmi). They were designed from one of the 2 lists of 160 stimuli. The comprehension trial was followed by a question. The purpose of this question was to ensure that the participant was reading and understanding the scenarios. The true or false question pertained to a detail in the previous sentence. The participant could answer by pressing the letter N (true) or M (false). A correct/incorrect answer was followed by feedback (green or red cross respectively). For our sample, the average rate of correct answers was 91.1%, indicating a high rate of comprehension. The experimental trials were presented in random order and without repetition. The blocks were separated by a break, the duration of which was determined by the participant. The scenarios used for the practice and comprehension trials were taken from a

list of 20 scenarios that differed from the lists of experimental scenarios. The words and fixation crosses were written in white, Helvetica font, size 14, bold, on a 17-inch (43.18 cm) black screen. The distance between the screen and the participant's eye was 70 cm. Three characters corresponded to approximately 1° of visual angle. The experimental task was created using E-Prime 2.0 software (*E-Prime*, Psychology Software Tools, Pittsburgh, PA).

ELECTROPHYSIOLOGICAL METHODS

The electroencephalography took place in a Faraday cage and under medium brightness. The brain activity of the participants was captured using 64 Ag/AgCI electrodes in an elastic cap. The position of the electrodes was done according to the International 10-10 System (Sharbrough et al., 1991). The right and left mastoids were used as references. One electrode was placed below the left eye to capture blinking and vertical eye movements. Two other electrodes were placed at the outer canthi of the eyes to capture horizontal eye movements. The signals were processed and recorded via a Biosemi ActiveTwo amplifier system (Amsterdam, Netherlands) at a sampling frequency of 512 Hz. Online, a 0.16 Hz high-pass filter and a 100 Hz low-pass filter were applied to the EEG signals. On Matlab, a 0.1 Hz high-pass filter and a 30 Hz low-pass filter were applied during offline analyses. The resulting signals were segmented in trials according to a time window of from 200 ms before, to 800 ms after the target word onset. The baseline time window ranged from -200 ms to 0 ms. Trials containing too many artifacts (i.e. eye or muscle movements) were rejected using an independent component analysis (Drisdelle et al., 2017). Rejection thresholds were applied for blink (i.e > 80 μV within a time window of 150 ms) and for eye movement (i.e > 35 μV within a time window of 300 ms). Electrodes with a noisy EEG signal (i.e. exceeding +/- 100 μV voltage) were interpolated by spherical spline. When more than 7 electrodes were noisy in a trial, the trial was rejected. When the number of rejected trials was greater than 20 per condition, the participant was excluded from the sample. In our final sample, the percentage of rejected trials was less than 17,5% in the four conditions (i.e. 0 to 17,5% for Hma, 0 to 15% for Hmi, 0 to 12,5% for NHma and 0 to 12,5% for NHmi). The trials were then averaged by condition (Hma, Hmi, NHma, NHmi) and for each participant. On average, there were 39 trials per condition. The ERP amplitudes captured by the electrodes were averaged over 6 lateral regions and 3 midline regions on the scalp. The lateral electrodes were separated as follows: anterior left (AF3, AF7, F1, F3, F5, F7, FT7, FC1, FC3, FC5), central left (TP7, T7, C1, C3, C5, CP1, CP3, CP5), posterior left (P1,

P3, P5, P7, PO3, PO7, O1), anterior right (AF4, AF8, F2, F4, F6, F8, FT8, FC2, FC4, FC6), central right (TP8, T8, C2, C4, C6, CP2, CP4, CP6) and posterior right (P2, P4, P6, P8, PO4, PO8, O2). The midline electrodes were analyzed as follows: anterior median (AFZ, FZ, FCZ), central median (CZ, CPZ), posterior median (PZ, POZ, OZ).

STATISTICAL ANALYSES

Statistical analyses were performed to evaluate the voltage of the ERP amplitudes (dependent variable) according to the conditions (Hma, Hmi, NHma, NHmi) and location of sensors on the scalp. Each subject being its own control, two repeated measures ANOVAs with Huynh-Feldt corrections were performed. The first ANOVA was for the lateral electrodes. The independent variables were Intention (hostile, non-hostile), Consistency (match, mismatch), Hemisphere (Left, right) and Location (Anterior, Central, Posterior). Mean ERP amplitudes observed at midline regions were analyzed in a second ANOVA. The independent variables were Intention, Consistency and Location. Given that our first objective was to demonstrate the presence of an N400 during expectancy violations (mismatch - match conditions), interaction effects involving the Consistency factor were looked at in the ANOVAs. In order to assess the role of the N400 in our mediation models, we selected regions showing greater negative amplitude (as shown in Gagnon et al., 2016; Leuthold et al., 2012). Pearson's correlations were performed between all variables. Therefore, several multiple linear regressions were conducted to assess whether antisocial characteristics, borderline characteristics, and the HAB (as measured by self-report or EEG) predicted scores on reactive and proactive aggression. Age, sex, education, paranoid ideation and depression traits served as covariates.

6.3. RESULTS

N400

Figure 1 shows mean ERP amplitudes at the 9 scalp regions, post target word presentations, for each of the four conditions (Hma, Hmi, NHma, NHmi). The ERPs consisted of two negative deflections at approximately 100 ms and 350 ms and two positive deflections emerging at around 200 ms and 600 ms. **Figure 2** shows differences in mean ERP amplitudes between mismatch and match conditions for the 9 scalp regions. Mean amplitude differences indicate a negative deflection

(N400) at around 350 to 650 ms during hostile expectancy violations (NHmi-NHma conditions). Based on visual inspection, greatest deflections were at midline and right sites in central and posterior regions. When non-hostile expectancies were violated (Hmi-Hma conditions), mean perceived amplitudes neared zero in the central and posterior regions. **Figure 3** shows topography of mean amplitude differences observed on the scalp from 350 to 650 ms (post target onset) during hostile and non-hostile expectancy violations. During the hostile expectancy violations, the N400 seems to appear in the central and posterior regions of the right hemisphere and the midline sites.

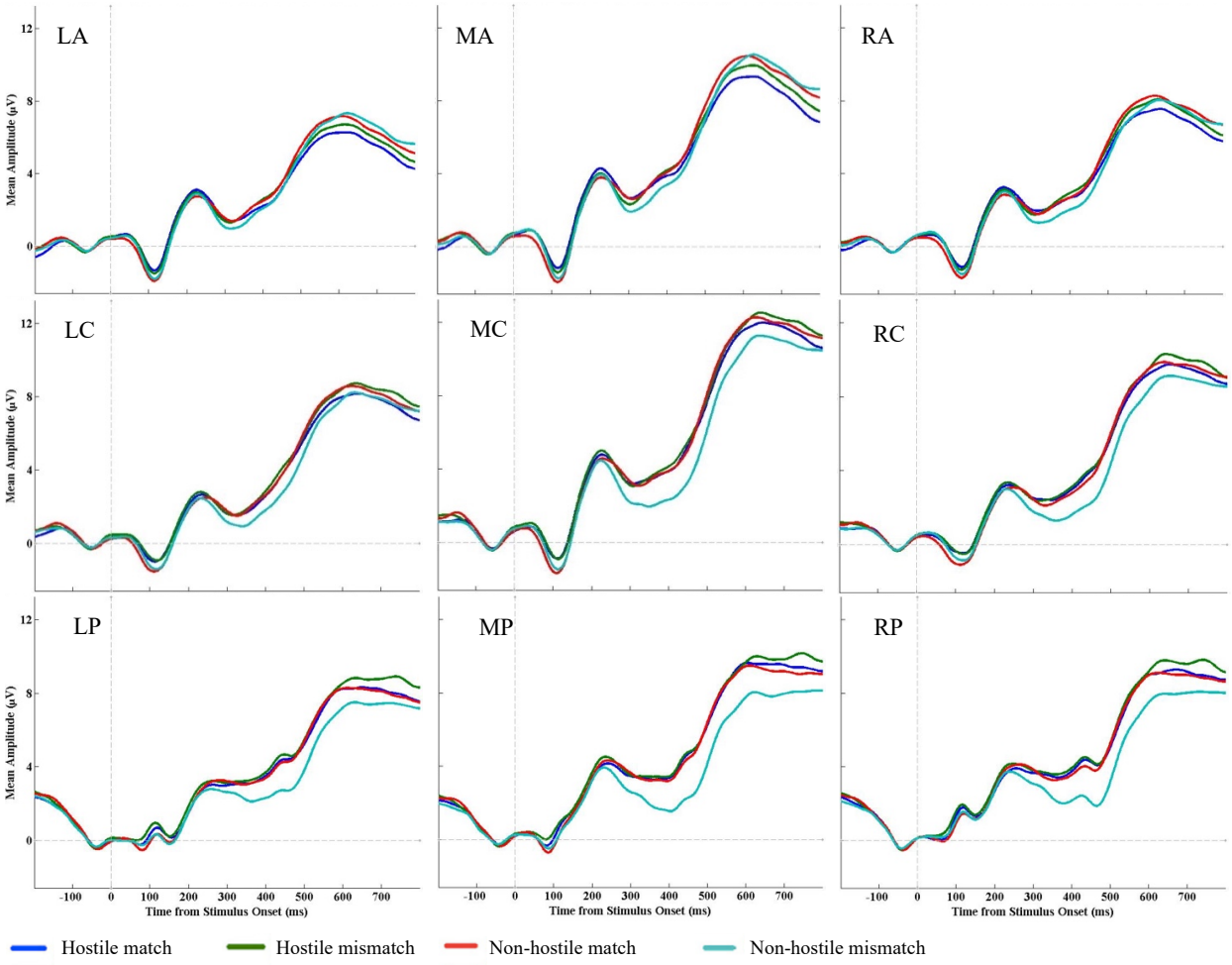

Figure 1. **Grand average ERPs recorded when presenting hostile match, hostile mismatch, non-hostile match and non-hostile mismatch target words for 9 brain regions.**

LA = anterior left; LC = central left; LP = posterior left; MA = anterior median; MC = central median; MP = posterior median; RA = anterior right; RC = central right; RP = posterior right.

For the lateral electrodes ANOVA, there was an interaction effect between Intention, Consistency and Location (F (2.124) = 5.90; p = .01), and between Intention, Consistency and

Hemisphere (F (1,62) = 5.21; p = .03). For these interactions, effect sizes were moderate (partial R^2 = .08; partial R^2 = .09 respectively). Simple effects for these last two interactions were assessed for Consistency factor by paired comparisons with post hoc Bonferroni adjustment. The levels of the Consistency factor (match and mismatch) differed significantly for the non-hostile Intention in the central and posterior regions, with an adjusted alpha of .004. Voltages were significantly more negative in the non-hostile mismatch condition at central (M = 4.83) and posterior regions (M = 4.39) than in the non-hostile match condition (M = 5.89, M = 5.82 respectively) (see Figure 1). There was no difference between hostile mismatch and hostile match at anterior, central and posterior sites. Thus, at the lateral electrodes, the negative deflection N400 was greater in the non-hostile mismatch conditions at central and posterior regions than in the non-hostile match conditions. Additionally, the levels of Consistency differed significantly for non-hostile Intention regardless of the Hemisphere with an adjusted alpha of .006. Voltages were significantly more negative in the non-hostile mismatch condition in the left (M = 4.69) and right (M = 4.66) hemispheres than in the non-hostile match condition (M = 5.47, M = 5.84 respectively) (see **Figure 1**). For hostile Intention, there was no difference between mismatch and match in the left and right hemispheres at anterior, central and posterior regions. There was also an interaction effect between Intention and Consistency (F (1,62) = 14.50; p = .00), between Consistency and Hemisphere (F (1,62) = 7.39; p = .01) and between Consistency and Location (F (1.124) = 8.73; p = .00). The magnitude of these interactions was high (partial R^2 = .19; partial R^2 = .11; partial R^2 = .12 respectively).

For the midline regions ANOVA, there was an interaction effect between factors of Intention and Consistency (F (1,62) = 16.16; p = .00), between factors of Consistency and Location (F (2,124) = 10.59; p = .00) and between factors of Intention, Consistency and Location (F (2,124) = 5.60; p = .01). Effect sizes for these interactions were high to moderate (partial R^2 = .21; partial R^2 = .15; partial R^2 = .08). Simple effects for the last interaction was evaluated for the Consistency factor by paired comparisons with post hoc Bonferroni adjustment. Level of Consistency factor differed significantly for non-hostile Intention at central and posterior regions on the scalp with an adjusted alpha of .004. Voltages were significantly more negative in the non-hostile mismatch condition at central (M = 6.37) and posterior regions (M = 4.52) than in the non-hostile match condition (M = 7.96, M = 6.37 respectively) (see **Figure 1**). There was no difference between mismatch and match for hostile Intention at anterior, central and posterior regions. Then, at midline

electrodes, the N400 was greater in the non-hostile mismatch conditions at central and posterior regions than in the non-hostile match conditions.

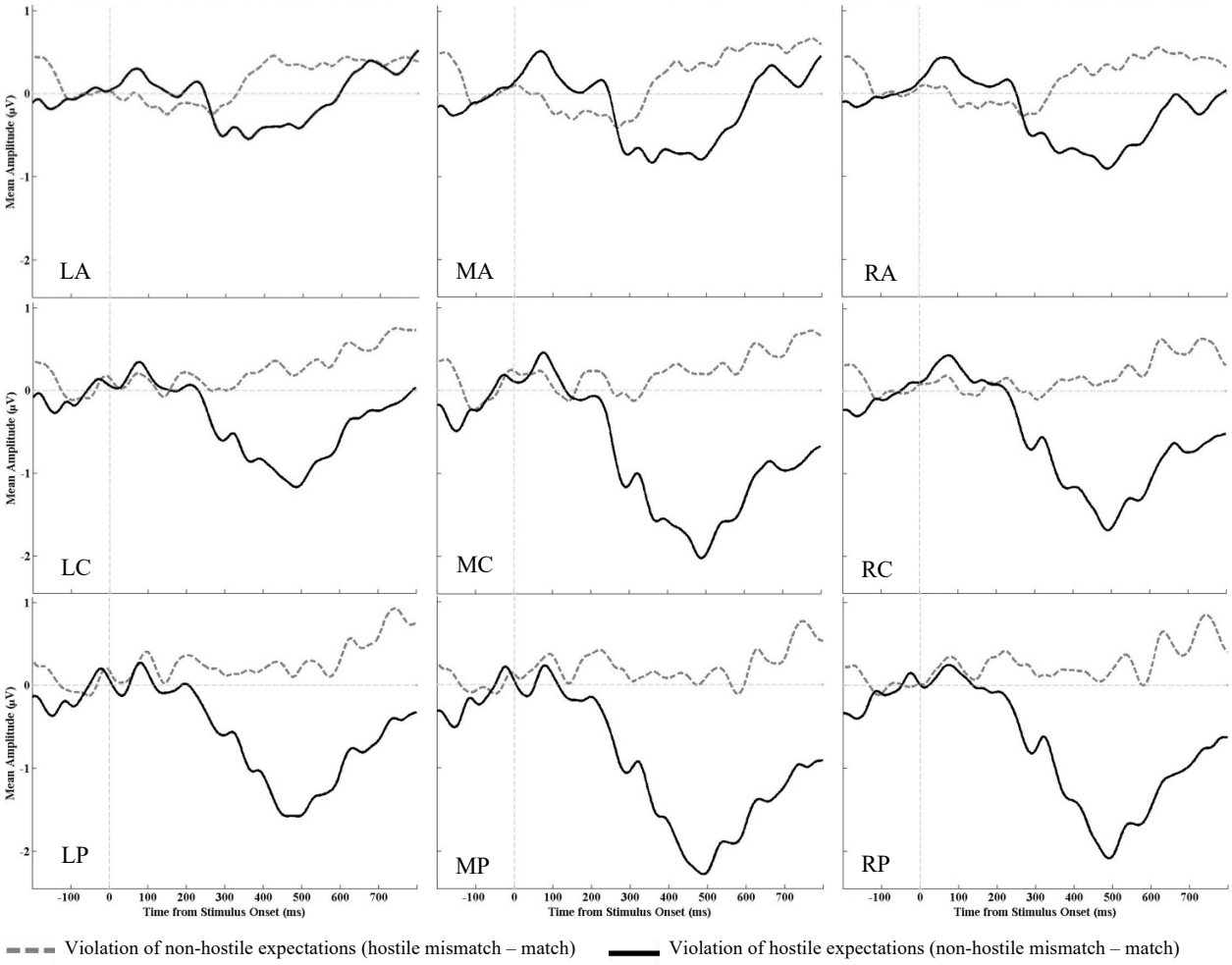

Figure 2. **Difference between the mismatch and match conditions of the grand ERP averages obtained after presentation of the hostile or non-hostile target word for 9 brain regions.**

LA = anterior left; LC = central left; LP = posterior left; MA = anterior median; MC = central median; MP = posterior median; RA = anterior right; RC = central right; RP = posterior right.

These results confirm the presence of the N400 at central and posterior regions in the non-hostile condition (i.e. when hostile expectations were violated). In the hostile condition (i.e. during non-hostile expectancy violations), the N400 was not significantly visible.

Since the ERP waveform differences and the topographic map indicated a stronger N400 effect in the central and posterior regions of the right and midline sites (**Figure 2 and 3**), we selected MC, RC, MP and RP regions for further analysis.

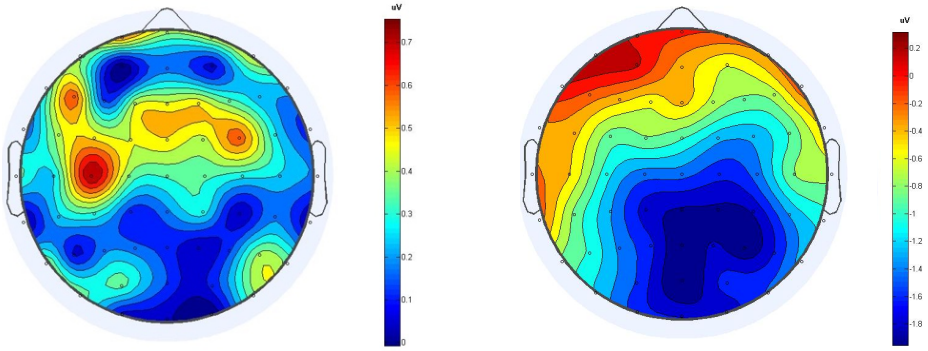

Figure 3. **The topographic map of ERP mean differences between mismatch and match conditions from 350 to 650 ms after presentation of hostile or non-hostile target words.**

On the left, non-hostile expectancy violation (hostile mismatch - match). On the right, hostile expectancy violation (non-hostile mismatch - match).

PREDICTION OF REACTIVE AGGRESSION

The distributions of self-report scores were inspected, and external validity was not compromised. The scores had good variability around the mean and extreme data were in continuity with the data set. The variables had less than 5% missing data and met the criteria for normality. However, the variables measuring proactive aggression and antisocial behavior had positive skewed distributions, which reflects the distribution observed in the general population.

Significant Pearson's correlations were found between different variables in the study. Correlation coefficients of the variables of interest are presented in **Table 2**. REAG was correlated with age, depression, antisocial traits, borderline traits and non-hostile expectancy violations (hostile condition) in the MC ($r = -.29$, $p \leq .05$ two-tailed), MP ($r = -.25$, $p \leq .05$ two-tailed), RC ($r = -.35$, $p \leq .01$ two-tailed) and RP ($r = -.32$, $p \leq .01$ two-tailed) region. In addition, antisocial traits were correlated with gender, paranoid ideation and borderline traits and indirect hostile attribution bias ($r = .30$, $p \leq .05$ two-tailed). Borderline traits were correlated with paranoid ideation, depression and self-reported HAB. REAG was not correlated with hostile expectancy violations (non-hostile condition) in MP ($r = .09$, $p \geq .05$ two-tailed), MC ($r = .11$, $p \geq .05$ two-tailed), RC ($r = .20$, $p \geq .05$ two-tailed) and RP ($r = .12$, $p \geq .05$ two-tailed).

Because the correlation between hostile or non-hostile expectancy violations and aggression scores (REAG) was more strongly consistent in the RC region than in the MC, MP and RP regions, regression analyses were performed with the RC region.

Table 2. ***Correlation matrix.***

Variables	1	2	3	4	5	6	7	8	9	10	11	12
1. Age	-											
2. Gender	**.27***	-										
3. Education	-.11	-.12	-									
4. Paranoid ideation	.13	.01	.09	-								
5. Depression	.07	.06	.08	**.49****	-							
6. ASPD features	.08	**.39****	.04	**.26***	.19	-						
7. BPD features	.05	.03	.21	**.58****	**.68****	**.48****	-					
8. REAG	**.26***	.23	.06	.24	**.41****	**.47****	**.52****	-				
9. HAB	.13	.11	-.11	**.58****	**.37****	.25	**.34****	.14	.16	-		
10. HN400RC	-.01	-.08	.03	.00	.10	-.10	-.11	**-.35****	**-.27***	.18	-	
11. NHN400RC	.14	**.26***	-.09	-.13	-.01	-.00	-.21	.19	.01	-.08	-.03	-

* $p \leq .05$; ** $p \leq .01$

Following the procedures of Hayes (2018), a first regression was conducted with ASPD features as the independent variable. Hostile expectancy violation (non-hostile condition) and self-reported HAB were the mediator variables and BPD features was introduced as a covariate. Antisocial characteristics did not predict hostile expectancy violation and self-reported HAB (see **Figure 4**). REAG was significantly predicted by antisocial characteristics ($R^2 = .41$) and hostile expectancy violation. Self-reported HAB did not predict REAG. With the addition of the mediators, antisocial characteristics predicted REAG ($R^2 = .46$). The total effect was not greater than the direct effect and the indirect effect was not significant for the hostile expectancy violation (indirect = .00, *SE* = .01, 95% CI [-.02; .04]) and self-reported HAB (indirect = -.00, *SE* = .01, 95% CI [-.02; .01]). Similarly, the Sobel Aroian test indicated that the indirect effect was not significant for hostile expectancy violation ($Z = .30$, $p = .76$) and self-reported HAB ($Z = -.37$, $p = .71$). The same regression was assessed with non-hostile expectancy violation (hostile condition) and self-reported HAB as mediator variables. Antisocial characteristics did not predict non-hostile expectancy violation. Non-hostile expectancy violation significantly predicted REAG. With the addition of the mediators, antisocial characteristics predicted REAG ($R^2 = .50$). The total effect was not greater than the direct effect and the indirect effect was not significant for hostile expectancy violation (indirect = -.00, *SE* = .02, 95% CI [-.04; .04]) and self-reported HAB (indirect = .00, *SE* = .01, 95% CI [-.02; .02]). The Sobel Aroian test indicated that the mediating effect for non-hostile expectancy violation ($Z = -.32$, $p = .75$) and self-reported HAB ($Z = -.05$, $p = .96$) was not significant.

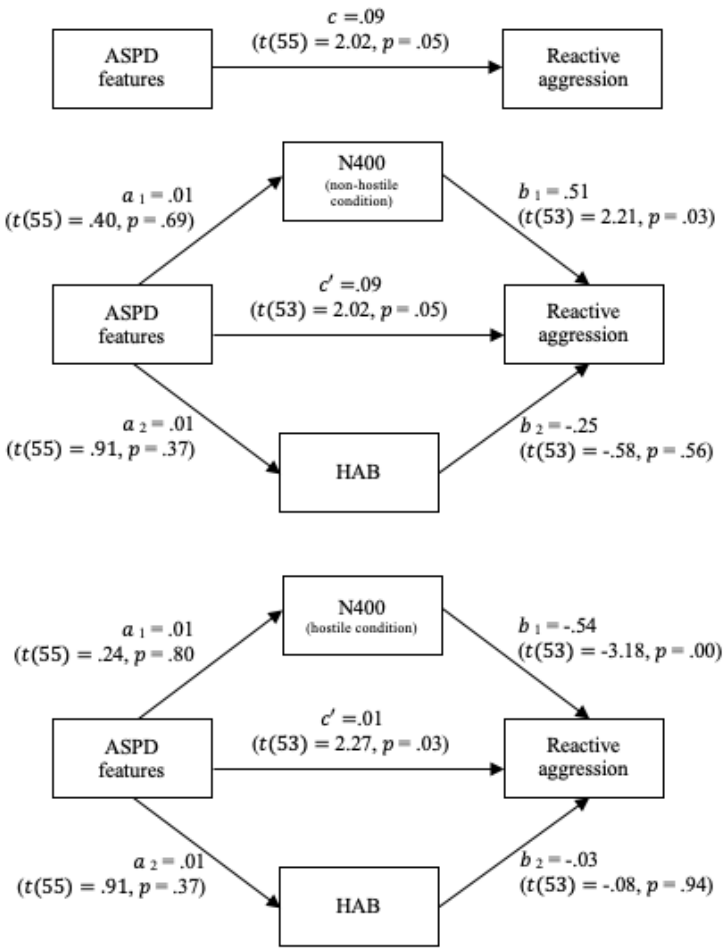

Figure 5. **Mediation of antisocial characteristics - reactive aggression relationship by the hostile attribution bias and the N400.**

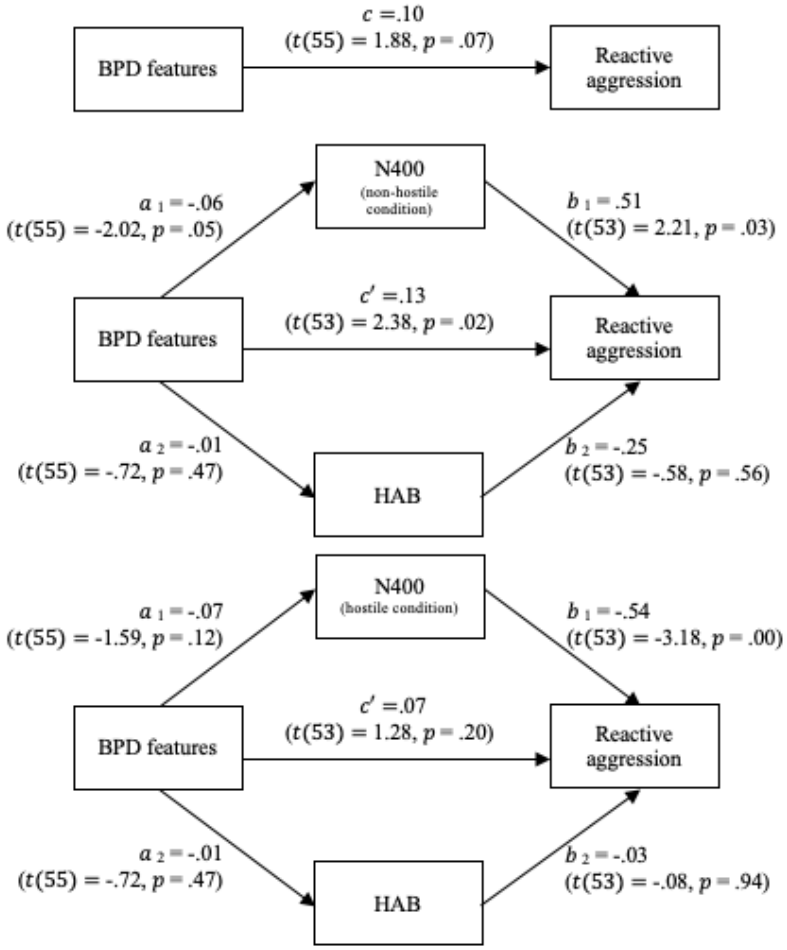

Figure 4. **Mediation of borderline characteristics - reactive aggression relationship by the hostile attribution bias and the N400.**

A second regression was conducted with BPD features as the independent variable, hostile expectancy violation (non-hostile condition) and self-reported HAB as mediator variables and ASPD features as the covariate. Borderline characteristics predicted hostile expectancy violation, but not self-reported HAB. Borderline characteristics did not predict REAG. Hostile expectancy violation predicted REAG whereas self-reported HAB did not predict REAG. When the mediator variables were added, borderline characteristics predicted REAG (R^2 = .46). The direct effect was greater than the total effect and the indirect effect was not significant for the hostile expectancy violation (indirect = -.03, *SE* = .03, 95% CI [-.10; .00]) and self-reported HAB (indirect = .00, *SE* = .01, 95% CI [-.01; .03]). The Sobel Aroian test also indicated that the indirect effect was not significant for hostile expectancy violation (Z = -1.24, p = .22) and self-reported HAB (Z = .23, p = .82). Therefore, there was no mediating effect of self-reported HAB or hostile expectancy violation on the relationship between borderline characteristics and REAG. Once again, the same regression was assessed with non-hostile expectancy violation (hostile condition) and self-reported HAB as mediator variables. Borderlines characteristics did not predict non-hostile expectancy violation. Non-hostile expectancy violation predicted REAG. With addition of the mediator variables, borderlines characteristics did not predict REAG (R^2 = .50). The total effect was greater than the direct effect, but the indirect effect was almost significant for the non-hostile expectancy violation (indirect = .04, *SE* = .03, 95% CI [-.01; .10]). The indirect effect was not significant for the self-reported HAB (indirect = -.00, *SE* = .01, 95% CI [.01; -.01]). With the Sobel Aroian test, the mediating effect was not significant for non-hostile expectancy violation (Z = 1.48, p = .14) and self-reported HAB (Z = .03, p = .98). Therefore, the non-hostile expectancy violations had no mediating effect on the relationship between borderline characteristics and REAG.

6.4. DISCUSSION

The first objective of this study was to replicate the measurement method of Gagnon et al. (2016) and to validate their results. The aim was to present scenarios describing social interactions while measuring brain activity. In each scenario, characters acted in a provocative and ambiguous manner in both hostile and non-hostile contexts. Participants were asked to read the scenarios on a screen and imagine the intentions behind the behaviors presented. Subsequently, the characters' hostile or non-hostile intentions were revealed through a final target word. The hostile or non-

hostile nature of the context could be in agreement or disagreement with the hostile or non-hostile nature of the intention. When the context was hostile, but the intention presented was non-hostile, the experimental manipulation was considered a hostile expectations violation. In contrast, when the context was non-hostile, but the intent was hostile, the manipulation was considered a non-hostile expectancy violation. As in the study by Gagnon et al. (2016), we were able to observe the N400 ERP component in a time window ranging from 350 to 650 ms post-stimulus onset. Moreover, the amplitude of this deflection was more pronounced during the hostile expectancy violation in the central and posterior cerebral regions at the medial and right electrodes. This meant that participants were surprised when a word signifying non-hostility was presented following a hostile context. This implied that participants attributed hostile intent to the characters when the context was hostile. This result has been corroborated by several other studies reporting maximum amplitude N400 in the centro-parietal regions when expectations are violated (Baetens et al., 2011; Bartholow et al., 2016; Gagnon et al., 2016; Leuthold et al., 2012; Van Overwalle et al., 2009). As demonstrated in Leuthold et al. (2012) and Gagnon et al. 2016, in our study, the N400 was observed post-stimulus due to target word being presented unexpectedly or inconsistently with the context of the scenario.

During the non-hostile expectancy violation, ERP amplitudes neared zero at approximately 350 to 650 ms. Therefore, when a non-hostile context was followed by ambiguous and provocative behavior, participants did not attribute a non-hostile intent to the behavior. Although consistent with findings reported in Gagnon et al. (2016), this result appears inconsistent with the established assumption that the N400 would reflect an expectancy violation. Gagnon et al. (2016) suggest this phenomenon possibly reflects a cautious interpretation, based on perceived cues, on the part of non-aggressive students. It is indeed possible that, in our study, non-hostile contextual cues conflicted with the ambiguous and provocative nature of the behavior. Therefore, the type of intent attribution depended on the weight the participant gave non-hostile cues versus provocative cues. In the end, in scenarios designed to violate non-hostile expectations, the participant may have had mixed views and not been systematically surprised to see hostile intent appear after a non-hostile context.

The second objective of this study was to demonstrate the predictive role of ASPD traits, BPD traits and self-reported HAB on the occurrence of aggressive behaviors. As expected, ASPD traits were highly correlated with REAG. Also, they positively predicted REAG, which is consistent with the scientific literature (Lobbestael et al., 2013; Ostrov & Houston, 2008; Ross & Babcock, 2009; Walters, 2007). The presence of high antisocial traits therefore predicted the occurrence of reactive aggressive behaviors. Thus, we were able to confirm our initial hypothesis that ASPD traits are good predictors of REAG.

BPD traits were highly correlated with REAG. However, when controlling for age, gender, education, depression, paranoid ideation and ASPD traits, they did not significantly predict REAG. This result was surprising given that several studies have shown BPD to be a good predictor of REAG (Lobbestael et al., 2015; Ostrov & Houston, 2008; Soliman & Reza, 2001; Zanarini et al., 2017). In a recent longitudinal study, however, Penson et al. (2018) reached the same conclusions as us, showing that BPD characteristics were not sufficient in significantly predicting aggressive behaviors and rather, that ASPD characteristics were better predictors. Thus, it is likely that, in our regressions, ASPD traits were more effective predictors of REAG than BPD traits and, as a result, received more of the variance of REAG. In addition, BPD traits and ASPD traits shared a high percentage of common variance ($r^2 = .24$), indicating some redundancy. It is therefore likely that the ASPD traits inherited the majority of the REAG variance because of the weak predictive role and redundancy of information in BPD traits. This would explain why the regression coefficient for BPD traits was insignificant for REAG. Thus, we were unable to confirm our third hypothesis that BPD traits are good predictors of REAG.

ASPD traits failed to predict both HAB, as measured by self-report, and hostile or non-hostile expectancy violations. These findings are not consistent with the few studies evaluating HAB in ASPD (Lobbestael et al., 2013; Smeijers et al., 2017). Indeed, Smeijers et al. (2017) reported that patients with ASPD exhibit a HAB regarding ambiguous facial expressions. Similarly, Lobbestael et al. (2013) used vignettes (verbal and pictorial stimuli) and successfully showed that ASPD traits were moderate predictors of the HAB. However, it is important to mention that the methodology used to measure the HAB in the latter two studies were not entirely similar to ours. We used the SIP-AEQ questionnaire to measure self-reported HAB and an electrophysiology method developed

by Gagnon et al. (2016) to measure hostile and non-hostile expectancies violations. As such, it is possible to expect different results in all three studies. In addition, it is possible that the individuals with ASPDs in our study showed fewer self-reported HABs as a result of the scenarios described in the SIP-AEQ and those selected in the EEG task. According to social-cognitive personality theory (Dodge et al., 2002), cognitive processes—such as hostile intent attribution processes—are situation-specific. We can therefore consider the possibility that situations provoking a (self-reported) HAB or a hostile expectancies violation in ASPD was not facilitated in our study, or that, depending on the scenario, the occurrence of the (self-reported) HAB or the hostile expectancy violation is not always systematic. It is possible, for example, to imagine that ASPD individuals were never really confronted with the situations described in our experience because they depict interactions that are too common or too conforming of societal norms. Thus, perhaps the situations described in our study were not provocative enough. Given that ASPD is characterized by a lack of conformity to societal norms (APA, 2013), it is possible that the individuals with ASPD in our study did not relate to the characters or that they experienced difficulty imagining the situations described in our experiment.

BPD traits also did not predict self-reported HAB and non-hostile expectancy violation. In contrast, they positively predicted the hostile expectancy violation (i.e., BPD traits negatively predicted N400). Thus, the higher the BPD traits, the stronger the hostile expectancy violation. In other words, when the context was hostile, people with high BPD traits made more hostile intent attributions than people with lower BPD traits. This result partially confirmed our expectations and was consistent with findings in Smeijers et al. (2017). In addition, several researchers have provided arguments regarding the meaning of such a prediction (Arntz et al., 2011; Baer et al., 2012; Barnow et al., 2009; Lobbestael & McNally, 2016). For example, Lobbestael & McNally (2016) demonstrated that people with BPD were subject to interpretive biases related to rejection and anger. According to Baer et al. (2012), people with BPD have negative beliefs about themselves and their environment. They also interpret and evaluate neutral and ambiguous stimuli negatively. Finally, according to Barnow et al. (2009) et Arntz et al. (2011), people with BPD judge other people as negative, aggressive and malicious. Another interesting finding is that in our correlation analyses, BPD traits were not correlated with the hostile expectancy violation. This may seem surprising given that, in our regression analyses, we were able to prove its role as a significant

predictor of the hostile expectancy violation. However, we would like to point out that our regressions controlled for several variables (including age, gender, education, depression and paranoid ideation). The fact that we did not find a significant correlation between BPD traits and non-hostile expectancy violation leads us to believe that the presence of high BPD traits was not associated with greater attribution of non-hostile intention when the context was non-hostile. Regardless of the level of BPD traits, the level of non-hostile expectancy violation remained the same. The fact that BPD traits did not predict self-reported HAB was surprising, given a mean positive correlation observed between the two variables. Among all the variables tested, paranoid ideation was most strongly correlated with self-reported HAB. In our regression models, it was also the variable that best predicted self-reported HAB ($\beta = .15$, $p = .00$; not reported in results section). Thus, it is likely that paranoid ideation received the majority of the variance of the self-reported HAB. Moreover, paranoid ideation shared a high common variance with BPD traits ($r^2 = .34$). Thus, it is possible that the BPD traits are too redundant compared to paranoid ideation in the prediction of HAB. This would explain why BPD traits did not predict self-reported HAB in our regressions. In the end, we were able to partially confirm our fourth hypothesis that ASPD and BPD traits significantly predict self-reported HAB and hostile expectancies violation.

Self-reported HAB did not predict REAG, which was surprising when considering the numerous studies showing that self-reported HAB is positively related to REAG (Bailey & Ostrov, 2008; Dodge et al., 2015; Gagnon & Rochat, 2017). However, there is a major bias in the measurement of the HAB using self-reports. This bias relates to an inability to capture early levels of information processing and therefore, spontaneous intention attribution processes. Thus, it is conceivable that the self-reported HAB is too biased to be associated with REAG.

Hostile expectancies violations showed a strong negative prediction of REAG (i.e., N400 positively predicted REAG). This result is somewhat unexpected. We assumed that hostile expectancies violation served as a measure of the HAB and thus would positively predict REAG, as reported in the literature (Bailey & Ostrov, 2008; Dodge et al., 2015; Gagnon & Rochat, 2017). This was not the case. We interpret the observed phenomenon as follows. First, let us recall that the hostile expectancies violation represents the subtraction of concordant non-hostile condition scenarios from discordant non-hostile condition scenarios (NHmi-NHma). In an NHmi scenario,

aggressors would be surprised to see a non-hostile intention appear after a hostile context. In an NHma scenario, an aggressive person would also be surprised to see a non-hostile intention appear after seeing a non-hostile context. Indeed, because of the chronic accessibility to hostile patterns, an aggressive person would see aggression in all their social interactions (Kim et al., 2019). Since NHmi and NHma conditions have a similar effect, their subtraction should have the effect of reducing the N400 (more positive amplitude going up) as aggressive traits increase. Thus, as hostile expectancies violation increases, REAG decreases. In addition to this interpretation, it appears that two other studies have found a negative relationship between HAB and REAG (Bowen et al., 2016; Coccaro, Fanning, & Lee, 2017). One possibility is that the relationship between HAB and REAG is more complex than we may think (Coccaro, Fanning, Fisher, et al., 2017). For Bowen et al. (2016), mean HAB scores were very low and they assumed that the aggressive individuals in their sample did not make sufficient HABs. Thus, another explanation is that variability in HAB scores was probably too small in our study to observe correlations established by most of the studies in the literature (Bailey & Ostrov, 2008; Dodge et al., 2015; Gagnon & Rochat, 2017).

Unexpectantly, non-hostile expectancies violation was a strong positive predictor of REAG (i.e., N400 negatively predicted REAG). In contrast to our assumption regarding the predictive role of hostile expectancies violation on REAG, we assumed that non-hostile expectancies violation would predict less REAG. A possible explanation can be found by recalling that the non-hostile expectancies violation was obtained by subtracting the concordant hostile condition scenarios from the concordant non-hostile condition (Hmi-Hma) scenarios. When the scenario was Hma, the non-aggressive participants were surprised to see hostile intent following a hostile context because they likely looked for the good in humans. When the scenario was Hmi, non-aggressive people were also surprised to see hostile intent after a non-hostile context because of inconsistent information (Leuthold et al., 2012). Thus, the difference between Hmi and Hma cancelled out the effect of surprise (which was roughly the same in both conditions) and the N400 appeared less negative. Therefore, the lower the non-hostile expectancies violation (more positive amplitude going up), the lower the REAG. As a result, we were unable to confirm our fifth hypothesis.

Finally, our final hypothesis that self-reported HAB, the hostile expectancy violation and the non-hostile expectancy violation were mediators of the relationship between cluster B

personality traits and REAG, could not be confirmed. The regression coefficient of the relationship between ASPD traits and REAG did not decrease significantly with the addition of self-reported HAB, hostile and non-hostile expectancy violation as mediators. Similarly, the regression coefficient of the relationship between BPD traits and REAG did not decrease significantly with the addition of self-reported HAB, hostile and non-hostile expectancy violation mediators. These findings were inconsistent with the few studies that have evaluated the mediating role of HAB in the relationship between ASPD and REAG and between BPD and REAG (Lobbestael et al., 2013; Smeijers et al., 2017). It is possible that the N400 effect (expectancy violation) may be influenced by other mediators, like sensitivity to rejection, impulsivity and dysfunctional beliefs (Baer et al., 2012; Lobbestael & McNally, 2016), which were not included in our study. Further studies are therefore needed to better understand the cognitive and affective processes underlying aggressive behavior in antisocial and borderline personality disorders.

This research project had several methodological limitations, such as sample size and heterogeneity. Indeed, according to G-Power analysis, the number of participants required to conduct a regression analysis with 8 predictors (self-reported expectancies violations, self-reported HAB, age, gender, education level, depression, paranoid ideation, BPD and ASPD traits) is 160. This is much higher than our current sample size (N = 63). Future analyses using a larger sample would be more adequate to better understand the nature of the observed relationships. In addition, our sample potentially over-represented students in the general population. Out of 63 participants, 49 were from an academic background. It would be interesting and beneficial to evaluate our measures on samples more representative of the general population.

6.5. CONCLUSION

In conclusion, our study replicated the measurement of expectation violations by electrophysiology and validated the presence of a strong negative deflection of ERP amplitudes at the time of hostile expectation violations, as demonstrated in the study by Gagnon et al. (2016). Additionally, our results show that antisocial traits, borderline traits and violation of high non-hostile expectations were positively associated with the frequency of reactive aggressive behaviors. Also, antisocial traits and violation of non-hostile expectations were positively correlated with the

frequency of proactive aggressive behavior. Our mediation models involving intention attribution processes as mediators could not be confirmed. Nevertheless, our data indicated that individuals with high borderline traits experienced more hostile expectation violations and that violation of hostile expectations predicted the frequency of reactive aggressive behaviors. To better understand the meaning of the relationship between expectancy violation and reactive aggression, it would be interesting to construct two groups of participants according to level (low or high) on reactive aggression and compare them according to level of surprise (i.e., strength of N400 deflection) experienced in each of the four scenario conditions.

6.6. Declaration of Interest

There is no conflict of interest. This project was approved by the Research Ethics Board of Education and Psychology of the University of Montreal. All procedures were consistent with the Énoncé de politique des trois conseils (EPTC-2, 2018).

6.7. Acknowledgements

This study was supervised by Dr. Jean Gagnon, PhD, from the Department of Psychology at the University of Montreal. The study was funded by a portion of the supervisor's research funds. The project was supported by Dr. Monique Bessette, PhD, psychologist and director of the Victoria Institute for participant recruitment. The author would like to thank Dr. Pierre Jolicoeur, PhD, of the Department of Psychology for his support in the analysis of EEG signals. The author would also like to thank the entire research team of the Laboratoire d'électrophysiologie en Neuroscience sociale for welcoming the participants and handing over the experimental tasks: Wan Seo Kim, MD, Gasser Saleh, MD, Joséphine Aldebert, B.Sc., Valérie Lapointe, B.Sc., Rodrigo Ahumada, B.Sc.

6.8. References

Adawi, M., Zerbetto, R., Re, T. S., Bisharat, B., Mahamid, M., Amital, H., Del Puente, G., & Bragazzi, N. L. (2019). Psychometric properties of the Brief Symptom Inventory in nomophobic subjects : Insights from preliminary

confirmatory factor, exploratory factor, and clustering analyses in a sample of healthy Italian volunteers. *Psychology Research and Behavior Management*, *12*, 145-154. https://doi.org/10.2147/PRBM.S173282

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5e éd.). American Psychiatric Publishing.

Arntz, A., Weertman, A., & Salet, S. (2011). Interpretation bias in Cluster-C and borderline personality disorders. *Behaviour Research and Therapy*, *49*(8), 472-481. https://doi.org/10.1016/j.brat.2011.05.002

Baer, R. A., Peters, J. R., Eisenlohr-Moul, T. A., Geiger, P. J., & Sauer, S. E. (2012). Emotion-related cognitive processes in borderline personality disorder : A review of the empirical literature. *Clinical Psychology Review*, *32*(5), 359-369. https://doi.org/10.1016/j.cpr.2012.03.002

Baetens, K., der Cruyssen, L. V., Achtziger, A., Vandekerckhove, M., & Van Overwalle, F. (2011). N400 and LPP in spontaneous trait inferences. *Brain Research*, *1418*, 83-92. https://doi.org/10.1016/j.brainres.2011.08.067

Bailey, C. A., & Ostrov, J. M. (2008). Differentiating Forms and Functions of Aggression in Emerging Adults : Associations with Hostile Attribution Biases and Normative Beliefs. *Journal of Youth and Adolescence*, *37*(6), 713-722. https://doi.org/10.1007/s10964-007-9211-5

Barnow, S., Stopsack, M., Grabe, H. J., Meinke, C., Spitzer, C., Kronmüller, K., & Sieswerda, S. (2009). Interpersonal evaluation bias in borderline personality disorder. *Behaviour Research and Therapy*, *47*(5), 359-365. https://doi.org/10.1016/j.brat.2009.02.003

Bartholow, B. D., Fabiani, M., Gratton, G., & Bettencourt, B. A. (2016). A Psychophysiological Examination of Cognitive Processing of and Affective Responses to Social Expectancy Violations: *Psychological Science*. https://journals.sagepub.com/doi/10.1111/1467-9280.00336

Bowen, K., Roberts, J., & Kocian, E. (2016). Decision making of inmates : Testing social information processing concepts using vignettes. *Applied Psychology in Criminal Justice*, *2016*, 1-17.

Bushman, B. J., & Huesmann, L. R. (2010). Aggression. Dans *Handbook of Social Psychology*. American Cancer Society. https://doi.org/10.1002/9780470561119.socpsy002023

Camodeca, M., & Goossens, F. A. (2005). Aggression, social cognitions, anger and sadness in bullies and victims. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *46*(2), 186-197. https://doi.org/10.1111/j.1469-7610.2004.00347.x

Chabrol, H., Valls, M., van Leeuwen, N., & Bui, E. (2012). *Callous-unemotional and borderline traits in nonclinical adolescents : Personality profiles and relations to antisocial behaviors*.

https://doi.org/10.1016/j.paid.2012.07.017

Chen, P., Coccaro, E. F., & Jacobson, K. C. (2012). Hostile Attributional Bias, Negative Emotional Responding, and Aggression in Adults : Moderating Effects of Gender and Impulsivity. *Aggressive Behavior*, *38*(1), 47-63. https://doi.org/10.1002/ab.21407

Cima, M., Raine, A., Meesters, C., & Popma, A. (2013). Validation of the Dutch Reactive Proactive Questionnaire (RPQ) : Differential Correlates of Reactive and Proactive Aggression From Childhood to Adulthood. *Aggressive Behavior*, *39*(2), 99-113. https://doi.org/10.1002/ab.21458

Coccaro, E. F., Fanning, J., & Lee, R. (2017). Development of a social emotional information processing assessment for adults (SEIP-Q). *Aggressive Behavior*, *43*(1), 47-59. https://doi.org/10.1002/ab.21661

Coccaro, E. F., Fanning, J. R., Fisher, E., Couture, L., & Lee, R. J. (2017). Social emotional information processing in adults : Development and psychometrics of a computerized video assessment in healthy controls and aggressive individuals. *Psychiatry Research*, *248*, 40-47. https://doi.org/10.1016/j.psychres.2016.11.004

Coccaro, E. F., Noblett, K. L., & McCloskey, M. S. (2009). Attributional and emotional responses to socially ambiguous cues : Validation of a new assessment of social/emotional information processing in healthy adults and impulsive aggressive patients. *Journal of Psychiatric Research*, *43*(10), 915-925. https://doi.org/10.1016/j.jpsychires.2009.01.012

Conroy, S., Burczycka, M., & Savage, L. (2019). *Family violence in Canada : A statistical profile, 2018*. *85*, 62.

Crick, N. R., & Dodge, K. A. (1996). Social Information-Processing Mechanisms in Reactive and Proactive Aggression. *Child Development*, *67*(3), 993-1002. JSTOR. https://doi.org/10.2307/1131875

De Castro, B. O., Veerman, J. W., Koops, W., Bosch, J. D., & Monshouwer, H. J. (2002). Hostile Attribution of Intent and Aggressive Behavior : A Meta-Analysis. *Child Development*, *73*(3), 916-934. https://doi.org/10.1111/1467-8624.00447

Derogatis, L. R. (1993). *BSI, Brief Symptom Inventory : Administration, scoring & procedures manual*. National Computer Systems.

Dodge, K. A., Laird, R., Lochman, J. E., & Zelli, A. (2002). Multidimensional Latent-Construct Analysis of Children's Social Information Processing Patterns : Correlations With Aggressive Behavior Problems. *Psychological assessment*, *14*(1), 60-73.

Dodge, K. A., Malone, P. S., Lansford, J. E., Sorbring, E., Skinner, A. T., Tapanya, S., Tirado, L. M. U., Zelli, A., Alampay, L. P., Al-Hassan, S. M., Bacchini, D., Bombi, A. S., Bornstein, M. H., Chang, L., Deater-Deckard, K., Di Giunta, L., Oburu, P., & Pastorelli, C. (2015). Hostile attributional bias and aggressive behaviour in global context. *Proceedings of the National Academy of*

Sciences of the United States of America, *112*(30), 9310-9315. https://doi.org/10.1073/pnas.1418572112

Drisdelle, B. L., Aubin, S., & Jolicoeur, P. (2017). Dealing with ocular artifacts on lateralized ERPs in studies of visual-spatial attention and memory : ICA correction versus epoch rejection. *Psychophysiology*, *54*(1), 83-99. https://doi.org/10.1111/psyp.12675

Duberstein, P. R., & Conwell, Y. (1997). Personality Disorders and Completed Suicide : A Methodological and Conceptual Review. *Clinical Psychology: Science and Practice*, *4*(4), 359-376. https://doi.org/10.1111/j.1468-2850.1997.tb00127.x

Dunsieth, N. W., Nelson, E. B., Brusman-Lovins, L. A., Holcomb, J. L., Beckman, D., Welge, J. A., Roby, D., Taylor, P., Soutullo, C. A., & McElroy, S. L. (2004). Psychiatric and legal features of 113 men convicted of sexual offenses. *The Journal of Clinical Psychiatry*, *65*(3), 293-300.

Edens, J., Campbell, J., & Weir, J. (2007). Youth Psychopathy and Criminal Recidivism : A Meta-Analysis of the Psychopathy Checklist Measures. *Law and human behaviour*, *31*, 53-75. https://doi.org/10.1007/s10979-006-9019-y

E-Prime (Version 2.0.10.356). (s. d.). [Computer Software]. Pittsburgh, PA: Psychology Software Tools.

Fazel, S., & Danesh, J. (2002). Serious mental disorder in 23000 prisoners : A systematic review of 62 surveys. *Lancet (London, England)*, *359*(9306), 545-550. https://doi.org/10.1016/S0140-6736(02)07740-1

Fitz, H., & Chang, F. (2019). Language ERPs reflect learning through prediction error propagation. *Cognitive Psychology*, *111*, 15-52. https://doi.org/10.1016/j.cogpsych.2019.03.002

Fossati, A., Raine, A., Borroni, S., Bizzozero, A., Volpi, E., Santalucia, I., & Maffei, C. (2009). A cross-cultural study of the psychometric properties of the Reactive–Proactive Aggression Questionnaire among Italian nonclinical adolescents. *Psychological Assessment*, *21*(1), 131. https://doi.org/10.1037/a0014743

Frank, S. L., Otten, L. J., Galli, G., & Vigliocco, G. (2015). The ERP response to the amount of information conveyed by words in sentences. *Brain and Language*, *140*, 1-11. https://doi.org/10.1016/j.bandl.2014.10.006

Frick, P. J., Ray, J. V., Thornton, L. C., & Kahn, R. E. (2014). Can callous-unemotional traits enhance the understanding, diagnosis, and treatment of serious conduct problems in children and adolescents? A comprehensive review. *Psychological Bulletin*, *140*(1), 1-57. https://doi.org/10.1037/a0033076

Gagnon, J., Aubin, M., Emond, F. C., Derguy, S., Bessette, M., & Jolicoeur, P. (2016). Neural mechanisms underlying attribution of hostile intention in nonaggressive individuals : An ERP study. *International Journal of Psychophysiology*, *110*, 153-162.

https://doi.org/10.1016/j.ijpsycho.2016.08.007

Gagnon, J., McDuff, P., Daelman, S., & Fournier, S. (2015). Is hostile attributional bias associated with negative urgency and impulsive behaviors? A social-cognitive conceptualization of impulsivity. *Personality and Individual Differences*, *72*, 18-23. https://doi.org/10.1016/j.paid.2014.08.011

Gagnon, J., & Rochat, L. (2017). Relationships between hostile attribution bias, negative urgency, and reactive aggression. *Journal of Individual Differences*, *38*(4), 211-219. https://doi.org/10.1027/1614-0001/a000238

Gardner, K. J., Archer, J., & Jackson, S. (2012). Does Maladaptive Coping Mediate the Relationship Between Borderline Personality Traits and Reactive and Proactive Aggression? *Aggressive Behavior*, *38*(5), 403-413. https://doi.org/10.1002/ab.21437

Gosselin, M., & Bergeron, J. (1993). *Évaluation des qualités psychométriques du questionnaire de santé mentale : SCL-90-R*. Recherche et intervention sur les substances psychoactives-Québec.

Grant, B. F., Chou, S. P., Goldstein, R. B., Huang, B., Stinson, F. S., Saha, T. D., Smith, S. M., Dawson, D. A., Pulay, A. J., Pickering, R. P., & Ruan, W. J. (2008). Prevalence, Correlates, Disability, and Comorbidity of DSM-IV Borderline Personality Disorder : Results from the Wave 2 National Epidemiologic Survey on Alcohol and Related Conditions. *The Journal of clinical psychiatry*, *69*(4), 533-545.

Grant, B. F., Hasin, D. S., Stinson, F. S., Dawson, D. A., Chou, S. P., Ruan, W. J., & Pickering, R. P. (2004). Prevalence, correlates, and disability of personality disorders in the United States : Results from the national epidemiologic survey on alcohol and related conditions. *The Journal of Clinical Psychiatry*, *65*(7), 948-958. https://doi.org/10.4088/jcp.v65n0711

Haller, J., & Kruk, M. R. (2006). Normal and abnormal aggression : Human disorders and novel laboratory models. *Neuroscience & Biobehavioral Reviews*, *30*(3), 292-303. https://doi.org/10.1016/j.neubiorev.2005.01.005

Hayes, A. F. (2018). *Introduction to Mediation, Moderation, and Conditional Process Analysis, Second Edition : A Regression-Based Approach: Vol. Second edition*. The Guilford Press. http://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=1617631&lang=fr&site=ehost-live

Karlin, B. E., Creech, S. K., Grimes, J. S., Clark, T. S., Meagher, M. W., & Morey, L. C. (2005). The Personality Assessment Inventory with chronic pain patients : Psychometric properties and clinical utility. *Journal of Clinical Psychology*, *61*(12), 1571-1585. https://doi.org/10.1002/jclp.20209

Khalil, A. A., Hall, L. A., Moser, D. K., Lennie, T. A., & Frazier, S. K. (2011). The Psychometric Properties of the Brief Symptom Inventory Depression and Anxiety Subscales in Patients With

Heart Failure and With or Without Renal Dysfunction. *Archives of Psychiatric Nursing*, *25*(6), 419-429. https://doi.org/10.1016/j.apnu.2010.12.005

Kim, W. S., Jolicoeur, P., & Gagnon, J. (2019). *Les effets non réciproques de l'activation du schéma hostile et non hostile sur le processus d'attribution d'intention chez les individus non agressifs*. https://papyrus.bib.umontreal.ca/xmlui/handle/1866/22736

Krug, E. G. (Éd.). (2002). *World report on violence and health*.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting : Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*, 621-647. https://doi.org/10.1146/annurev.psych.093008.131123

Lansford, J. E., Malone, P. S., Dodge, K. A., Crozier, J. C., Pettit, G. S., & Bates, J. E. (2006). A 12-Year Prospective Study of Patterns of Social Information Processing Problems and Externalizing Behaviors. *Journal of abnormal child psychology*, *34*(5), 715-724. https://doi.org/10.1007/s10802-006-9057-4

Lenzenweger, M. F., Lane, M. C., Loranger, A. W., & Kessler, R. C. (2007). DSM-IV personality disorders in the National Comorbidity Survey Replication. *Biological psychiatry*, *62*(6), 553-564. https://doi.org/10.1016/j.biopsych.2006.09.019

Leuthold, H., Filik, R., Murphy, K., & Mackenzie, I. G. (2012). The on-line processing of socio-emotional information in prototypical scenarios : Inferences from brain potentials. *Social Cognitive and Affective Neuroscience*, *7*(4), 457-466. https://doi.org/10.1093/scan/nsr029

Lobbestael, J., Cima, M., & Arntz, A. (2013). The Relationship Between Adult Reactive and Proactive Aggression, Hostile Interpretation Bias, and Antisocial Personality Disorder. *Journal of Personality Disorders*, *27*(1), 53-66. https://doi.org/10.1521/pedi.2013.27.1.53

Lobbestael, J., Cima, M., & Lemmens, A. (2015). The relationship between personality disorder traits and reactive versus proactive motivation for aggression. *Psychiatry Research*, *229*(1-2), 155-160. https://doi.org/10.1016/j.psychres.2015.07.052

Lobbestael, J., & McNally, R. J. (2016). An Empirical Test of Rejection- and Anger-Related Interpretation Bias in Borderline Personality Disorder. *Journal of Personality Disorders*, *30*(3), 307-319. https://doi.org/10.1521/pedi_2015_29_194

Miller, J. D., & Lynam, D. R. (2006). Reactive and proactive aggression : Similarities and differences. *Personality and Individual Differences*, *41*(8), 1469-1480. https://doi.org/10.1016/j.paid.2006.06.004

Morey, L. C. (1991). Inventaire de la personnalité (translate by C. Routhier), Personality Assessment Inventory: Professional manual. Odessa: FL: Psychological Assessment Ressources.

Morey, L. C. (2014). The Personality Assessment Inventory. Dans *Personality assessment, 2nd ed* (p. 181-228). Routledge/Taylor & Francis Group.

Newhill, C. E., Eack, S. M., & Mulvey, E. P. (2009). Violent Behavior in Borderline Personality. *Journal of Personality Disorders*, *23*(6), 541-554. https://doi.org/10.1521/pedi.2009.23.6.541

Ostrov, J. M., & Houston, R. J. (2008). The Utility of Forms and Functions of Aggression in Emerging Adulthood : Association with Personality Disorder Symptomatology. *Journal of Youth and Adolescence*, *37*(9), 1147-1158. https://doi.org/10.1007/s10964-008-9289-4

Penson, B. N., Ruchensky, J. R., Morey, L. C., & Edens, J. F. (2018). Using the Personality Assessment Inventory Antisocial and Borderline Features Scales to Predict Behavior Change : A Multisite Longitudinal Study of Youthful Offenders. *Assessment*, *25*(7), 858-866. https://doi.org/10.1177/1073191116680292

Putkonen, H., Komulainen, E. J., Virkkunen, M., Eronen, M., & Lönnqvist, J. (2003). Risk of repeat offending among violent female offenders with psychotic and personality disorders. *The American Journal of Psychiatry*, *160*(5), 947-951. https://doi.org/10.1176/appi.ajp.160.5.947

Raine, A., Dodge, K., Loeber, R., Gatzke-Kopp, L., Lynam, D., Reynolds, C., Stouthamer-Loeber, M., & Liu, J. (2006). The Reactive–Proactive Aggression Questionnaire : Differential Correlates of Reactive and Proactive Aggression in Adolescent Boys. *Aggressive behaviour*, *32*(2), 159-171. https://doi.org/10.1002/ab.20115

Ramirez, J., & Andreu, J. (2006). Aggression, and some related psychological constructs (anger, hostility and impulsivity) Some comments from a research project. *Neuroscience and Biobehavioural Reviews*, *30*, 276-291.

Ross, J. M., & Babcock, J. C. (2009). Proactive and Reactive Violence among Intimate Partner Violent Men Diagnosed with Antisocial and Borderline Personality Disorder. *Journal of Family Violence*, *24*(8), 607-617. https://doi.org/10.1007/s10896-009-9259-y

Ruiter, C. de, & Greeven, P. G. J. (2000). Personality Disorders in a Dutch Forensic Psychiatric Sample : Convergence of Interview and Self-Report Measures. *Journal of Personality Disorders*, *14*(2), 162-170. https://doi.org/10.1521/pedi.2000.14.2.162

Sharbrough, F., Chatrian, G. E., Lesser, R., Luders, H., Nuwer, M., & Picton, T. (1991). American Electroencephalographic Society guidelines for standard electrode position nomenclature. *Clinical Neurophysiology*, *8*, 200-202.

Slavin-Mulford, J., Sinclair, S. J., Stein, M., Malone, J., Bello, I., & Blais, M. A. (2012). External Validity of the Personality Assessment Inventory (PAI) in a Clinical Sample. *Journal of Personality Assessment*, *94*(6), 593-600.

https://doi.org/10.1080/00223891.2012.681817

Smeijers, D., Rinck, M., Bulten, E., Heuvel, T. van den, & Verkes, R.-J. (2017). Generalized hostile interpretation bias regarding facial expressions : Characteristic of pathological aggressive behaviour. *Aggressive Behavior*, *43*(4), 386-397. https://doi.org/10.1002/ab.21697

Soliman, A. E.-D., & Reza, H. (2001). Risk Factors and Correlates of Violence Among Acutely Ill Adult Psychiatric Inpatients. *Psychiatric Services*, *52*(1), 75-80. https://doi.org/10.1176/appi.ps.52.1.75

Soloff, P. H., Lynch, K. G., Kelly, T. M., Malone, K. M., & Mann, J. J. (2000). Characteristics of suicide attempts of patients with major depressive episode and borderline personality disorder : A comparative study. *The American Journal of Psychiatry*, *157*(4), 601-608. https://doi.org/10.1176/appi.ajp.157.4.601

Tiihonen, J., Eronen, M., & Hakola, P. (1993). Criminality Associated With Mental Disorders and Intellectual Deficiency. *Archives of General Psychiatry*, *50*(11), 917-918. https://doi.org/10.1001/archpsyc.1993.01820230087010

Ullrich, S., Deasy, D., Smith, J., Johnson, B., Clarke, M., Broughton, N., & Coid, J. (2008). Detecting personality disorders in the prison population of England and Wales : Comparing case identification using the SCID-II screen and the SCID-II clinical interview. *The Journal of Forensic Psychiatry & Psychology*, *19*(3), 301-322. https://doi.org/10.1080/14789940802045182

Van Overwalle, F., Van den Eede, S., Baetens, K., & Vandekerckhove, M. (2009). Trait inferences in goal-directed behaviour : ERP timing and localization under spontaneous and intentional processing. *Social Cognitive and Affective Neuroscience*, *4*(2), 177-190. https://doi.org/10.1093/scan/nsp003

Walters, G. D. (2007). Measuring Proactive and Reactive Criminal Thinking With the PICTS : Correlations With Outcome Expectancies and Hostile Attribution Biases. *Journal of Interpersonal Violence*, *22*(4), 371-385. https://doi.org/10.1177/0886260506296988

Warren, J. I., Burnette, M., South, S. C., Chauhan, P., Bale, R., & Friend, R. (2002). Personality disorders and violence among female prison inmates. *The Journal of the American Academy of Psychiatry and the Law*, *30*(4), 502-509.

Werner, N. E., & Crick, N. R. (1999). Relational aggression and social-psychological adjustment in a college sample. *Journal of Abnormal Psychology*, *108*(4), 615-623. https://doi.org/10.1037/0021-843X.108.4.615

World Health Organization. (2017, mai). *10 facts about violence prevention*. WHO; World Health Organization. http://www.who.int/features/factfiles/violence/en/

Yeager, D. S., Miu, A. S., Powers, J., & Dweck, C. S. (2013). Implicit theories of personality and attributions of hostile intent : A meta-analysis, an experiment,

and a longitudinal intervention. *Child Development*, *84*(5), 1651-1667. https://doi.org/10.1111/cdev.12062

Zanarini, M. C., Temes, C. M., Ivey, A. M., Cohn, D. M., Conkey, L. C., Frankenburg, F. R., & Fitzmaurice, G. M. (2017). The 10-year Course of Adult Aggression toward Others in Patients with Borderline Personality Disorder and Axis II Comparison Subjects. *Psychiatry research*, *252*, 134-138. https://doi.org/10.1016/j.psychres.2017.02.054

7. CHAPITRE III. CONCLUSION GÉNÉRALE

Ce projet de recherche a été conçu dans le but de répondre à deux principaux objectifs.

Le premier objectif était de répliquer et d'évaluer une méthode de mesure du HAB créée par Gagnon et al. (2016). Cette méthode impliquait l'administration d'une tâche de lecture de scénarios pendant l'enregistrement de l'activité cérébrale des participants. Les scénarios décrivaient des personnages interagissant dans des scènes de la vie quotidienne. Pour chaque scénario, les participants devaient évaluer des comportements ambigus et potentiellement provocateurs dans un contexte décrit (hostile ou non hostile) et imaginer les intentions alimentant ces comportements. Par la suite, une dernière phrase apparaissait dans le but de clarifier les intentions derrière les comportements des personnages. L'intention révélée pouvait être hostile ou non hostile. Comme décrit précédemment, les scénarios pouvaient être de l'ordre de quatre conditions : hostile concordante, hostile discordante, non hostile concordant et non hostile discordant. Lorsque les conditions étaient hostiles, le mot cible décrivait une intention hostile derrière le comportement du personnage. À l'inverse, lorsque les conditions étaient non hostiles, l'intention décrite était non hostile. Les conditions étaient dites concordantes lorsque la nature hostile ou non hostile de l'intention était cohérente avec la nature hostile ou non hostile du contexte. De même, les conditions étaient dites discordantes, lorsque la nature hostile ou non hostile de l'intention différait de la nature hostile et non hostile du contexte. Tout comme obtenu dans l'étude de Gagnon et al. (2016), nous nous attendions à observer une déflexion négative des amplitudes des ERP (composante N400) enregistrées lors de la violation des attentes hostiles dans les régions cérébrales postérieures droites et c'est effectivement ce que nous avons trouvé. Plus précisément, la déflexion négative N400 était la plus prononcée lors de la violation des attentes hostiles dans les régions centrales et postérieures, aux électrodes médianes et droites. Ce résultat a été corroboré par de nombreuses autres études dans la littérature scientifique (Baetens et al., 2011; Bartholow et al., 2016; Leuthold et al., 2012; Van Overwalle et al., 2009).

Le deuxième objectif de ce projet était d'examiner le rôle médiateur de la violation des attentes (hostiles ou non) dans la relation entre le TPAS et la REAG. Pour y répondre, les traits de personnalité, les processus d'attribution d'intention et les comportements agressifs des participants ont été évalués par l'administration de questionnaires autorapportés. Puis, toutes les variables

d'intérêts, dont les mesures d'électrophysiologie de la violation des attentes ont été étudiées dans des analyses de corrélation et de régression. Nous avons émis cinq hypothèses pour guider notre interprétation des résultats. La première hypothèse formulée était celle selon laquelle le TPAS prédisait positivement les comportements agressifs. Cette première hypothèse a pu être confirmée comme dans d'autres études (Lobbestael et al., 2013; Ostrov & Houston, 2008; Ross & Babcock, 2009; Tweed & Dutton, 1998; Walters, 2007). Ainsi, nous étions capables de dire que le TPAS prédisait positivement et fortement les comportements agressifs réactifs et proactifs. Selon notre deuxième hypothèse, le TPAS prédisait positivement le HAB et la violation des attentes hostiles. Toutefois, ce postulat n'a pas pu être validé dans notre étude. Contrairement à ce qui fut rapporté par la communauté scientifique (Lobbestael et al., 2013; Smeijers et al., 2017), le TPAS ne prédisait pas significativement le HAB ou la violation des attentes. Pour notre troisième hypothèse, nous nous attendions à ce que le HAB et la violation des attentes prédisent positivement les comportements agressifs réactifs comme l'ont observé de nombreux chercheurs (Bailey & Ostrov, 2008; Basquill et al., 2004; Camodeca & Goossens, 2005; De Castro et al., 2002; Dodge, 2006; Dodge et al., 2015; Gagnon & Rochat, 2017; Hubbard et al., 2001; MacBrayer et al., 2003; Matthews & Norris, 2002; Miller & Lynam, 2006). Contre toute attente, la violation des attentes hostiles prédisait négativement la REAG tandis que la violation des attentes non hostiles prédisait positivement la REAG. Le HAB lui, ne prédisait pas la REAG. Selon notre quatrième hypothèse, la violation des attentes était reliée positivement et faiblement à la PEAG. Ce postulat a pu être confirmé puisque nous sommes parvenus à trouver une corrélation positive et moyenne entre la violation des attentes non hostiles et la PEAG, comme d'autres études présentées dans la méta-analyse de De Castro et al. (2002),. Enfin, pour notre cinquième hypothèse, nous avons postulé que l'ajout du HAB ou de la violation des attentes comme médiateurs dans les régressions changerait la qualité de la relation entre le TPAS et la REAG. Cette supposition n'a pas pu être prouvée, car, dans nos modèles de régression, l'ajout de la variable HAB ou la violation des attentes (hostiles ou non) ne modifiait pas significativement la relation entre le TPAS et les comportements agressifs réactifs.

Afin d'expliquer les résultats inattendus ou non conformes à ceux de la communauté scientifique, nous avons élaboré des pistes de réflexion et d'interprétations de nos données. En ce qui concerne le rôle du TPAS dans la prédiction du HAB et de la violation des attentes, nous

spéculons que les personnes ayant plus de traits antisociaux éprouvaient des difficultés à s'identifier aux situations décrites dans la tâche EEG de Gagnon et al. (2016) et le questionnaire évaluant les processus d'attribution d'intention (SIP-AEQ; Coccaro, Noblett, & McCloskey, 2009). Selon le DSM-5, le TPAS se caractérise par des interactions marginales et peu conformes aux normes sociétales (APA, 2013). Or dans nos tâches expérimentales, les interactions sociales décrites sont quotidiennes et socialement courantes. Ainsi, il est fort probable que les personnes avec des traits du TPAS n'aient presque jamais été confrontées aux situations décrites et aient eu du mal à les interpréter. Par ailleurs, selon la théorie de la personnalité sociocognitive, les processus cognitifs tels que les processus d'attribution d'intention hostile, sont spécifiques à une situation donnée (Dodge et al., 2002). Alors, nous pouvons envisager que nos scénarios ne déclenchaient pas systématiquement de HAB ou de violation des attentes chez le TPAS. À propos de l'absence de lien entre le HAB et la REAG, nous supposons que le questionnaire SIP-AEQ était trop biaisé pour être associé à la REAG. Étant une mesure autorapportée subjective, ce questionnaire ne permet pas de capter les premiers niveaux du traitement de l'information sociale (tel que les processus d'attribution d'intention spontanés) de Crick & Dodge (1996). Ainsi, il nous est facile de concevoir que les personnes complétant le SIP-AEQ aient suffisamment le temps d'explorer différentes interprétations et d'évaluer les patrons de réponses préférées avant de porter un jugement et donner leur réponse. Pour le sens des relations observées entre la violation des attentes hostiles et la REAG et, entre la violation des attentes non hostiles et REAG, nous interprétons le phénomène comme découlant de la manière dont est opérationnalisée la violation des attentes. La violation des attentes hostiles représente la soustraction entre les scénarios de condition non hostile concordante et les scénarios de condition non hostile discordante (NHD-NHC). Quant à la violation des attentes non hostiles, il s'agit plutôt d'une soustraction entre les scénarios de type hostile concordant et les scénarios de type hostile discordant (HD-HC). Ainsi, l'ampleur de la déflexion négative N400 dépend de l'ampleur des différences entre les conditions discordantes et les conditions concordantes. En considérant cela, la violation des attentes hostiles (NHD-NHC) peut paraître plus faible (i.e. N400 moins négative) chez les individus agressifs si le niveau de surprise ressentie chez cette population est le même dans les scénarios de type NHD que dans les scénarios de type NHC. Cela peut par exemple se produire dans le cas d'une accessibilité chronique à des schémas hostiles (Kim et al., 2019). Percevant de l'hostilité dans toutes leurs interactions, les agressifs peuvent être surpris de voir apparaître une intention non hostile après la présentation d'un

contexte qu'il soit hostile ou non. Dans le même ordre d'idées, la violation des attentes non hostiles (HD-HC) peut paraître plus faible (i.e. N400 moins négative) chez les individus non agressifs, si le niveau de surprise éprouvé dans un scénario hostile discordant est le même que celui éprouvé dans un scénario hostile discordant. Ceci peut se produire dans le cas d'une résistance à l'accumulation d'indices hostiles. En voulant systématiquement percevoir l'être humain comme un être bon, les participants non agressifs vont être surpris de voir une intention hostile apparaître, quelle que soit la nature (hostile ou non hostile) du contexte.

Notre projet de recherche comportait plusieurs limites méthodologiques, telles que la taille et l'hétérogénéité de l'échantillon. En effet, le nombre de participants nécessaires pour conduire une analyse de régression avec une variable indépendante, une variable dépendante, deux médiateurs (violation des attentes et HAB autorapportés) et six covariables (âge, sexe, niveau d'éducation, traits de dépression, d'idéation paranoïde et limites) est de 166 selon G-Power. Ce qui est bien plus élevé que la taille actuelle de notre échantillon (N = 63). Des analyses futures sur un plus large échantillon serait alors adéquates pour mieux comprendre la nature des relations observées. De plus, notre échantillon était très hétérogène et sensible à une potentielle surreprésentation de la quantité d'étudiants dans la population générale. Sur 63 participants, 49 provenaient d'un milieu universitaire. Il serait intéressant d'évaluer nos mesures sur des échantillons plus représentatifs de la population générale.

En conclusion, notre étude a permis de développer nos connaissances actuelles sur les cognitions agressives et leurs influences sur la survenue des comportements agressifs. Par ailleurs, elle conscientise davantage la communauté scientifique au sujet de l'importance d'investiguer sur les processus cognitifs à l'origine de l'agressivité réactive chez la personnalité antisociale. Ceci permettrait de développer des axes de traitement et de contribuer à long terme à la diminution de la criminalité dans le monde.

8. BIBLIOGRAPHIE

Allen, J. J., & Anderson, C. A. (2017). Aggression and Violence : Definitions and Distinctions. Dans *The Wiley Handbook of Violence and Aggression* (p. 1-14). American Cancer Society. https://doi.org/10.1002/9781119057574.whbva001

Allen, J. J., Anderson, C. A., & Bushman, B. J. (2018). The General Aggression Model. *Current Opinion in Psychology*, *19*, 75-80. https://doi.org/10.1016/j.copsyc.2017.03.034

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5e éd.). American Psychiatric Publishing.

Anderson, C. A., & Bushman, B. J. (2002). Human Aggression. *Annual Review of Psychology*, *53*(1), 27-51. https://doi.org/10.1146/annurev.psych.53.100901.135231

Anderson, C. A., & Carnagey, N. L. (2004). Violent evil and the General Aggression Model. Dans *The social psychology of good and evil* (p. 168-192). The Guilford Press.

Baetens, K., der Cruyssen, L. V., Achtziger, A., Vandekerckhove, M., & Van Overwalle, F. (2011). N400 and LPP in spontaneous trait inferences. *Brain Research*, *1418*, 83-92. https://doi.org/10.1016/j.brainres.2011.08.067

Bailey, C. A., & Ostrov, J. M. (2008). Differentiating Forms and Functions of Aggression in Emerging Adults : Associations with Hostile Attribution Biases and Normative Beliefs. *Journal of Youth and Adolescence*, *37*(6), 713-722. https://doi.org/10.1007/s10964-007-9211-5

Bandura, A. (1978). Social Learning Theory of Aggression. *Journal of Communication*, *28*(3), 12-29. https://doi.org/10.1111/j.1460-2466.1978.tb01621.x

Bandura, A. (2001). Social Cognitive Theory : An Agentic Perspective. *Annual Review of Psychology*, *52*(1), 1-26. https://doi.org/10.1146/annurev.psych.52.1.1

Barratt, E. S., Stanford, M. S., Dowdy, L., Liebman, M. J., & Kent, T. A. (1999). Impulsive and premeditated aggression : A factor analysis of self-reported acts. *Psychiatry Research*, *86*(2), 163-173. https://doi.org/10.1016/s0165-1781(99)00024-4

Barratt, E. S., Stanford, M. S., Kent, T. A., & Alan, F. (1997). Neuropsychological and cognitive psychophysiological substrates of impulsive aggression. *Biological Psychiatry*, *41*(10), 1045-1061. https://doi.org/10.1016/S0006-3223(96)00175-8

Bartholow, B. D., Fabiani, M., Gratton, G., & Bettencourt, B. A. (2016). A Psychophysiological Examination of Cognitive Processing of and Affective Responses to Social Expectancy Violations: *Psychological Science*. https://journals.sagepub.com/doi/10.1111/1467-9280.00336

Basquill, M. F., Nezu, C. M., Nezu, A. M., & Klein, T. L. (2004). Aggression-related hostility bias and social problem-solving deficits in adult males with mental retardation. *American Journal of Mental Retardation: AJMR*, *109*(3), 255-263. https://doi.org/10.1352/0895-8017(2004)109<255:AHBASP>2.0.CO;2

Bowen, F., Levasseur, C., & Desbiens, N. (2014). L'agressivité et la nature humaine : Perspectives évolutionnistes et éthologique du développement de l'intimidation. *Revue québécoise de psychologie*, *35*(3), 63-88.

Bowen, K., Roberts, J., & Kocian, E. (2016). Decision making of inmates : Testing social information processing concepts using vignettes. *Applied Psychology in Criminal Justice*, *2016*, 1-17.

Brent, D. A., Johnson, B. A., Perper, J., Connolly, J., Bridge, J., Bartle, S., & Rather, C. (1994). Personality Disorder, Personality Traits, Impulsive Violence, and Completed Suicide in Adolescents. *Journal of the American Academy of Child & Adolescent Psychiatry*, *33*(8), 1080-1086. https://doi.org/10.1097/00004583-199410000-00003

Bushman, B. J., & Anderson, C. A. (2001). Is it time to pull the plug on hostile versus instrumental aggression dichotomy? *Psychological Review*, *108*(1), 273. https://doi.org/10.1037/0033-295X.108.1.273

Bushman, B. J., & Huesmann, L. R. (2010). Aggression. Dans *Handbook of social psychology, Vol. 2, 5th ed* (p. 833-863). John Wiley & Sons, Inc. https://doi.org/10.1002/9780470561119.socpsy002023

Camodeca, M., & Goossens, F. A. (2005). Aggression, social cognitions, anger and sadness in bullies and victims. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *46*(2), 186-197. https://doi.org/10.1111/j.1469-7610.2004.00347.x

Coccaro, E. F., Fanning, J. R., Fisher, E., Couture, L., & Lee, R. J. (2017). Social emotional information processing in adults : Development and psychometrics of a computerized video assessment in healthy controls and aggressive individuals. *Psychiatry Research*, *248*, 40-47. https://doi.org/10.1016/j.psychres.2016.11.004

Coccaro, E. F., Noblett, K. L., & McCloskey, M. S. (2009). Attributional and emotional responses to socially ambiguous cues : Validation of a new assessment of social/emotional information processing in healthy adults and impulsive aggressive patients. *Journal of Psychiatric Research*, *43*(10), 915-925. https://doi.org/10.1016/j.jpsychires.2009.01.012

Conroy, S., Burczycka, M., & Savage, L. (2019). *Family violence in Canada : A statistical profile, 2018*. *85*, 62.

Crick, N. R., & Dodge, K. A. (1996). Social Information-Processing Mechanisms in Reactive and Proactive Aggression. *Child Development*, *67*(3), 993-1002. JSTOR. https://doi.org/10.2307/1131875

De Castro, B. O., Veerman, J. W., Koops, W., Bosch, J. D., & Monshouwer, H. J. (2002). Hostile Attribution of Intent and Aggressive Behavior : A Meta-Analysis. *Child Development*, *73*(3), 916-934. https://doi.org/10.1111/1467-8624.00447

DeWall, N. C., Anderson, C. A., & Bushman, B. (2011). The General Aggression Model : Theoretical Extensions to Violence. *Psychology of Violence*, *1*, 245-258. https://doi.org/10.1037/a0023842

DeWall, N. C., Anderson, C. A., & Bushman, B. J. (2013). Aggression. Dans *Handbook of psychology : Personality and social psychology, Vol. 5, 2nd ed* (p. 449-466). John Wiley & Sons, Inc.

Dodge, K. A. (1991). The structure and function of reactive and proactive aggression. Dans *The development and treatment of childhood aggression* (p. 201-218). Lawrence Erlbaum Associates, Inc.

Dodge, K. A. (2006). Translational science in action : Hostile attributional style and the development of aggressive behavior problems. *Development and psychopathology*, *18*(3), 791-814.

Dodge, K. A., Malone, P. S., Lansford, J. E., Sorbring, E., Skinner, A. T., Tapanya, S., Tirado, L. M. U., Zelli, A., Alampay, L. P., Al-Hassan, S. M., Bacchini, D., Bombi, A. S., Bornstein, M. H., Chang, L., Deater-Deckard, K., Di Giunta, L., Oburu, P., & Pastorelli, C. (2015). Hostile attributional bias and aggressive behavior in global context. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(30), 9310-9315. https://doi.org/10.1073/pnas.1418572112

Duberstein, P. R., & Conwell, Y. (1997). Personality Disorders and Completed Suicide : A Methodological and Conceptual Review. *Clinical Psychology: Science and Practice*, *4*(4), 359-376. https://doi.org/10.1111/j.1468-2850.1997.tb00127.x

Dunsieth, N. W., Nelson, E. B., Brusman-Lovins, L. A., Holcomb, J. L., Beckman, D., Welge, J. A., Roby, D., Taylor, P., Soutullo, C. A., & McElroy, S. L. (2004). Psychiatric and legal features of 113 men convicted of sexual offenses. *The Journal of Clinical Psychiatry*, *65*(3), 293-300.

Fazel, S., & Danesh, J. (2002). Serious mental disorder in 23000 prisoners : A systematic review of 62 surveys. *Lancet (London, England)*, *359*(9306), 545-550. https://doi.org/10.1016/S0140-6736(02)07740-1

Fitz, H., & Chang, F. (2019). Language ERPs reflect learning through prediction error propagation. *Cognitive Psychology*, *111*, 15-52. https://doi.org/10.1016/j.cogpsych.2019.03.002

Frank, S. L., Otten, L. J., Galli, G., & Vigliocco, G. (2015). The ERP response to the amount of information conveyed by words in sentences. *Brain and Language*, *140*, 1-11. https://doi.org/10.1016/j.bandl.2014.10.006

Gagnon, J., Aubin, M., Emond, F. C., Derguy, S., Bessette, M., & Jolicoeur, P. (2016). Neural mechanisms underlying attribution of hostile intention in nonaggressive individuals : An ERP study. *International Journal of Psychophysiology*, *110*, 153-162. https://doi.org/10.1016/j.ijpsycho.2016.08.007

Gagnon, J., & Rochat, L. (2017). Relationships between hostile attribution bias, negative urgency, and reactive aggression. *Journal of Individual Differences*, *38*(4), 211-219. https://doi.org/10.1027/1614-0001/a000238

Gardner, B. O., Boccaccini, M. T., Bitting, B. S., & Edens, J. F. (2015). Personality Assessment Inventory scores as predictors of misconduct, recidivism, and violence : A meta-analytic review. *Psychological Assessment*, *27*(2), 534-544. https://doi.org/10.1037/pas0000065

Houston, R. J., Stanford, M. S., Villemarette-Pittman, N. R., Conklin, S. M., & Helfritz, L. E. (2003). Neurobiological Correlates and Clinical Implications of Aggressive Subtypes. *Journal of Forensic Neuropsychology*, *3*(4), 67-87. https://doi.org/10.1300/J151v03n04_05

Hubbard, J. A., Dodge, K. A., Cillessen, A. H., Coie, J. D., & Schwartz, D. (2001). The dyadic nature of social information processing in boys' reactive and proactive aggression. *Journal*

of Personality and Social Psychology, *80*(2), 268-280. https://doi.org/10.1037/0022-3514.80.2.268

Huesmann, L. R. (1988). An information processing model for the development of aggression. *Aggressive Behavior*, *14*(1), 13-24. https://doi.org/10.1002/1098-2337(1988)14:1<13::AID-AB2480140104>3.0.CO;2-J

Huesmann, L. R. (1998). The Role of Social Information Processing and Cognitive Schema in the Acquisition and Maintenance of Habitual Aggressive Behavior. Dans R. G. Geen & E. Donnerstein (Éds.), *Human Aggression* (p. 73-109). Academic Press. https://doi.org/10.1016/B978-012278805-5/50005-5

Huesmann, L. R., & Taylor, L. D. (2006). The Role of Media Violence in Violent Behavior. *Annual Review of Public Health*, *27*(1), 393-415. https://doi.org/10.1146/annurev.publhealth.26.021304.144640

Johnson, B. D. (2010). Multilevel Analysis in the Study of Crime and Justice. Dans A. R. Piquero & D. Weisburd (Éds.), *Handbook of Quantitative Criminology* (p. 615-648). Springer. https://doi.org/10.1007/978-0-387-77650-7_30

Kim, W. S., Jolicoeur, P., & Gagnon, J. (2019). *Les effets non réciproques de l'activation du schéma hostile et non hostile sur le processus d'attribution d'intention chez les individus non agressifs*. https://papyrus.bib.umontreal.ca/xmlui/handle/1866/22736

Krahé, B. (2013). *The social psychology of aggression, 2nd ed* (p. xv, 398). Psychology Press. https://doi.org/10.4324/9781315804521

Krug, E. G. (Éd.). (2002). *World report on violence and health*.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting : Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*, 621-647. https://doi.org/10.1146/annurev.psych.093008.131123

Lansford, J. E., Malone, P. S., Dodge, K. A., Crozier, J. C., Pettit, G. S., & Bates, J. E. (2006). A 12-Year Prospective Study of Patterns of Social Information Processing Problems and Externalizing Behaviors. *Journal of abnormal child psychology*, *34*(5), 715-724. https://doi.org/10.1007/s10802-006-9057-4

Leuthold, H., Filik, R., Murphy, K., & Mackenzie, I. G. (2012). The on-line processing of socio-emotional information in prototypical scenarios : Inferences from brain potentials. *Social Cognitive and Affective Neuroscience*, *7*(4), 457-466. https://doi.org/10.1093/scan/nsr029

Lobbestael, J., Cima, M., & Arntz, A. (2013). The Relationship Between Adult Reactive and Proactive Aggression, Hostile Interpretation Bias, and Antisocial Personality Disorder. *Journal of Personality Disorders*, *27*(1), 53-66. https://doi.org/10.1521/pedi.2013.27.1.53

Lösel, F., Bliesener, T., & Bender, D. (2007). Social Information Processing, Experiences of Aggression in Social Contexts, and Aggressive Behavior in Adolescents. *Criminal Justice and Behavior*, *34*(3), 330-347. https://doi.org/10.1177/0093854806295833

MacBrayer, E. K., Milich, R., & Hundley, M. (2003). Attributional Biases in Aggressive Children and Their Mothers. *Journal of Abnormal Psychology*, *112*(4), 698. https://doi.org/10.1037/0021-843X.112.4.598

Matthews, B. A., & Norris, F. H. (2002). When Is Believing "Seeing"? Hostile Attribution Bias as a Function of Self-Reported Aggression1. *Journal of Applied Social Psychology*, *32*(1), 1-31. https://doi.org/10.1111/j.1559-1816.2002.tb01418.x

Merk, W., De Castro, B. O., Koops, W., & Matthys, W. (2005). The distinction between reactive and proactive aggression : Utility for theory, diagnosis and treatment? *European Journal of Developmental Psychology*, *2*(2), 197-220. https://doi.org/10.1080/17405620444000300

Miller, J. D., & Lynam, D. R. (2006). Reactive and proactive aggression : Similarities and differences. *Personality and Individual Differences*, *41*(8), 1469-1480. https://doi.org/10.1016/j.paid.2006.06.004

Moran, P. (1999). The epidemiology of antisocial personality disorder. *Social Psychiatry and Psychiatric Epidemiology*, *34*(5), 231-242. https://doi.org/10.1007/s001270050138

Moreau, G. (2019, juillet 22). *Police-reported crime statistics in Canada, 2018*. https://www150.statcan.gc.ca/n1/pub/85-002-x/2019001/article/00013-eng.htm

Murray-Close, D., Ostrov, J. M., Nelson, D. A., Crick, N. R., & Coccaro, E. F. (2010). Proactive, reactive, and romantic relational aggression in adulthood : Measurement, predictive validity, gender differences, and association with Intermittent Explosive Disorder. *Journal of Psychiatric Research*, *44*(6), 393-404. https://doi.org/10.1016/j.jpsychires.2009.09.005

Ostrov, J. M., & Houston, R. J. (2008). The Utility of Forms and Functions of Aggression in Emerging Adulthood : Association with Personality Disorder Symptomatology. *Journal of Youth and Adolescence*, *37*(9), 1147-1158. https://doi.org/10.1007/s10964-008-9289-4

Poulin, F., & Boivin, M. (2000). Reactive and proactive aggression : Evidence of a two-factor model. *Psychological Assessment*, *12*(2), 115-122. https://doi.org/10.1037/1040-3590.12.2.115

Raine, A., Dodge, K., Loeber, R., Gatzke-Kopp, L., Lynam, D., Reynolds, C., Stouthamer-Loeber, M., & Liu, J. (2006). The Reactive–Proactive Aggression Questionnaire : Differential Correlates of Reactive and Proactive Aggression in Adolescent Boys. *Aggressive behavior*, *32*(2), 159-171. https://doi.org/10.1002/ab.20115

Raine, A., Meloy, J. R., Bihrle, S., Stoddard, J., Lacasse, L., & Buchsbaum, M. S. (1998). Reduced prefrontal and increased subcortical brain functioning assessed using positron emission tomography in predatory and affective murderers. *Behavioral Sciences & the Law*, *16*(3), 319-332. https://doi.org/10.1002/(SICI)1099-0798(199822)16:3<319::AID-BSL311>3.0.CO;2-G

Ramirez, J., & Andreu, J. (2006). Aggression, and some related psychological constructs (anger, hostility and impulsivity) Some comments from a research project. *Neuroscience and Biobehavioural Reviews*, *30*, 276-291.

Ross, J. M., & Babcock, J. C. (2009). Proactive and Reactive Violence among Intimate Partner Violent Men Diagnosed with Antisocial and Borderline Personality Disorder. *Journal of Family Violence*, *24*(8), 607-617. https://doi.org/10.1007/s10896-009-9259-y

Smeijers, D., Rinck, M., Bulten, E., Heuvel, T. van den, & Verkes, R.-J. (2017). Generalized hostile interpretation bias regarding facial expressions : Characteristic of pathological aggressive behavior. *Aggressive Behavior*, *43*(4), 386-397. https://doi.org/10.1002/ab.21697

Tweed, R., & Dutton, D. (1998). A Comparison of Impulsive and Instrumental Subgroups of Batterers. *Violence and victims*, *13*, 217-230. https://doi.org/10.1891/0886-6708.13.3.217

Van Overwalle, F., Van den Eede, S., Baetens, K., & Vandekerckhove, M. (2009). Trait inferences in goal-directed behavior : ERP timing and localization under spontaneous and intentional processing. *Social Cognitive and Affective Neuroscience*, *4*(2), 177-190. https://doi.org/10.1093/scan/nsp003

Walters, G. D. (2007). Measuring Proactive and Reactive Criminal Thinking With the PICTS : Correlations With Outcome Expectancies and Hostile Attribution Biases. *Journal of Interpersonal Violence*, *22*(4), 371-385. https://doi.org/10.1177/0886260506296988

Warren, J. I., Burnette, M., South, S. C., Chauhan, P., Bale, R., & Friend, R. (2002). Personality disorders and violence among female prison inmates. *The Journal of the American Academy of Psychiatry and the Law*, *30*(4), 502-509.

World Health Organization. (2017). *10 facts about violence prevention*. WHO; World Health Organization. http://www.who.int/features/factfiles/violence/en/

9. ANNEXE A – FIGURES ADDITIONNELLES

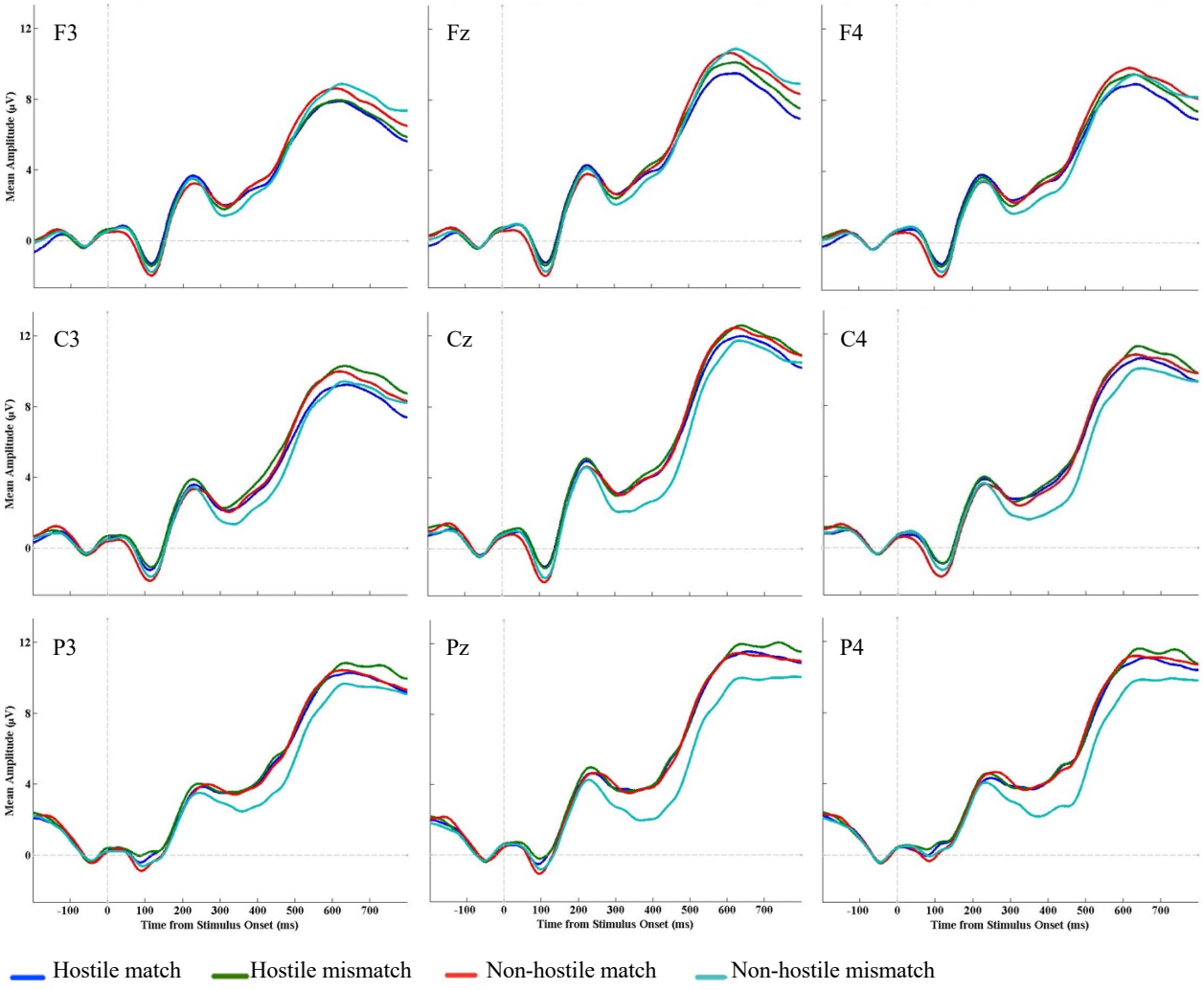

***Figure 6*. Grand average ERPs recorded when presenting hostile match, hostile mismatch, non-hostile match and non-hostile mismatch target words at electrodes of anterior, central and posterior sites.**

F = anterior; C = central ; P = posterior.

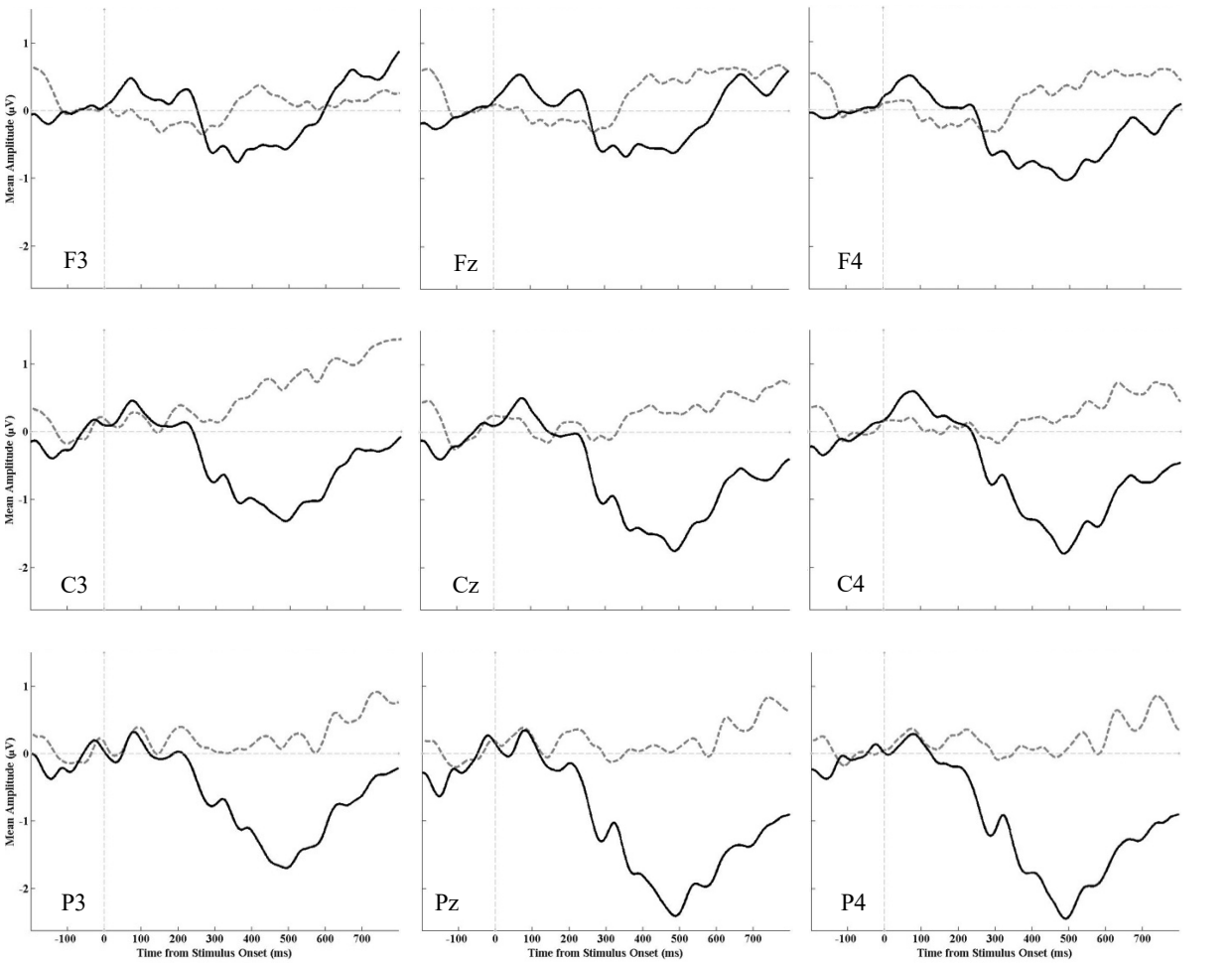

Figure 7. **Difference between the mismatch and match conditions of the grand ERP averages obtained after presentation of the hostile or non-hostile target word at electrodes of anterior, central and posterior sites.**

F = anterior; C = central; P = posterior.