Université de Paris

École doctorale
« Cognition, Comportements, Conduites Humaines » (ED 261)

Laboratoire de Psychopathologie et Processus de Santé (EA 4057)

The multimodal assessment of adult attachment: development of a biometric test

Par Federico Parra Silva

Thèse de doctorat de Psychologie

Dirigée par Susana Tereno

Présentée et soutenue publiquement le 25 novembre 2019 devant un jury composé de :

- Antoine Guédeney, PU-PH, Université Paris Diderot, Président du jury
- Mario Mikulincer, Professor, IDC Herzliya, Rapporteur
- Manuela Verissimo, Professor, ISPA – Instituto Universitário William James Center for Research, Rapporteur
- Stefan Scherer, Assistant Professor, Institute for Creative Technologies, University of Southern California, Membre du jury
- Yannick Benezeth, MCU, Université de Bourgogne Franche-Compté, Membre du jury
- Susana Tereno, PAST-HDR, Université Paris Descartes, la Directrice.

Except where otherwise noted, this is work licensed under
https://creativecommons.org/licenses/by-nc-nd/3.0/fr/

Title: The multimodal assessment of adult attachment: development of a biometric test

Abstract:

Context: Psychometric tests are the cornerstone of psychological research and precious tools for clinical practice. One of the essential psychological dimensions to measure in mental health is the quality of adult attachment, which reveals the degree to which we are capable of establishing healthy intimate and durable bonds with our loved ones. However, existing assessments for adult attachment suffer from several problems: they are costly, time-consuming, and in the case of interviews, somewhat subjective, or in the case of questionnaires, they are controversial because of possible self-appreciation biases.

Objectives: In this context, our study proposes to create and validate the Biometric Attachment Test (BAT). Inspired by projective tests, the BAT uses music clips and pictures to evoke attachment-related situations and asks participants to express their feelings towards the stimuli. But the difference with already existing tests is that the BAT will analyze the participant's behavioral, linguistic, and physiological responses automatically, using state-of-the-art computer algorithms.

Methodology: Our computer test software will be capable of presenting the music and picture stimuli and will be able to analyze in real-time the facial expressions, head movements, gaze direction, voice stress, heart rate variability, and language used as the participant takes the test. Powered by artificial intelligence, our goal is that the BAT can integrate and interpret all this different data to deduct the quality of attachment of the person, in a fast, economical and holistic fashion, taking into account several dimensions of the response, all automatically and objectively.

Results: A functional version of the BAT software was created. Three validation experiments were conducted, exploring the test's content validity, convergent validity, cross-cultural validity, and test-retest reliability. Overall, these experiments showed that the BAT is a valid measure for adult attachment. The software will soon be made available for other researchers to try and replicate these findings.

Keywords: adult attachment, artificial intelligence, biometrics, psychometrics

Titre : L'évaluation multimodale de l'attachement chez l'adulte : développement d'un test biométrique

Résumé :

Contexte : Les tests psychométriques sont la pierre angulaire de la recherche en psychologie, et des outils précieux dans la clinique. Une des dimensions psychologiques importante à mesurer en santé mentale est la qualité de l'attachement chez l'adulte, qui révèle le degré dans lequel nous sommes capables d'établir des liens intimes et durables avec nos proches. Cependant, les tests existants pour mesurer l'attachement chez l'adulte posent un certain nombre de problèmes : soit ils sont très coûteux, chronophages et, dans le cas des interviews, un peu subjectifs, soit ils font polémique, dans le cas des autoquestionnaires, en raison d'éventuels biais d'auto-appréciation.

Objectives : Dans ce contexte, notre étude propose de créer et de valider le Biometric Attachment Test (BAT). Inspiré des tests du type projectifs, le BAT utilise des clips de musique et des images pour évoquer des situations liées aux relations intimes, et demande aux participants d'exprimer leurs sentiments vis-à-vis de ces stimuli. Mais à la différence des tests déjà existants, le BAT analysera automatiquement les réponses comportementales, linguistiques et physiologiques du participant à ces stimuli, à l'aide de puissants algorithmes informatiques.

Méthodologie : Ainsi, notre test par ordinateur sera capable à lui seul de présenter les images et musiques ainsi que d'analyser en temps réel les expressions du visage, les mouvements de la tête, la direction des yeux, la tension dans la voix, la variabilité du rythme cardiaque, et les mots utilisés par le participant pendant qu'il passe le test. Aidé par l'intelligence artificielle, notre objectif est que le BAT soit capable ensuite d'intégrer ces différentes données pour arriver à déduire la qualité d'attachement de la personne, et ceci d'une manière rapide, économique et plus holistique, prenant en compte plusieurs dimensions de la réponse, le tout de manière automatique et objective.

Résultats : Une version fonctionnelle du logiciel BAT a été créé. Le test a été validé à trois reprises, y compris sa validité de contenu, sa validité convergente, sa validité transculturelle, et test-retest. Ces validations ont montré que le BAT est une mesure valide d'attachement chez l'adulte. Le logiciel sera bientôt mis à disposition des chercheurs pour qu'ils puissent essayer de reproduire ces résultats.

Mots clefs : attachement adulte, intelligence artificielle, biometrie, psychometrie

1 Acknowledgements

I want to thank my PhD advisor, Susana Tereno, for her trust and encouragement at the most difficult of times: thank you. I want to thank my partner, Pauline Bandelier, for having been there for me day in and day out throughout three years of hard work, and I want to thank our baby girl in the belly, who I await with impatience. I want to thank Stefan Scherer for his mentoring in the computer sciences and his continuous advice. I want to thank Yannick Benezeth for having trusted me with his outstanding method for detecting heart rate from video. I want to thank Carol George for having taught me the Adult Attachment Projective and so many other things related to adult attachment. I want to thank David Elliott, for having been my mentor, colleague, and friend. I want to thank my family in Argentina for their support: Seba and mom, thank you. I want to thank the team at Ville-Evrard for having supported me throughout these years. I want to thank the Attachment Seminar of Nicole and Antoine Guedeney, as well as their DU, which helped me cook some of the ideas hereby exposed. And I want to thank dear friends and colleagues that supported me greatly throughout this project: Gwenaëlle Persiaux, Khalid Kalalou, and Wilfrid Nkodia. Finally, a special thanks goes out to “Dr. Dan” (Daniel P. Brown), for starting the fire of attachment in me. Thank you all.

2 List of main abbreviations

AAI; Adult Attachment Interview

AAP; Adult Attachment Projective

AAQ; Adult Attachment Questionnaire

BAT; Biometric Attachment Test

HRV; Heart rate variability

RSQ; Relationship Scales Questionnaire

SSP; Strange Situation Procedure

DERS; Difficulties in Emotion Regulation Scale

ACE; Adverse Childhood Experience questionnaire

RSES; Rosenberg Self-esteem scale

GSE; General Self-efficacy scale

Summary

3 Résume en Français

3.1 Introduction

La théorie de l'attachement est née des travaux du psychiatre britannique John Bowlby et de la psychologue américano-canadienne Mary Ainsworth. Bowlby a théorisé que les chances de survie des gènes humains avaient augmenté grâce à la sélection naturelle de comportements augmentant la proximité et les liens entre les nourrissons et leurs soignants, ce qui a accru la probabilité de protection des enfants (Bowlby, 1969; Cassidy & Shaver, 2016).

Une figure d'attachement est la personne qui interagit le plus régulièrement et le plus fréquemment avec un enfant. Bien que la théorie postule une prédisposition génétiquement déterminée chez les soignants à réagir avec le plus grand soin aux comportements de quête de proximité des enfants, ceux-ci ne sont parfois pas en mesure de le faire de manière appropriée et cohérente (Ainsworth, 1985).

La qualité, ainsi que le résultat, de ces interactions précoces répétées entre le comportement de quête de proximité des enfants et la réponse des figures d'attachement laissent une marque physiologique, comportementale et cognitive durable chez l'enfant en développement (de Wolff & van IJzendoorn, 1997; Sroufe, Coffino, & Carlson, 2010; Fraley & Roisman, 2015; Lorber & Egeland, 2009).

La théorie de l'attachement a joué un rôle essentiel dans deux des études longitudinales en psychologie les plus importantes et les plus rigoureuses à ce jour: The Minnesota Longitudinal Study of Risk and Adaptation (MLSRA; (Sroufe, Coffino, & Carlson, 2010)), qui a étudié 180 enfants américains de familles pauvres dès avant la naissance (pendant la grossesse de la mère) jusqu'à 35 ans; et l'étude NICHD Study of Early Child Care and Youth Development (SECCYD, Institut national de la santé), qui a suivi 1 300 enfants de famillesordinaires dans plusieurs endroits des États-Unis, de la naissance jusqu'à 19 ans. Ces études "monstres" ont révélé que la qualité des interactions précoces dans l'enfance prédisent la sécurité de l'attachement chez l'enfant (de Wolff & van IJzendoorn, 1997), ce qui prédit partiellement la psychopathologie chez l'adulte (Sroufe, Coffino, & Carlson, 2010; Carlson, 1998). Par exemple, dans l'étude MLSRA, l'association entre l'attachement désorganisé dans l'enfance et la dépression à 18 ans chez les hommes était de $r = 0{,}4$, après contrôle de tous les facteurs de confusion (Sroufe, Coffino, & Carlson, 2010). Des études transversales avec des adultes semblent avoir confirmé cette découverte développementale, trouvant que l'attachement insécure est lié à la dépression (Cantazaro & Wei, 2010), au syndrome de stress post-traumatique (Ein-Dor, Doron, Solomon, Mikulincer, & Shaver, 2010) et au trouble de stress post-traumatique complexe (Parra, George, Kalalou, & Januel, 2017), à la personnalité limite (Levy, Meehan, Weber, Reynoso, & Clarkin, 2005) et d'autres troubles de la personnalité (Crawford et al., 2007), aux troubles de l'alimentation (Illing, Tasca, Balfour, & Bissada, 2010), au trouble obsessionnel-compulsif (Doron et al., 2011), à l'idéation suicidaire (Gormley & McNiel, 2009) et même à la schizophrénie (Mikulincer & Shaver, 2007). Au contraire, et renforçant davantage la notion de lien entre l'attachement et la psychopathologie, les expériences d'attachement positives à l'âge adulte, qu'elles soient naturelles ou le résultat d'une intervention thérapeutique méthodique telle que l'Ideal Parent Figure Method de Brown et Elliott, se sont avérées augmenter la sécurité d'attachement, ce qui améliore la santé mentale (Brown & Elliott, 2016; Parnell & Siegel, 2015; Mikulincer & Shaver, 2015).

Les études longitudinales susmentionnées ont également montré que les compétences en matière de développement, essentielles au maintien de la santé mentale et à la gestion des troubles de la santé mentale, telles que la régulation émotionnelle, les aptitudes sociales ou les capacités cognitives, sont associées et interdépendantes de l'attachement tout au long de la vie (Bergman, Sarkar, Glover, & O'Connor, 2010; de Wolff & van IJzendoorn, 1997; Sroufe, Coffino, & Carlson, 2010).

L'attachement des adultes a également un impact sur la capacité des adultes à créer et à maintenir des relations intimes saines à long terme (Feeney, 2016).

Nous croyons que l'attachement est une dimension vitale de la santé mentale des adultes et que mesurer correctement l'attachement des adultes devrait être une priorité dans tous les contextes liés à la santé mentale.

Depuis 1984 (conclusions de George C, Kaplan N et Main M, données inédites, 1984), divers instruments validés d'évaluation de l'attachement chez l'adulte se sont développés parallèlement dans les domaines de la psychologie sociale et de la psychologie du développement.

La psychologie sociale a stimulé le développement de multiples mesures d'auto-évaluation sous forme de questionnaires. Des exemples de cette approche sont le Adult Attachment Questionnaire (AAQ; (Simpson, Rholes, & Phillips, 1996)), le Adult Attachment Scale (AAS; (Collins & Read, 1990)) et le Relationships Scale Questionnaire (RSQ, (Griffin & Bartholomew, 1994)).

La psychologie du développement, en revanche, s'est appuyée sur une variété de méthodes d'entretien semi-structurées au sens large, à commencer par l'AAI, qui était la première et qui est toujours considérée comme le «standard de référence» (Ravitz, Maunder, Hunter, Sthankiya, & Lancee, 2010). Des exemples plus récents sont le Patient Attachment

Coding System (PACS; (Talia, Miller-Bottome, & Daniel, 2015)) et le Adult Attachment Projective (AAP; (George & West, 2012)).

Les questionnaires d'attachement sont des mesures d'auto-évaluation. En tant que tels, ils sont sujets à des distorsions liées à l'auto-évaluation qui ont été bien décrits dans la littérature (par exemple (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003; van de Mortel, 2008; Hunt, Auriemma, & Cashaw, 2003; Niedenthal, Brauer, Robin, & Åse H. Innes-Ker, 2002)).

En termes de validité de construction, aucune association longitudinale n'a été démontrée entre l'attachement dans l'enfance mesuré par exemple avec la Strange Situation Procedure (SSP; (Ainsworth & Bell, 1970)) et l'attachement à l'âge adulte mesuré à l'aide de questionnaires (KE Grossmann, 2005). De plus, presque aucune validité convergente n'a été trouvée entre les questionnaires et les évaluations de l'attachement d'un adulte basées sur des entretiens, ce qui ajoute à la critique de la validité de construction (Roisman et al., 2007). Cependant, les mesures d'auto-évaluation sont corrélées entre elles et avec des échelles mesurant des concepts théoriquement liés tels que la régulation des émotions ou l'estime de soi (Ravitz, Maunder, Hunter, Sthankiya, & Lancee, 2010).

Les questionnaires d'attachement adulte ne sont pas holistiques : ils ne peuvent pas saisir les nombreuses dimensions de l'expression de l'attachement adulte, telles que ses traits physiologiques, ses représentations cognitives et ses caractéristiques comportementales multiformes (par exemple, la voix qui change lorsqu'on est confronté à un stress relationnel).

Et pourtant, les questionnaires d'attachement adulte sont faciles, économiques et rapides à administrer et à noter. L'administration peut être effectuée à distance et un scoring automatique est possible.

Les évaluations de l'attachement adulte fondées sur des entretiens reposent sur une forme d' entretien semi-structuré, qui est ensuite transcrit et noté par un juge qualifié, qui a suivi une formation approfondie dans le cadre d'une tradition de notation normalisée spécifique.

Cette forme d'évaluation est plus proche des évaluations d'attachement chez l'enfant, dans la mesure où elles s'appuient également sur des experts tiers pour la notation et la classification. Cependant, au cours des évaluations d'attachement dans l'enfance telles que le SSP, les experts en notation observent le comportement en général, sous plusieurs dimensions telles que le mouvement, les expressions faciales, les réponses vocales, etc., alors que dans les méthodes d'entretien, seul le langage transcrit est analysé.

Contrairement aux questionnaires, les méthodes d'entretien sont difficiles, coûteuses et longues à administrer et à noter. Elles ajoutent des étapes supplémentaires au processus, à savoir la transcription manuelle et le codage de l'interview. Une formation est requise pour l'administration et la notation. Ce processus est coûteux et prend du temps. En particulier, lecoût financier prohibitif ne les rends pas accessible à à la recherche à faible budgetet rend leur mise en œuvre dans la santé mentale publique inimaginable. Les méthodes d'entretien sont affectées par un problème supplémentaire : la subjectivité inhérente à tout juge expert humain (Cassidy, Sherman, & Jones, 2012). Enfin, contrairement aux questionnaires, les méthodes basées sur des entretiens ne peuvent pas être administrées à distance, ce qui limite leur application, par exemple dans la recherche numérique.

En termes de validité de construction, les mesures basées sur les entretiens ont été considérées depuis toujours comme la meilleure pratique dans le domaine de l'attachement adulte. En particulier, l'AAI a toujours montré un lien entre les modèles d'attachement des parents et de leurs enfants, ce qui est considéré comme une preuve solide de sa validité.

Le BAT est un test basé sur l'exposition dont la conception psychométrique a été fortement influencée par trois mesures d'attachement antérieures: le Separation Anxiety Test de Bowlby (SAT; (Klagsbrun & Bowlby, 1976)), le SSP précédemment mentionné et le AAP.

La SAT a été le premier test d'attachement (Klagsbrun & Bowlby, 1976). Il a été conçu par Klagsbrun et Bowlby comme un test d'attachement projectif destiné aux enfants âgés de 4 à 7 ans et consistant en un ensemble de 6 images illustrant des situations dans lesquelles un enfant, séparé de sa famille, doit faire face à la situation sans l'aide de ses parents. On demande à l'enfant testé d'imaginer les sentiments du protagoniste et de prédire son comportement. Leurs réponses transcrites sont ensuite notées et classées selon différentes sous-échelles telles que la sécurité de l'attachement, qui évalue la capacité de l'enfant à exprimer la vulnérabilité et les besoins du personnage. Ce test repose sur le principe d'identification empathique (Zillmann, 1995). Le participant s'identifie avec le personnage de l'image, comme nous le faisons avec les personnages de films, puis fait l'expérience par procuration Ddes expériences du personnage, qui dans ce cas visent à l'activation de l'attachement. Pour le BAT, nous avons repris l'idée du SAT d'utiliser des photos réelles représentant de vraies personnes dans des situations clairement liées à l'attachement.

Le SSP est un protocole d'observation structuré créé par Mary Ainsworth pour évaluer l'attachement chez les enfants âgés de 9 à 18 mois (Ainsworth & Bell, 1970). Pendant 20 minutes, l'enfant subit une série de séparations et de retrouvailles avec sa mère, tandis qu'il est également exposé à l'arrivée et à la présence d'un étranger. Le comportement de l'enfant est enregistré en vidéo, puis systématiquement analysé à l'aide d'un protocole normalisé de notation et de classification. L'analyse est effectuée par des juges experts hautement qualifiés et prend en compte plusieurs aspects du comportement de l'enfant, tels que les expressions faciales, les mouvements du corps et de la tête, la direction du regard et les expressions vocales.

Le SSP nous a inspiré de deux manières pour élaborer le BAT. Premièrement, il nous a incité à créer une situation de test qui provoquerait une activation de l'attachement (stress) chez le participant, mais qui lui offrirait également la possibilité d'être rassuré et apaisé, tout en surveillant attentivement dans quelle mesure le système d'attachement se désactive avec succès (relaxation). Deuxièmement, le SSP nous a donné l'idée de créer un test prenant en compte toute une gamme de dimensions comportementales observables, y compris la parole, mais aussi le ton de la voix, les expressions faciales, les mouvements de la tête et la direction du regard, non seulement comme des canaux d'information distincts, mais aussi en les les considérant sous forme combinées.

Le AAP est le deuxième test d'attachement pour adultes créé par Carol George, auteur du AAI (George & West, 2001). Le AAP est une mesure projective basée sur l'exposition à un ensemble de dessins en noir et blanc décrivant diverses situations connues pour activer le système d'attachement: séparation, perte, solitude et menace physique (George & West, 2012). Les participants sont invités à raconter une courte histoire sur les personnages des images, qui sera ensuite transcrite et analysée. Une classification d'attachement ainsi que des scores continus seront alors obtenus (George & West, 2012).

Dans la conception du BAT, nous avons emprunté plusieurs concepts au AAP de George. L'un de ces concepts est l'idée d'incorporer au test l'imagerie liée aux traumatismes de l'enfance afin de pouvoir mieux distinguer le modèle d'attachement non résolu des autres et de mieux rendre compte de la désorganisation de l'esprit.

Une autre idée empruntée à l'AAP est la notion qu'à l'âge adulte, diverses situations peuvent activer le système d'attachement: perte d'un être cher, séparation d'un partenaire, souvenir d'une enfance solitaire ou souvenir d'un traumatisme dans l'enfance. Nous avons tenté de reproduire cela en concevant une pluralité de thèmes liés à l'attachement devant être invoqués par le BAT.

Les tests projectifs et basés sur l'exposition reposent généralement sur un ensemble fixe de stimuli sélectionnés par les auteurs, à l' instar du AAP (Tuber, 2014; George & West, 2012). Nous formulons les critiques suivantes à cette approche: premièrement, les stimuli peuvent éventuellement tomber dans le domaine public, comme dans le cas du Rorschach (Schultz & Loving, 2012), ce qui pourrait nuire à l'efficacité du test en raison des effets d'amorçage. Deuxièmement, les études longitudinales telles que les essais cliniques exigent que les participants soient testés plusieurs fois à l'aide des mêmes instruments. Si les stimuli sont toujours les mêmes, cela peut également entraîner des effets d'amorçage. De plus, nous pensons que les stimuli devraient être sélectionnés en fonction de l'opinion de la population générale qui en fera l'expérience.

Notre BAT innove en introduisant le concept de thèmes : des espaces réservés pour les stimuli.

Un thème est un récit simple qui décrit une situation spécifique devant être évoquée par un stimulus, avec un objectif spécifique. Par exemple, en termes d'attachement adulte, un thème pourrait être «la perte d'un proche», son objectif étant d'activer le système d'attachement (c'est-à-dire de provoquer une détresse liée à l'attachement).

L'utilisation de thèmes peut résoudre les problèmes susmentionnés liés aux conceptions de test à stimuli fixe.

Les thèmes ont été inspirés par les principes fondamentaux du SAT, du AAP, du SSP et de la théorie de l'attachement. Au total, 14 thèmes ont été créés dans le cadre de ce travail.

Ensuite, un ensemble de critères objectifs et subjectifs ont été développés pour chacun des thèmes. Les critères objectifs étaient directement dérivés des récits des thèmes : par exemple, pour qu'un stimulus soit approprié pour représenter le thème de la « dyade mère-enfant bien accordés », il devrait y avoir une mère et un enfant dans l'image. Les critères subjectifs sont des notions qui nécessitent des jugements plus complexes : par exemple, pour qu'un stimulus soit approprié pour représenter le thème de la « dyade mère-enfant bien accordés », l'enfant et la mère doivent sembler s'accorder émotionnellement et, grâce à cette harmonisation, les deux devraient sembler relativement détendus.

Premièrement, nous avons utilisé les critères objectifs pour présélectionner les stimuli.

Nous avons effectué une recherche exhaustive en utilisant ces critères dans trois grandes bases de données d'images conçues pour l'étude de l'émotion: le Nencki Affective Picture System (NAPS; (Marchewka, Łukasz Żurawski, Jednoróg, & Grabowska, 2013)), le International Affective Picture System (IAPS; (Lang, Bradley, & Cuthbert, n.d.)) et le Geneva Affective Picture Database (GAPED ; (Dan-Glauser & Scherer, 2011)). Nous nous sommes retrouvés avec 140 images présélectionnées, soit environ 12 images présélectionnées par thème (les thèmes 9 et 11 n'utilisent pas d'images).

Nous avons ensuite décidé de lancer une enquête Web anonyme utilisant les services de SurveyGizmo. L'enquête a introduit au hasard chacune des 140 images présélectionnées, accompagnées de curseurs que les participants pouvaient ajuster à droite ou à gauche, signalant ainsi un jugement sur un critère subjectif spécifique pour le thème auquel l'image était destinée. Dans le cas du "couple bien accordé" (thème 7), par exemple, l'un des curseurs disponibles dans l'enquête permettait de noter le niveau de connexion perçu dans le couple, la mesure dans laquelle une telle connexion était plutôt érotique ou tendre, et si l'image était authentique ou scriptée (ce dernier curseur était présent pour toutes les images).

Au total, 520 participants (femmes = 72,3%, 376/520, hommes = 27,7%, 144/520), de différents âges (moyenne 37,53, SD 10,87) et pays (Argentine, France et États-Unis) ont répondu.

À notre connaissance, le BAT est le premier test psychométrique à utiliser la musique comme stimuli. La musique a été incluse en raison de sa capacité à déclencher des émotions et des expériences émotionnelles fortes, en surmontant les défenses psychologiques (Baumgartner, Lutz, Schmidt, & Jäncke, 2006).

La sélection de musique était facile et n'a pas nécessité de sondage. Certains thèmes ont été conçus pour transmettre des émotions pures (ex. thème 9, « tristesse pure »). Au total, 25 courts extraits musicaux ont été sélectionnés depuis une librairie musicale conçue pour susciter des émotions, qui fournissait déjà des scores par émotion (URL: https://www.jyu.fi/hytk/fi/laitokset/mutku/en/research/projects2/past-projects/coe/materials/emotion/soundtracks/Index).Nous avons simplement choisi les clips de musique correspondants le mieux à chaque thème et son émotion associée.

L'un de nos objectifs avec le BAT était de créer un test concis afin de minimiser la charge pesant sur les participants, en particulier les patients cliniques, ainsi que de réduire la charge de temps pour les chercheurs. Nous avons décidé d'exposer les participants à chaque thème pendant 15 secondes, après quoi nous leur demandions de commenter ce qu'ils avaientvécu, pendant 20 secondes. Nous avons décidé d'utiliser le langage de l'émotion, posant la simple question : "Qu'avez-vous ressenti ?" après chaque stimulus. Enfin, nous avons choisi d'offrir une pause de 5 secondes avant de passer au stimulus suivant, afin que la personne puisse se récupérer. La durée totale de ce test est de 9 minutes et 30 secondes.

La recherche de marqueurs psychophysiologiques et comportementaux liés aux troubles psychologiques gagne du terrain dans la psychiatrie traditionnelle (Cummins et al., 2015; Scherer, Gale, Gratch, Rizzo, & Morency, 2016), dans le cadre d'une quête visant à fournir des évaluations cliniques plus objectives et plus précises aux patients.

L'Institut national américain de la santé mentale a publié une déclaration en 2013 (URL: https://www.nimh.nih.gov/about/directors/thomas-insel/blog/2013/transforming-diagnosis.shtml) dans laquelle il explicite son ambition de favoriser des méthodologies de diagnostic plus objectives et plus précises .

Plusieurs tentatives pour s'attaquer à ce problème ont été soulevées par l'informatique. Dans une revue récente, Cummins et al. (Cummins et al., 2015) ont examiné l'état actuel des connaissances en matière de détection automatique de la dépression et des tendances suicidaires via l'analyse de la parole et de ses caractéristiques acoustiques paralinguistiques. Scherer et al. ont décrit, en 2013, un ensemble de caractéristiques comportementales non verbales et audiovisuelles automatiquement extraites, utiles pour l'identification de la dépression, de l'anxiété et du ESPT (Scherer et al., 2013).

Le lien manquant entre la mesure objective de marqueurs biométriques ou comportementaux singuliers, et la capacité recherchée à offrir des diagnostics plus précis, repose sur l'utilisation d'algorithmes d'apprentissage automatique capables de fusionner plusieurs modalités de données à la fois (Novak, Mihelj, & Munih, 2012). Cela permet de découvrir des patterns complexes de données multimodales pouvant servir à l'évaluation automatique de troubles mentaux spécifiques.

Dans certaines études récentes, de tels systèmes multimodaux ont dépassé la performance humaine en matière de détection d'indicateurs de stress post-traumatique (Lucas, Gratch, Scherer, Boberg, & Stratou, 2015).

En 2012, Sharma et ses collègues ont mené une enquête sur les mesures physiologiques et comportementales disponibles pour détecter le stress (Sharma & Gedeon, 2012). Le stress n'est pas l'attachement, mais être capable de détecter avec précision si un stimulus lié à l'attachement augmente le stress ou induit une relaxation, est l'énigme au cœur du BAT. Le BAT utilise huit des douze meilleures mesures de stress selon le classement de Sharma et al. : variabilité de la fréquence cardiaque, activité électro dermique, analyse de la voix, direction du regard, expressions du visage, température de la peau, pouls sanguin, et clignement des yeux.

Pour intégrer toutes ces différentes sources de données dans un score d'attachement cohérent, le BAT utilise la fusion multimodale, une technique d'apprentissage automatique qui permet de créer des modèles non linéaires pour une intégration optimale des données provenant de différents capteurs ou modalités afin de prédire un résultat donné (Poria, Cambria, Bajpai, & Hussain, 2017). On dit qu'il imite le fonctionnement du cerveau humain, en combinant différentes sources d'informations par rapport au même phénomène, de sorte que chaque canal d'information ajoute des détails et des perspectives différentes, permettant ainsi de mieux juger le phénomène que ce qui pourrait être fait si basé sur des informations provenant d'un seul canal.

3.2 Articles publiés / soumis

Évaluation multimodale de la sécurité de l'attachement chez l'adulte : mise au point du Biometric Attachment Test

Notre premier article publié sur le BAT répondait à un appel à contributions du Journal of Medical Internet Research (facteur d'impact 4.94), pour un numéro spécial consacré à l'informatique et à la santé mentale.

Dans cet article, nous avons exploré les relations entre les scores bruts de toutes les modalités du BAT (expressions faciales, EDA, par exemple) et le score de sécurité de l'attachement obtenu par trois mesures validées d'attachement adulte, dont deux basées sur des entretiens. Nous avons également testé l'hypothèse selon laquelle différents ensembles de stimuli conçus à l'aide de notre procédure standardisée et notre enquête produiraient des réponses similaires chez les participants.

Au total, 46 corrélations cohérentes avec la théorie ont été trouvées au cours de l'exploration (sur 65 corrélations significatives totales). Des analyses répétées ANOVA effectuées individuellement sur chacune des caractéristiques cohérentes avec la théorie ont révélé que seules 7 des 46 caractéristiques (15%) présentaient des valeurs significativement différentes entre les réponses à trois ensembles de stimuli différents.

Nous avons pu valider ainsi deux des hypothèses fondamentales de l'instrument : sa capacité à mesurer la sécurité de l'attachement, et la viabilité de l'utilisation de thèmes comme substituts pour des stimuli rotatifs.

Développement et évaluation interculturelle d'un algorithme de notation pour le Biometric Attachment Test : relever les défis de la fusion multimodale avec de « petites données»

Cet article a été publié en 2019 dans la revue Transactions on Affective Computing (facteur d'impact 6.28), une revue interdisciplinaire dédiée à l'intersection de la psychologie, en particulier l'étude de l'émotion humaine, et de l'informatique.

Notre article le plus technique à ce jour, il décrit en détail une combinaison originale de méthodes d'apprentissage automatique que nous avons mises en place afin que le BAT puisse apprendre à évaluer l'attachement adulte. Le document traitait d'un problème commun concernant l'utilisation de l'apprentissage automatique avec des échantillons de psychologie, à savoir comment appliquer des algorithmes conçus pour traiter des quantités énormes de données sur les petits échantillons communs dans la recherche en psychologie.

Mais le document propose également des interprétations psychologiques originales de ces algorithmes et méthodes d'apprentissage automatique, ce qui en fait une bonne introduction au "cœur" du BAT pour les psychologues non expérimentés en informatique.

Enfin, cet article est également le premier à présenter des preuves psychométriques de la validité convergente et multiculturelle du BAT, à partir d'échantillons provenant de la France et des États-Unis.

Du point de vue de la théorie de l'attachement et de la psychométrie, notre principale conclusion était que le BAT v2 mesurait effectivement des constructions très similaires par rapport à la AAQ. La corrélation élevée entre les dimensions de l'anxiété d'attachement (0,79) et de l'évitement (0,74) entre BAT v2 et AAQ dans l'échantillon anglophone était remarquable si l'on considérait que l'algorithme de notation BAT v2 avait été entraîné sur un échantillon francophone. Ces résultats constituent la première preuve psychométrique en faveur de l'utilisation du BAT v2 et de son algorithme de notation automatique comme une mesure valable pour l'attachement adulte.

Propriétés psychométriques du Biometric Attachment Test : validation française, américaine et tunisienne du premier test d'attachement adulte alimenté par l'intelligence artificielle

Notre dernier article à ce jour est actuellement en cours de révision dans le European Journal of Psychological Assessment (facteur d'impact 2.22). Cet article représente une étude de validation psychométrique plus classique pour le BAT, testant sa validité convergente et interculturelle ainsi que sa fiabilité test-retest sur des échantillons provenant de trois pays différents : la Tunisie, la France et les États-Unis.

Nos résultats ont montré une bonne preuve de la validité convergente entre le modèle RSQ du BAT et le questionnaire RSQ aussi bien pour l'anxiété que pour les dimensions d'évitement dans tous nos échantillons, ce qui suggère que les scores obtenus à l' aide du BAT avec son modèle RSQ sont des mesures valides de l'attachement adulte. De plus, le fait que les résultats soient comparables pour les trois échantillons suggère que le modèle BAT RSQ peut être considéré comme valable sur le plan interculturel au moins en ce qui concerne la France, la Tunisie et les États-Unis.

Nous avons également trouvé des preuves de la fiabilité test-retest, ce qui suggère que le

BAT peut être utilisé en toute confiance dans les études longitudinales.

3.3 Discussion

Les études que nous avons menées ont permis de mettre en évidence la flexibilité du BAT par rapport à d'autres tests psychométriques, dans différents cas de figure.

Cette flexibilité concerne tout d'abord les stimuli que nous utilisons pour déclencher des réactions liées à l'attachement. Nous avons montré que, tant que les stimuli sont choisis en suivant la procédure et l'enquête normalisées, les stimuli peuvent changer sans affecter les performances du test. En particulier, dans la première étude, des mesures répétées de l'ANOVA ont été utilisées pour tester la variabilité des réponses de mêmes participants à différents ensembles de stimuli. Aucune différence significative n'a été trouvée pour la majorité des fonctionnalités. Dans la troisième étude, BAT v3 a été utilisée, qui randomise automatiquement les stimuli chaque fois que le test est effectué de sorte qu'aucune administration du BAT ne puisse jamais être exactement identique. Et même ainsi, la fiabilité test-retest a été établie pour les dimensions anxiété et évitement sur un sous-échantillon ($n = 30$), montrant une stabilité malgré les différents stimuli.

Cette souplesse implique qu'à l'avenir, les chercheurs ou les cliniciens pourraient décider de créer de nouveaux ensembles de stimuli, éventuellement dans le but d'accroître l'effet d'identification empathique, en sélectionnant des sujets plus représentatifs d'une population donnée. Cela différencie le BAT de tous les autres tests basés sur l'exposition qui reposent sur un ensemble fixe de stimuli, tel que le AAP.

Une autre application a été révélée dans les deuxième et troisième études : la procédure de fusion multimodale en trois étapes peut aider à créer des modèles de scoring qui imitent d'autres mesures. Pour des raisons pratiques, nous avons utilisé deux questionnaires, l'AAQ et le RSQ, pour ces études ; mais le même mécanisme pourrait être utilisé pour imiter l'AAI ou même le SSP (en entrainant le BAT avec un échantillon longitudinal). À mesure que de plus en plus de modèles de notation sont développés, le BAT devient plus utile: avec un seul test de 9 minutes, un clinicien ou un chercheur pourrait obtenir plusieurs résultats (par exemple, résultat du modèle BAT SSP, résultat du modèle BAT AAI, résultat du modèle BAT RSQ). Selon le problème examiné, les utilisateurs du BAT pourront choisir le score qui leur convient le mieux ou en combiner plusieurs dans une mesure composite. Les scores pourraient aussi être comparés. Pour obtenir les mêmes résultats sans le BAT, il faudrait un investissement important en temps et en budget afin qu'un échantillon de participants subisse tous les tests nécessaires.

Une autre flexibilité que nous avons observée dans les deuxième et troisième articles est la capacité du BAT à s'adapter facilement à une nouvelle langue. En principe, le BAT peut s'adapter à toute langue pour laquelle il existe un dictionnaire validé pour le logiciel LIWC, qu'on utilise pour l'analyse linguistique. À ce jour, cela inclut l'espagnol, l'anglais, le français, le brésilien et le portugais, le néerlandais, l'allemand, l'italien, le russe, le chinois simplifié et traditionnel. Si un chercheur souhaite rendre le BAT disponible dans l'une de ces langues, il lui suffira de traduire ses instructions, un ensemble de phrases simples du type "Qu'as-tu ressenti ?" qui, au total, équivaut à une page de texte. De plus, nous essayons maintenant une solution pour les langues pour lesquelles un dictionnaire LIWC validé n'existe pas encore. Mario Mikulincer utilisera les BAT dans le centre IDC Herzliya, où les participants israéliens seront testés dans leur langue maternelle, l'hébreu. Comme il n'existe pas de dictionnaire LIWC validé pour l'hébreu, nous insérons une étape supplémentaire dans la procédure d'analyse linguistique : après la transcription automatique de Google Cloud Speech, nous avons implémenté Google Cloud Translation Services pour traduire de l'hébreu à l'anglais. Ensuite, le texte traduit est analysé en tant que texte anglais normal. Les analyses empiriques de nos résultats avec l'échantillon hébreu détermineront s'il s'agit d'une bonne approche. Si tel est le cas, cela pourrait signifier que presque toutes les langues du monde seront couvertes par le BAT, car Google Cloud Speech peut détecter un total de 120 langues et dialectes et que Google Cloud Translation Services peut en traduire 104 en anglais.

Une dernière flexibilité du BAT est sa capacité à interpréter l'attachement à la fois comme une construction dimensionnelle et une construction catégorique. La procédure de fusion multimodale en trois étapes que nous avons développée peut utiliser dans sa dernière étape soit un réseau de neurones de régression, soit un réseau de neurones de classification, ce qui permet au test d'imiter non seulement les scores d'autres tests, mais également les classifications d'autres tests. À l'avenir, une personne prenant le BAT aura non seulement des résultats dimensionnels, tels que le niveau d'anxiété d'attachement, mais également une classification, telle que sécure ou préoccupé.

Nous voudrions souligner ici les efforts de recrutement nécessaires à ce travail. Si nous additionnons des échantillons de toutes les expériences citées, le nombre total de participants s'élève à $N = 912$. Si nous divisons ces chiffres en termes de pays, une autre caractéristique de ce travail apparaît : sa propagation interculturelle. 536 participants venaient de France, 172 des États-Unis, 33 de Tunisie, 131 d'Argentine, 40 du Portugal. Ces 5 pays couvrent 4 continents différents.

Outre les limitations de chacune des études que nous avons menées et qui ont été abordées dans les articles respectifs, il existe des limitations plus générales de ce travail que nous aimerions examiner.

La première est lié à notre utilisation des mesures d'auto-évaluation pour vérifier la validité du BAT. Si les mesures d'auto-évaluation sont biaisées, et que le BAT apprend à les imiter, le BAT ne serait-il pas aussi partial ? La réponse est bien sûr oui. Pour que le BAT devienne meilleur que les questionnaires en termes de validité de construction, un modèle de notation BAT doit être formé en utilisant quelque chose de mieux qu'un questionnaire, tel que l'AAI. Les raisons qui nous ont obligés à choisir des questionnaires pour nos deux études de validation étaient purement pratiques : nous n'avions pas les fonds nécessaires pour tester toutes ces personnes avec des évaluations basées sur des entretiens, et même si cela avait été le cas, il aurait été difficile d'interviewer, de transcrire, et de noter les transcriptions de 392 personnes dans les délais prescrits pour ce programme de doctorat. Bien que nous ayons testé la validité du contenu du BAT vis-à-vis de deux mesures basées sur l'entretien (le AAP et l'AMMI) lors de notre première étude, nous pourrions néanmoins avancer un argument selon lequel jusqu'à ce que nous entrainions un système de notation pour le BAT basé sur les scores AAI et démontrer avec succès sa validité convergente, dans un sens strictement développemental, notre test BAT n'a pas encore fait ses preuves.

À cette critique, nous répondons de deux manières. La première consiste à rappelerqu'en effet l'élaboration de modèles de notation BAT basées sur des mesures fondées sur des entrevues est une priorité et définitivement dans notre feuille de route. La seconde est de mentionner que, même si nous partageons quelque peu les préoccupations relatives à la validité conceptuelle des questionnaires, nos propres analyses semblent montrer un niveau élevé de cohérence théorique pour l'attachement, mesuré par les questionnaires RSQ et AAQ, montrant les associations attendues avec la régulation des émotions, l'estime de soi, l'efficacité personnelle et les traumatismes de l'enfance dans tous nos échantillons. Nous concluons ainsi que nos choix de mesures de validité convergentes ne sont pas si mauvais.

Enfin, une critique de notre travail pourrait être que nous n'avons pas, même dans la discussion de cette thèse, comparé de manière systématique les propriétés psychométriques du BAT à celles des mesures existantes. Cette remarque semble raisonnable étant donné les critiques que nous avons formulées à l'égard de ces autres mesures tout au long de ce travail, et l'implication constante que le BAT serait meilleur qu'elles. Le test-retest du BAT est-il meilleur que le test-retest des mesures codées par l'homme ? Possède-t-il une meilleure validité interculturelle ou une validité de convergence plus élevée que les questionnaires ?

La vérité est que nous n'avons pas encore assez de résultats pour tenter une comparaison systématique. Par exemple, nous n'avons pas encore de modèle de notation entraîné à produire des classifications, pour le comparer avec l'AAI ou des mesures catégoriques similaires. Comparer les niveaux de convergence entre la BAT et d'autres mesures serait également déraisonnable, car la plupart des mesures le font par rapport à la norme AAI, ce que nous n'avons pas encore été en mesure de faire. La seule comparaison que nous puissions faire avec les données limitées dont nous disposons est l'évaluation test-retest du BAT modèle RSQ par rapport à d'autres mesures dimensionnelles de l'attachement qui évaluent l'anxiété et l'évitement. Nous proposons une comparaison non exhaustive sur le tableau ci-dessous, qui suggère que les modèles de notation du BAT doivent encore améliorer leur stabilité temporale pour pouvoir espérer être plus fiables que les mesures d'attachement classiques pour adultes.

En lisant cette thèse, et en particulier le dernier article, on pourrait avoir l'impression que le BAT est un outil prêt à être utilisé par les chercheurs et les cliniciens. Ce n'est pas encore tout à fait vrai.

Quelles sont les étapes à franchir pour que le BAT devienne un outil utile ?

La première étape consiste à déboguer le logiciel. Le logiciel BAT v3 mis en œuvre pour tester les 274 recrues de notre troisième étude (plus les N = 40 recrues du Portugal) a bien fonctionné pour les tester, mais a généré des erreurs avec beaucoup d'autres participants qui n'ont pas pu terminer la procédure. Les erreurs qui nous ont été signalées étaient variées, allant d'effondrements complets de l'application à un moment donné pendant le test, aux incompatibilités de caméra, microphone ou système, aux erreurs apparues à la dernière minute pendant la phase de transfert des données, lorsque les résultats anonymisés nous étaient transmis pour analyse avant d'être effacés. Bien que, pour nos besoins, cette application bêta soit suffisamment bonne, il est évident que le logiciel nécessite un processus de débogage complet jusqu'à ce que l'application devienne plus stable.

Deuxièmement, dans sa version actuelle, l'application BAT v3 est capable de présenter les images et les clips musicaux et d'effectuer toutes les analyses d'extraction de données : fréquence cardiaque, expressions faciales, caractéristiques de la parole (paralinguistique et linguistique), direction du regard et pose de la tête. Cependant, il ne peut pas encore fusionner les données et ne peut donc pas produire de scores par lui-même.

Tous les algorithmes de fusion fonctionnent toujours sur notre ordinateur à l'aide de Matlab, traitant les données multimodales générées à distance par l'application. Pour que l'application BAT soit totalement autonome, les algorithmes de fusion, qui produisent les scores d'attachement, doivent être traduits de Matlab en code Xamarin et inclus dans l'application utilisée par les utilisateurs. Cette tâche ne doit pas être sous-estimée, les deux codes (source Matlab versus traduction Xamarin) doivent être comparés de manière exhaustive dans plusieurs situations afin de prouver qu'ils produisent exactement les mêmes scores.

Enfin, une fois que l'application est stable et peut générer par elle-même les scores finaux, nous aurions besoin de faciliter le logiciel à d'autres chercheurs afin qu'ils puissent mener des études de validation indépendantes. En effet, ne souhaitons pas que le BAT soit utilisé dans les milieux cliniques ou de recherche sans avoir d'abord été soumis à un minimum d'études de validation externes, pour le bien de la reproduction scientifique et de la déontologie.

Nous envisageons notre travail dans le cadre d'un mouvement plus général qui touche de nombreux domaines, les humains apprenant lentement à apprivoiser le pouvoir de l'intelligence artificielle, au profit de l'homme.

La psychométrie, une discipline née autour du 19ème siècle dans ses deux courants allemands et victoriens, n'a depuis pas beaucoup progressé.

Comme nous l'avons vu tout au long de ce travail, une injection d'intelligence artificielle peut faire évoluer la tradition psychométrique.

Cela ne va pas sans dilemmes éthiques, même si nous pensons que les avantages sont sur le point de surpasser les risques.

Dans nos rêves, dans quelques années, lorsqu'on rentre dans un hôpital psychiatrique, un ordinateur ou une tablette pourrait nous présenter des images, des scènes de film ou de la musique, et dix minutes plus tard, nous révéler les dimensions psychologiques derrière nos symptômes, permettant de mieux orienter notre traitement.

Mais peut-être cela n'arrivera t-il jamais. Un cher collègue de la Harvard Divinity School nous a dit un jour que l'obtention d'un doctorat était comme gravir la cime d'une montagne étroite dépassant les nuages, loin du sol. Le sol représente notre connaissance commune, et la cime, l'effet euphorisant qu'on atteint lorsqu'on focalise son attention sur un seul sujet pendant une longue période. L'étroitesse de la montagne représente elle la fragilité d'un tel savoir acquis, dans la mesure où sa portée reste incroyablement limitée.

Nous souscrivons à cette ordonnance d'humilité. Qui sait, par exemple, si l'humanité acceptera que la santé mentale soit mesurée par des ordinateurs, aussi apprivoisés soient-ils?

Une seule chose reste sûre: notre espoir sincère que ce travail puisse éventuellement contribuer à alléger un peu les souffrances associées à un attachement insécure chez nos semblables.

4 Introduction

4.1 Why measure adult attachment? A developmental response

Attachment theory originated with the work of British psychiatrist John Bowlby and American-Canadian psychologist Mary Ainsworth. Bowlby became inspired by ethological observations and evolution theory. He theorized that the chance for survival of human genes had increased by the natural selection of behaviors that augmented proximity and bonding between infants and their caregivers, leading to a higher probability of protection for the children (Bowlby, 1969; Cassidy & Shaver, 2016).

Attachment theory posits an innate behavioral system, the attachment system, which activates mainly in times of a perceived threat, inciting the child to seek the proximity and care of the caregivers, the attachment figures. The attachment behavioral system deactivates once a felt sense of security and safety is reestablished (Sroufe & Waters, 2017; Bowlby, 1969).

This genetically determined predisposition manifests in the form of proximity-seeking behaviors, first studied by Mary Ainsworth in Uganda (Ainsworth, Blehar, Waters, & Wall, 1978). She and others have found that these behaviors are universal and can be observed in children from very different cultures around the world (Grossmann, Grossmann, & Keppler, 2005; Ainsworth, Blehar, Waters, & Wall, 1978). An exhaustive list of these behaviors can be found in Ainsworth's coding instructions for the Strange Situation Procedure, a test she developed and that we'll explore later: for babies, these behaviors include purposefully approaching the attachment figure by creeping, crawling, or walking clambering up on or grasping hold of the attachment figure; signaling by reaching or equivalent behavior that he or she wants to be picked up. If the baby is being held by a stranger and is not able to approach the attachment figure through locomotion, he or she might actively and vigorously strain toward the attachment figure. Other forms of proximity-seeking behaviors include summoning attachment figures by vocalizing or "directed crying" (Barnas & Cummings, 1994; Ainsworth, Blehar, Waters, & Wall, 1978).

The attachment figure is the person who most consistently and extensively interacts with a child. Although the theory posits a genetically determined predisposition in caregivers to respond with attuned care to children's proximity-seeking behaviors, caregivers are, for a variety of reasons, sometimes not able to do so appropriately and consistently (Ainsworth, 1985).

The quality, as well as the outcome, of these repeated early interactions between children's proximity-seeking behavior and the attachment figures' response, leave an enduring physiological, behavioral, and cognitive mark in the developing child (de Wolff & van IJzendoorn, 1997; Sroufe, Coffino, & Carlson, 2010; Fraley & Roisman, 2015; Lorber & Egeland, 2009).

The mark is physiological because negative early attachment-related experiences can lead to adults with persistent differences in the response of the bilateral amygdala and left ventral striatum during stressful situations, and to an overall higher sympathetic activation baseline (Lemche et al., 2006; Hane & Fox, 2016). The impact is also behavioral: the child's innate attachment behaviors will tend to accommodate to the availability (or lack thereof) of responsive care in his or her environment (Ainsworth, 1985). For example, in cases in which the child might face persistent unavailability or unresponsiveness from the caregivers, he or she might stop proximity-seeking behaviors altogether, and begin acting as if he or she was fully independent. Such a behavioral pattern, not innate but an adaptive response to a deficiency in care, can be carried into adulthood and negatively impact adult intimate relationships (Mikulincer, Dolev, & Shaver, 2004). It is important to note that episodic memory as a developmental capacity emerges later than attachment (3 to 4 years of age for the former, 6 months to 2 years of age for the latter). That is why the majority of what attachment experiences leave in us is stored in procedural memory, the sort of "body memory" we use for driving or biking, and can't be easily conceptualized or even accessed by conscious reflection (Brown & Elliott, 2016). Finally, we mentioned an enduring cognitive mark because as the child grows, dynamic representational models of the attachment figures and the relationship with them tend to develop (Miljkovitch et al., 2013; Miljkovitch, Moss, Bernier, Pascuzzo, & Sander, 2015). These dynamic representations, called internal working models of attachment, tend to persist throughout development and contribute later in adulthood to appraisals of the self as worthy of care and of others as capable of providing care (Waters, Hamilton, & Weinfield, 2000; George & West, 2012).

Attachment theory took on an essential role in two of the largest and most rigorous longitudinal studies in psychology to date: the Minnesota Longitudinal Study of Risk and Adaptation (MLSRA; (Sroufe, Coffino, & Carlson, 2010)), which studied 180 American children of poor families from pre-birth (mother's pregnancy) to 35 years of age; and the NICHD Study of Early Child Care and Youth Development (SECCYD, National

Institute of Health), that followed 1300 children of normative families in multiple sites in the United States, from birth up to 19 years of age. These "monster" studies collected multiple kinds of data (self-rating scales, observational studies, interviews, third-party family/teacher/partner assessment) at a myriad time points (e.g. testing the mother before giving birth, testing the newborn, testing 12 times during the first 18 months of life, testing at 3, 4, 7, 9, 15 17 and 19 years of age) throughout the life span. The studies included different settings, such as the participants' home but also their school as well as controlled lab experiments; and different dimensions, such as temperament, genetic characteristics, physiological features, traumatic events, or school success rate. Data from both studies was made available to researchers around the world, and since, more than 100 and 300 articles respectively were written based on MLSRA and SECCYD data. They revealed that negative early attachment interactions in childhood predict childhood attachment security (de Wolff & van IJzendoorn, 1997), which in turn partially predicts adult psychopathology (Sroufe, Coffino, & Carlson, 2010; Carlson, 1998). For example, the association between disorganized attachment in childhood and depression at 18 years for men were found to be of r=.4 in the MLSRA study, after controlling for all confounding factors (Sroufe, Coffino, & Carlson, 2010). Cross-sectional studies with adults seem to have confirmed this developmental discovery, with insecure attachment linked to depression (Cantazaro & Wei, 2010), post-traumatic stress disorder (Ein-Dor, Doron, Solomon, Mikulincer, & Shaver, 2010) and complex post-traumatic stress disorder (Parra, George, Kalalou, & Januel, 2017), borderline (Levy, Meehan, Weber, Reynoso, & Clarkin, 2005) and other personality disorders (Crawford et al., 2007), eating disorders (Illing, Tasca, Balfour, & Bissada, 2010), obsessive-compulsive disorder (Doron et al., 2011), suicidal ideation (Gormley & McNiel, 2009), and even schizophrenia (Mikulincer & Shaver, 2007). On the contrary, and further cementing the notion of a link between attachment and psychopathology, positive attachment experiences in adulthood, whether naturally occurring or the outcome of a methodical therapeutic intervention such as Brown and Elliott's Ideal Parent Figure method, have been shown to increase attachment security, which in turn improves mental health (Brown & Elliott, 2016; Mikulincer & Shaver, 2015; Parnell & Siegel, 2015). With Carol George and colleagues at the Ville-Evrard Psychotherapy Center in Saint-Denis, France, we further tested this hypothesis in the clinical treatment of 17 high-risk complex post-traumatic stress disorder patients replicating prior findings: improving attachment quality improves mental health and decreases psychopathology (Parra, George, Kalalou, & Januel, 2017).

To be clear, these findings do not support a deterministic developmental pathway from childhood attachment insecurity towards adult psychopathology, and it would be reductionistic to suggest that a given past experience will unequivocally produce a predetermined result in all persons. However, these findings do support the notion that attachment insecurity, in both childhood and adulthood, is both a vulnerability factor as well as a maintaining factor for psychopathology, whereas attachment security is a protective factor against them.

The aforementioned MSLRA and SECCYD longitudinal studies have also show that developmental competencies, which are essential to sustain mental health and to cope with mental health disorders, such as emotional regulation, social skills, or cognitive ability, are associated and interdependent with attachment across the life span (Bergman, Sarkar, Glover, & O'Connor, 2010; de Wolff & van IJzendoorn, 1997; Sroufe, Coffino, & Carlson, 2010).

Adult attachment is also impactful for the capacity of adults to create and sustain healthy long-term intimate relationships (Feeney, 2016). Such relationships are, in turn, vital for mental health, for they tend to function as a buffer for coping with psychopathology and stress (Ditzen et al., 2008; Rodin et al., 2007). Secure attachment, in relationship with social support, has been acknowledged as a protective factor for psychological distress (Ditzen et al., 2008), with perceived social support mediating the relationship between attachment security and depressive symptoms (Rodin et al., 2007). Attachment insecurity has been associated with having more interpersonal problems in general (Bartholomew & Horowitz, 1991), and these problems explain insecure persons' self-reported loneliness, social isolation, low relationship satisfaction, more frequent relationship breakups, greater physiological stress reaction to interpersonal conflict, and more frequent conflicts and violence (Mikulincer & Shaver, 2012; Larose & Bernier, 2001; Powers, Pietromonaco, Gunlicks, & Sayer, 2006; Wallace & Vaux, 1993).

We believe therefore that attachment is a vital dimension of adult mental health and that measuring adult attachment correctly should be a priority in all mental health settings, whether psychiatric hospitals or other mental health-related institutions, psychologists and psychiatrists' private practices or psychology research labs.

The following dissertation stems from the combined insight of the importance of adult attachment, as underlined in the paragraphs above, and the inadequacy of available adult attachment measures, a topic we will now explore.

4.2 Limitations of current assessments for adult attachment

4.2.1 Overview

Since 1984 (Findings by George C, Kaplan N, and Main M, unpublished doctoral manuscript, 1984), various validated instruments for the assessment of adult attachment developed concomitantly within the fields of social psychology and developmental psychology.

Social psychology has spurred the development of multiple self-report measures in the form of questionnaires. Examples of this approach are the Adult Attachment Questionnaire (AAQ; (Simpson, Rholes, & Phillips, 1996)), the Adult Attachment Scale (AAS; (Collins & Read, 1990)), and the Relationship Scales Questionnaire (RSQ; (Griffin & Bartholomew, 1994)).

Developmental psychology, on the other hand, has relied on a variety of broadly defined semistructured interview methods, beginning with the AAI, which was the first and is still considered the "gold standard" in the field of adult attachment assessment (Ravitz, Maunder, Hunter, Sthankiya, & Lancee, 2010). More recent examples are the Patient Attachment Coding System (PACS; (Talia, Miller-Bottome, & Daniel, 2015)) and the Adult Attachment Projective (AAP; (George & West, 2012)).

In the next page we reproduced Ravitz and colleagues' table resulting from their systematic review of 25 years of attachment measures (Ravitz, Maunder, Hunter, Sthankiya, & Lancee, 2010), in which most of the existing validated questionnaire and interview-based measures for adult attachment are listed and compared in terms of their number of items (when applicable), time consumption for both interview (when applicable) and scoring, the attachment relationship each focuses on (e.g. attachment to partner), if the measure produces categorical or continuous results, as well as their reliability and validity evidence. Ravitz et al. paper includes references for each of the validation articles that support each measure (Ravitz, Maunder, Hunter, Sthankiya, & Lancee, 2010).

In the following paragraphs, we will explore a series of limitations of self-report measures and interview-based measures for adult attachment, that we believe impair both research and clinical assessments, and that more generally speaking are reflective of the current state of psychometrics as a field.

Scale[a]	Authors	Number of items	Subject time[b]	Scoring time[b]	Relationship focus	Yields categories/ dimensions	Categories/dimensions measured	Reliability[c]	Validity[d]
Interviewer-assessed instruments									
Adult Attachment Interview (AAI)	George et al. [17], Fonagy et al. [26,44,45], Kobak [42], Fyffe and Waters [43], Grossmann et al. [46]	20	L	VL	Parents	C	Secure/autonomous, dismissing, preoccupied, unresolved/disorganized with respect to trauma	+++	+++
Adult Attachment Interview as a Questionnaire (AAIQ)	Crandell et al. [109]	20	L	L	Parents	C	Secure/autonomous, dismissing, preoccupied, unresolved/disorganized with respect to trauma	++	++
Adult Attachment Projective (AAP)	George and West [23] and Buchheim et al. [52]	8	L	L	Nonspecific	C	Secure, dismissing, preoccupied, unresolved	+	++
Attachment Style Interview (ASI)	Bifulco et al. [114]	NK	VL	VL	Close relationships	C	Secure, enmeshed, fearful, angry/ dismissive, withdrawn	+	++
Couple Attachment Interview (CAI)	Cohn and Silver [101]	29	VL	VL	Partner	C	Secure, dismissing, preoccupied	++	++
Current Relationship Interview (CRI)	Crowell and Owens [47]	22	VL	VL	Couples	C	Secure, insecure/dismissing, insecure/preoccupied	++	+++
Marital Attachment Interview (MAI)	Dickstein et al. [122]	16	L	VL	Spouse	C	Secure, dismissing, preoccupied, unresolved with respect to loss or trauma, cannot classify	+	+
Secure Base Scoring System (SSBS)	Crowell et al. [49]	8	VL	VL	Couples	D	Secure base use, secure base support	++	+
Self-report questionnaires									
Adult Attachment Styles	Hazan and Shaver [19]	1	VS	VS	Intimate relationships	C	Secure, avoidant, anxious/ambivalent	++	+
Adult Attachment Questionnaire (AAQ)	Simpson [58] Simpson et al. [61]	17 13	S	VS	Partner	D	Attachment anxiety, attachment avoidance	++	+++
Avoidant Attachment Questionnaire for Adults (AAQA)	West and Sheldon-Kellor [82]	22	S	VS	General	D	Maintains distance in relationships, priority on self-sufficiency, attachment relationship is a threat to security, desire for close affectional bonds	+	+
Adult Attachment Scale (AAS) and Revised-Adult Attachment Scale (RAAS)	Collins and Read [56] Collins [57]	21 18	S	VS	Partner	D	Comfort with closeness, comfort with depending on others, anxious concern about abandonment	++	+++
Attachment History Questionnaire (AHQ)	Pottharst [132]	51	L	VS	Partner	C and D	Categories: secure, insecure; dimensions: secure attachment, parental discipline, peer system	+	++
Attachment and Object Relations Inventory (AORI)	Buelow et al. [135]	75	L	VS	Parents, peers, partners, and self	D	View of self as: warm, secure, interdependent, not anxious versus distant, dependent/preoccupied, anxious; view of others as: emotionally accessible, responsive versus not accessible, unresponsive	+++	++
Attachment Style Questionnaire (ASQ)	Feeney et al. [73]	40	S	VS	Close relationships	D	Discomfort with closeness, need for approval, preoccupation with relations, viewing relationships as secondary to achievement, lack of confidence	++	++
Continued Attachment Scale (CAS)	Berman et al. [138]	12	VS	VS	Parents	D	Cognitive and behavioral components of parental attachment	+	++
Client Attachment to Therapist Scale (CATS)	Mallinckrodt et al. [141]	36	S	VS	Therapist	C	Secure, avoidant/fearful, preoccupied/merger	++	+
Experiences in Close Relationships (ECR) and Experiences in Close Relationships-Revised (ECR-R)	Brennan et al. [59] Fraley and Shaver [27]	36 36	S	VS	Partner (or general)	D	Attachment anxiety, attachment avoidance	++	+++
Measure of Attachment Qualities (MAQ)	Carver [145]	14	VS	VS	General	C	Security, avoidance, ambivalence/worry, ambivalence/merger	++	+
Mother Father Peer Scale (MFPS)	Epstein [147]	70	L	VS	Parents and peers	D	Acceptance/rejection, independence/overprotection, defensive idealization	+	++
Maternal Separation Anxiety Scale (MSAS)	Hock et al. [149]	35	S	VS	Child	D	Maternal separation anxiety, perception of separation effects on child, employment-related separation concerns	++	++
Parental Attachment Questionnaire (PAQ)	Kenny [150]	55	L	VS	Parents (of adolescents)	D	Affective quality of relationships, fostering of autonomy, provision of emotional support	++	++
Parents of Adolescents Separation Anxiety Scale (PASAS)	Hock et al. [153]	35	S	VS	Adolescent children	D	Anxiety about adolescent distancing, comfort with secure base role	+	++
Parenting Bonding Instrument (PBI)	Parker et al. [84,85]	50	L	VS	Parents	D	Parental care, parental protection	++	+++
Reciprocal Attachment Questionnaire for Adults (RAQA)	West et al. [80], West and Sheldon [81], and West and Sheldon-Kellor [82,83]	15	S	VS	Most important attachment figure	D	Proximity seeking, separation protest, feared loss, perceived availability, angry withdrawal; compulsive: care giving, self-reliance, and care seeking	++	++
Reciprocal Questionnaire (RQ)	Bartholomew and Horowitz [21]	4	VS	VS	Partner	C and D	Secure, preoccupied, dismissing, fearful	+	++
Relationship Scales Questionnaire (RSQ)	Griffin and Bartholomew [71]	30	S	VS	Partner	C and D	Categories: secure, preoccupied, fearful, dismissing; dimensions: model of self and model of others	+	++
Revised Inventory of Parental Attachment (R-IPA)	Johnson et al. [160]	30	S	VS	Children	D	Trust/avoidance, symptom distress, social role, interpersonal relations, physical aggression	+	+
Vulnerable Attachment Style Questionnaire (VASQ)	Bifulco et al. [22]	23	S	VS	Support	D	Insecurity, proximity seeking	++	++

[a] For each scale, the original source is cited along with revised versions and alternate coding, where applicable.
[b] Subject and scoring time are categorized according to the following labels: VS (very short), <5 min; S (short), 5–15 min; L (long), 15–60 min; VL (very long), >1 h; NK (not known).
[c] Reliability scores: (+ to ++) adequate test–retest, interrater, or interitem; one "+" for each criterion; (+++) excellent properties.
[d] Validity scores: (+) convergent with other attachment scales; (++) other evidence of convergent, discriminant, and predictive validity; (+++) excellent properties.

4.2.2 Limitations of self-report questionnaires for adult attachment

Attachment questionnaires are self-report measures. As such, they are prone to self-report biases that have been well described in the literature (e.g. (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003; Hunt, Auriemma, & Cashaw, 2003; Niedenthal, Brauer, Robin, & Åse H. Innes-Ker, 2002; van de Mortel, 2008)).

In terms of construct validity, there has been no longitudinal association demonstrated between attachment in childhood as measured for example with the Strange Situation Procedure (SSP; (Ainsworth & Bell, 1970)) and adult attachment as measured with questionnaires (KE Grossmann, 2005). Furthermore, almost no concurrent validity has been found between questionnaires and interview-based assessments of adult attachment, adding to the construct validity critique (Roisman et al., 2007). But self-report measures have been preponderant in attachment research and on occasion have provided more theoretically consistent results than their interview counterparts. For example, Fortuna and colleagues found that high levels of attachment anxiety and avoidance measured using the RSQ predicted psychopathology under conditions of high and low life stress whereas, in comparison, the gold-standard AAI only did so under conditions of high stress (Fortuna & Roisman, 2008). More generally, self-report measures are correlated between them and with scales measuring theoretically related constructs such as emotion regulation or self-esteem (Ravitz, Maunder, Hunter, Sthankiya, & Lancee, 2010). Our second article reproduced in the Experiments section contains an exploratory study of correlates of the AAQ with emotion regulation and self-esteem that replicates this (see section 5.2).

In addition to construct validity critiques, Carol George, creator of both the gold standard AAI and of the more recent AAP once pointed out that since the attachment behavioral system gets activated only when a person experiences stress, in order to measure attachment correctly the test situation should achieve a certain level of stress in the testee, so that the attachment behavioral system is made operational and thus measurable (George, personal communication, 2016). Questionnaires do not implement any such mechanism to stir up relational stress.

Finally, by their very nature, self-report questionnaires of adult attachment are not holistic: they cannot capture the many dimensions of expression of adult attachment, such as its physiological traits, cognitive representations, and multifaceted behavioral characteristics (e.g., voice changes when confronted with relational stress).

And yet, adult attachment questionnaires are easy, economical, and fast to both administrate and score. Administration can be done remotely, and automatic scoring is possible.

These practical psychometric characteristics may explain the surge of studies that have chosen questionnaires of adult attachment as their measure (Ravitz, Maunder, Hunter, Sthankiya, & Lancee, 2010).

4.2.3 Limitations of interview-based measures for adult attachment

Interview-based assessments of adult attachment rely on some form of semistructured interview, which is later transcribed and scored by a trained judge, that has undergone substantial training in a specific standardized scoring tradition.

This form of assessment is closer to childhood attachment assessments in that they also rely on third-party experts for scoring and classification. However, during childhood attachment assessments such as the SSP, the scoring experts observe behavior in general, on multiple dimensions such as movement, facial expressions, vocalization, etc., whereas in interview methods only transcribed language is analyzed. Although Mary Ainsworth husband, Erik Hesse famously claimed that taking into account other dimensions such as facial expressions or tone of voice would produce bias in transcript raters of the AAI, proof has later shown that these dimensions contain unique and important information about attachment that could add to the assessment quality (Roisman, Tsai, & Chiang, 2004). In fact Roisman and his colleagues examined facial expressions and physiological markers during AAI interviews and concluded that "The results of this study emphasize strongly that such work must include multiple levels of analysis and methods in order to fully explicate the social and emotional legacy of childhood experiences in the developmentally salient contexts of adulthood-(Roisman, Tsai, & Chiang, 2004).

In contrast to their questionnaire-based counterparts, interview methods are difficult, costly, and lengthy to both administer and score. They add additional layers to the process, that is, the manual transcription and coding of the interview. There is a training load required for both administrating and scoring. This process is costly and time-consuming. In particular, their financial cost virtually closes the door to low-budget research, such as graduate research, and makes their implementation in public mental health care unimaginable. As an example, the gold-standard AAI takes about 90 minutes to administrate and about 12 hours to transcribe, plus about 3 hours to have it scored by an expert judge.

Interview methods are impacted by an additional problem: the subjectivity inherent to any human expert judge (Cassidy, Sherman, & Jones, 2012). This limitation might decrease the replicability of adult attachment studies, adding to psychology's current "replicability crisis" (Maxwell, Lau, & Howard, 2015).

Finally, and contrary to questionnaires, interview-based methods cannot be administered remotely, limiting their application, for instance, in internet-based research.

In terms of construct validity, interview-based measures have been considered the best-practice in the adult attachment field from its onset. In particular, the AAI has consistently shown a link between parents and their children's attachment patterns, which is considered strong evidence of its validity. Also, a substantial longitudinal link has been found between the Adult Multiple Model Interview (AMMI) and the SSP (Miljkovitch, Moss, Bernier, Pascuzzo, & Sander, 2015).

When creating the BAT, efforts were put into trying to overcome the limitations of both self-report and interview-based measures, described above, while conserving their respective best characteristics. Compared to questionnaires, the BAT is not a self-report measure and thus does not suffer from self-assessment biases. The BAT is holistic, taking into account behavior, language, and physiology. Contrary to interview-based attachment measures, the BAT can be taken in only 9 minutes, whereas scoring, which is automatic, takes only ten more minutes. It can also be administered remotely, using the BAT software on the participant's computer. In the next few sections, we will explore the creative process that led to the BAT, bringing light to other of its advantages vis-à-vis its predecessors.

4.3 Inspirations for the construction of the BAT

The BAT is an exposure-based test whose psychometric conception was strongly influenced by three previous attachment measures: Bowlby's Separation Anxiety Test (SAT; (Klagsbrun & Bowlby, 1976)), the previously mentioned SSP (Ainsworth & Bell, 1970), and the AAP (George & West, 2001). We describe here not only the key characteristics of these measures but, more importantly, the notions and practices that we borrowed from them and the rationales that led us to do so.

4.3.1 Elements from the Separation Anxiety Test

The SAT was the first attachment test (Klagsbrun & Bowlby, 1976). It was designed by Klagsbrun and Bowlby as a projective attachment test for children aged 4-7 years consisting of a set of 6 pictures depicting situations in which a child, separated from their family, must cope on their own without help from their parents. The tested child is asked to imagine the protagonist's feelings and predict their behavior, and their transcribed responses are later scored and classified according to different sub-scales such as attachment security, which evaluates the child's capacity to voice the character's vulnerability and needs.

For the BAT, we took SAT's idea of using actual photos depicting real people in clear-cut attachment-related situations. This idea runs somewhat counter the projective tradition of using rather ambiguous, undefined pictures as stimuli. Such tradition, which was adopted by other attachment tests such as George's AAP, is based on the rationale that in order for a participant to be moved by a stimulus, he or she must be able to see him or herself in it, filling the blanks with details from his or her own situation. Bowlby's SAT photo stimuli relies on a different psychological mechanism: empathetic identification (Zillmann, 1995). The participant identifies with the character in the picture, as we do with characters in movies, and then vicariously experiences whatever the character experiences, which in this case is attachment activation.

4.3.2 Elements from the Strange Situation Procedure

The SSP is a structured observation protocol created by Mary Ainsworth for assessing attachment in children aged between 9 and 18 months (Ainsworth & Bell, 1970). During 20 min, the child undergoes a series of separations and reunions from their caregiver, while they are also exposed to the arrival and presence of a stranger. The child's behavior is videotaped and then systematically analyzed using a standardized protocol for attachment scoring and classification. The analysis is performed by highly trained expert judges and takes into account several aspects of the child's behavior, such as facial expressions, body and head movements, gaze direction and vocalizations.

The SSP creates a lab-controlled situation in which the attachment behavioral system of the child is purposedly activated by having the mother leave the room and having the stranger stay. The SSP also provides opportunities for deactivating[1] the attachment system: the reunion episodes, when the mother comes back. In fact, those are the key moments for SSP judges to examine: secure children are able to use their mother as a "safe haven" that brings them comfort, reassurance and soothing; and after being soothed, they become calm anew and tend to roam around to play and explore. Meanwhile, insecure children exhibit different problems to connect with their mothers upon reunion, and

[1] We would like to clarify that throughout this dissertation we'll use the terms "attachment activation" and "attachment deactivation" in their literal sense, that is, the way in which the attachment system is activated when under specific relational stress and how it becomes deactivated when that relational stress is sufficiently addressed. This is not to be confused with "avoidant deactivation," a "Minimizing strategy (...) conceived by Mary Main (1990) as a shift of attention away from conditions normally eliciting attachment behavior, leading to the apparent absence of attachment behaviors in such circumstances" (Main 1990).

different difficulties to return to meaningful exploration thereafter.

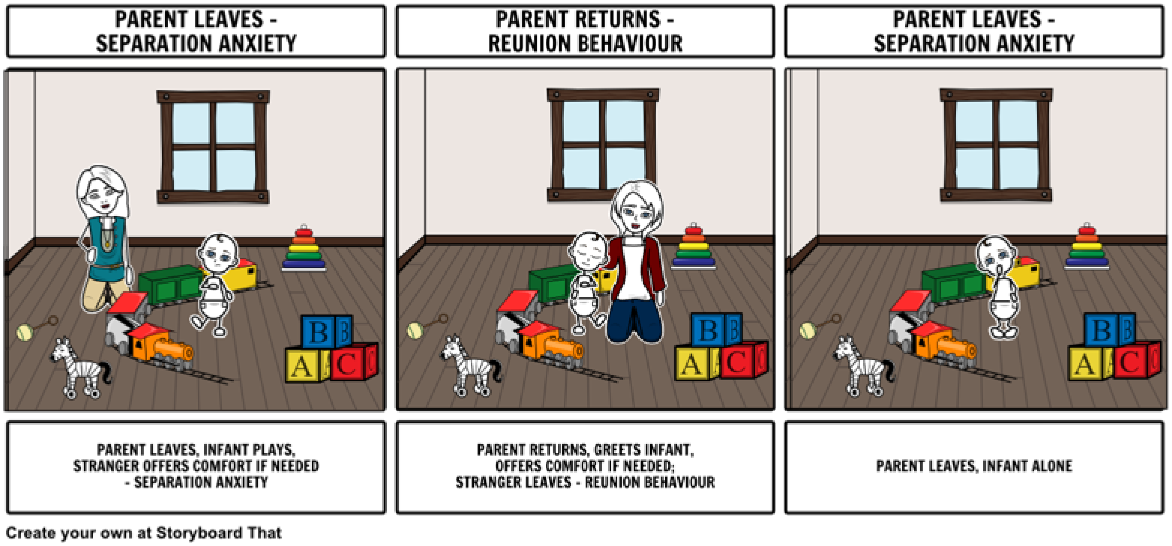


Figure 1: Mary Ainsworth's Strange Situation Procedure

We were inspired by the SSP in the development of the BAT in two ways. First, it inspired us to create a test situation that would bring forth attachment activation (stress) in the participant, but that would also offer him or her opportunities to be soothed and reassured, all the while carefully monitoring the extent to which the attachment system successfully deactivates. This strategy is in stark contrast with other interview-based attachment tests, such as the AAP, that generate a testing situation in which attachment activation increases throughout the interview, peaking towards the end of the test and culminating in a fairly attachment-activated state. There is a deontological argument to be made that it is not a best practice to let participants walk out of a testing situation feeling emotionally activated, and this argument becomes stronger when dealing with clinical patients. Our design, inspired by the SSP, mitigates this problem.

Secondly, we were inspired by the SSP to create a test that would take into account a whole range of observable behavioral dimensions, including language but also the tone of voice, facial expressions, head movements, and gaze direction, not only as separate channels of information but integratively by looking at their simultaneous combined meaning. In contrast with the SSP though, the BAT includes an analysis of physiological reactions and interprets them in conjunction with the aforementioned behavioral manifestations, offering and even more complete picture of attachment.

4.3.3 Elements from the Adult Attachment Projective

The AAP (George & West, 2001) is the second adult attachment test created by Carol George, co-author of the "gold standard" AAI. Throughout its development, George and West focused on keeping the testing time as short as possible, and succeeded to a degree, since the AAP interview is four times shorter than the AAI's and the AAP scoring takes half the time compared to the AAI's. Despite this, the AAP is 90% convergent with the AAI (George & West, 2012).

The AAP is a projective measure based on exposure to a set of black and white drawings depicting diverse situations known to activate the attachment system: separation, loss, solitude, and physical threat (George & West, 2012). Participants are asked to tell a short story about the characters in the pictures, which is later transcribed and analyzed, and an attachment classification and continuous scores are obtained (George & West, 2012). The last drawing depicts a situation that can be understood as suggestive of child abuse. This drawing helps in differentiating the unresolved pattern of attachment, which tends to get heavily disorganized when exposed to that scene.

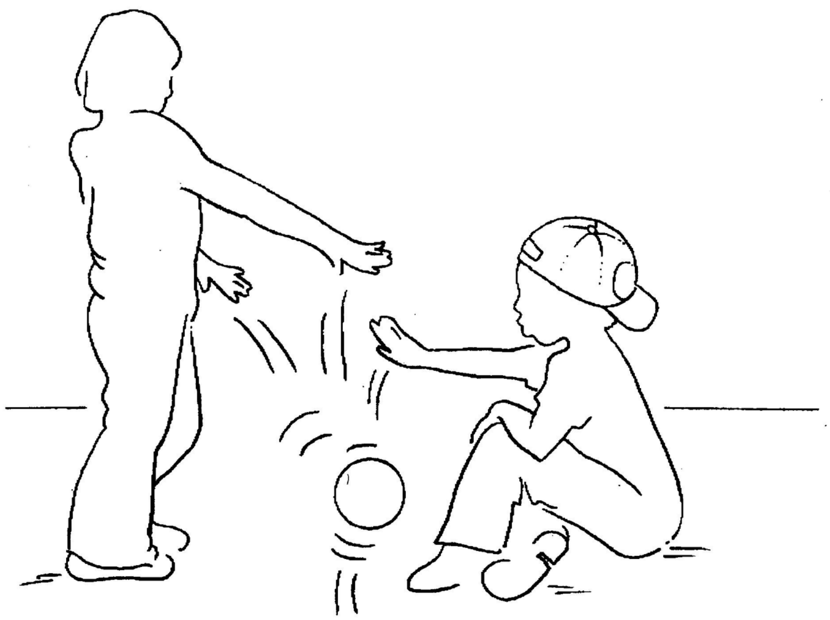


Figure 2: First picture of the AAP

In the conception of the BAT, we borrowed several concepts from George's AAP. One such concept is the idea of incorporating childhood-trauma-related imagery to the test in order to be able to better distinguish the unresolved pattern of attachment from the other ones, and to better capture disorganization of mind. However and unlike the AAP, we tried to improve deontologically by not finishing the test with that stimulus; rather, it is included in fifth place in a test with fourteen stimuli, with the seventh, eighth, tenth, thirteenth and fourteenth stimulus helping sooth the participant.

Another idea borrowed from the AAP is the notion that, in adulthood, a variety of situations can activate the attachment system: the loss of a loved one, separation from a partner, the memory of a lonely childhood, or the recollection of childhood trauma. We attempted to reproduce this by conceiving a plurality of attachment-related themes to be invoked by the BAT.

4.4 Building the Biometric Attachment Test

4.4.1 Overcoming priming: the concept of "themes" in the BAT

Exposure-based and projective tests typically rely on a fixed set of stimuli selected by the authors, as we saw with the AAP (George & West, 2012; Tuber, 2014). We pose the following critiques to this approach: first, stimuli can eventually leak into the public domain, such as in the case of the Rorschach (Schultz & Loving, 2012), and this might undermine a test's effectiveness due to priming effects. Second, longitudinal studies such as clinical trials require participants to be tested several times using the same instruments, and if stimuli are always the same this might also lead to priming effects. Additionally, we believe stimuli should be selected based on input from the general population that will consume it.

Our BAT innovates introducing the concept of themes: placeholders for actual stimuli.

A theme is a simple narrative that describes a specific situation to be evoked by a stimulus, with a specific objective. For example, in terms of adult attachment, a theme could be "the loss of a close one," its objective being to activate the attachment system (i.e., to cause attachment-related psychophysiological arousal).

The use of themes can solve the aforementioned problems with fixed-stimuli test designs: since themes are placeholders for stimuli as opposed to fixed stimuli, there is no risk if a stimuli set becomes widely known. All the contrary: stimuli in the BAT can—and probably should—be replaced from time-to-time and from context-to-context. In the case of clinical trials, stimuli sets in the BAT can rotate between assessments. Finally, the process for stimuli selection in the BAT is standardized and crowdsourced, as we will see briefly.

4.4.2 The rationale for each theme

About the themes' objectives, each is meant to evoke a reaction in the participant depending on the participant's attachment patterns. The themes were inspired by the SAT, the AAP, the SSP, and attachment theory core principles. In total, 14 themes resulted from this work.

Then, a set of objective and subjective criteria were developed for each of the BAT themes. The objective criteria were directly derived from the themes' narratives: for example, for a stimulus to be appropriate to represent the "attuned mother-child dyad" theme, there should be a mother and a child in the picture. Subjective criteria are notions that require more complex judgments: for example, for a stimulus to be appropriate to depict the "attuned mother-child dyad" theme, the child and the mother must seem attuned to each other and, thanks to said attunement, they should both seem relatively relaxed.

In the next two pages below we offer a description of each theme narrative as well as the objective and subjective criteria required for a stimulus to sufficiently evoke it.

Theme	Narrative	Manipulation intention	Objective criteria	Subjective criteria	Music
1	neutral object	To measure the participants' reactions to being in the test situation, before being exposed to attachment-specific stimuli, in order to provide baselines for behavioral and physiological measures. To warm-up the participant.	A neutral object	Neither emotionally arousing nor depressing.	Silence
2	child alone	To distinguish between more secure participants for which it will evoke introspection, and less secure participants for whom it will evoke loneliness and perhaps abandonment.	Child (1-5 years old) alone. Rain or cloudy lighting. No face, or undefined facial expression. The picture should not explicitly show physical abandonment (e.g. child all alone in supermarket). Pictures indoors are better; outdoors are fine if in situations where a child might normally be alone (e.g. tree house). Child is not performing an action, or the action doesn't define the picture (i.e. child eating)	Needs to be ambiguous enough to evoke both loneliness and introspection depending on who sees it.	Silence
3	couple in crisis	To elicit the negative feelings associated with couple crisis and/or separation, thus activating the attachment system. To gather data about the person's reaction to that activation and possible deactivation curve and strategies. Dismissing individuals might not show emotion, preoccupied ones might get hyperactivated and nervous, secure persons might have better psychophysiological emotion regulation, etc.	No violence or aggression shown in the picture. No other people are shown. The image shows explicitly two people in a situation of crisis or breaking up. They are not explicitly fighting.	Capable of evoking the feelings of a genuine crisis or break up.	Silence
4	attuned father-child dyad	To monitor how much this image can calm the attachment activation from the previous 2 images.	Father and child. No one else in the picture. Looking at each other or touching each other.	Father seems relaxed, child seems relaxed or is being soothed. There seems to be a deep connection between them.	Silence
5	child abuse	Intended to produce a specific and distinct reaction in unresolved attachment cases.	Child (1-7 years old) about to be physically abused. The child is alone except for the attacker, that can be fully shown or implied in a clear way. Otherwise, attacker alone about to attack.	Image must evoke fear.	Silence
6	unattuned mother-child dyad	Cumulative attachment activation, piling up from the previous picture, specially when the participant experienced this kind of care.	Visibly distressed young child (0-2 years old) with mother.	The child seems in distress. The mother seems ignoring this or incapable of soothing the child.	Silence
7	attuned couple	To observe how much this image helps soothe the attachment system after being activated in two previous images. Positive counterpoint to theme 3, theme 7 also allows us to infer which representation is more usual for the subject.	A couple. No explicit erotic content (to avoid activating sexual behavioral system).	The couple seems deeply connected and their connection seems tender.	Music evocative of tenderness.

Theme	Narrative	Manipulation intention	Objective criteria	Subjective criteria	Music
8	exploration	To secure individuals successfully soothed by the previous picture, to serve as an invitation to explore. Dismissing might explore while still have psychophysiological measurable stress. Preoccupied people might reject the stimulus.	Image of outward exploration: play, adventure, exotic, either in the first person or showing an adult character from behind.	Invokes a sense of freedom, doesn't seem too risky.	Silence
9	raw sadness	Music-only theme. Works as a projective test. Deep sadness/sorrow should activate the attachment system, evoking sad memories. Allows us to evaluate emotion regulation.	Blank screen		Music evocative of sadness.
10	baby	During the 15 seconds of exposure to this stimulus a secure person should have enough time to self-sooth from the sadness, and once the attachment system deactivates their caregiving system should activate, producing some spontaneous positive reaction to the baby (N. Guédeney, personal communication, 2015). Preoccupied participants, still sad, might miss the stimulus, whereas dismissing folks might avoid its implicit request for intimacy.	Baby alone, not in explicit distress.	Should seem vulnerable, relaxed, and cute.	Silence
11	raw fear	Music-only theme. Works as a projective. Raw fear is the main and most primal activation path for the attachment system.	Blank screen		Music evocative of fear.
12	death/loss	To cumulatively activate the attachment system, compounds with fear.	Tombs, cemeteries, black bag covering body, ECG showing no signal, etc. It should show death and not possible death.	The scene should symbolically mean death for most people.	Silence
13	attuned mother-child dyad	To evaluate whether this is enough to sooth the attachment system after being activated by the previous two themes. Positive counterpart to theme 6, this also allows us to infer which representation is more usual for the subject.	Mom and baby. No one else in the picture. Looking at each other or touching each other. Baby can either smile or be in the process of being soothed. Mother doesn't show obvious signals of distress.	Mother seems relaxed, baby seems relaxed/being soothed. There seems to be a deep connection between them.	Music evocative of tenderness.
14	joyful family/brothers	To help the person recover after the test and to measure how a connected-family or brothers image helps or doesn't help to further calm down the attachment system	A family or a group of brothers, no one else in the picture.	Needs to transmit joy and fun. There needs to be a sense of connection between them.	Music evocative of joy.

4.4.3 Pre-selection of picture stimuli

In order to preselect stimuli, we first used the straightforward objective criteria.

We run an exhaustive search using those criterions in three large picture databases conceived for the study of emotion: the Nencki Affective Picture System (NAPS; (Marchewka, Łukasz Żurawski, Jednoróg, & Grabowska, 2013)), the International Affective Picture System (IAPS; (Lang, Bradley, & Cuthbert, n.d.)), and the Geneva Affective Picture Database (GAPED; (Dan-Glauser & Scherer, 2011)). When, in some cases, none of these databases had suitable enough pictures for some of the themes, we turned to a stock picture service, iStockPhoto. We ended up with 140 preselected pictures, or about 12 pictures preselected per theme (themes 9 and 11 do not use pictures).

An essential feature of the three databases used for sourcing photos for the BAT is their ethnic diversity, with the subjects captured in their photos coming from all around the world. Anecdotally, we've found said diversity has facilitated the empathic identification of many of our participants with the BAT.

4.4.4 Final selection of highest ranked pictures: an international survey

When we look at the stimuli selection process adopted in psychological projective tests, most of them are based on the subjective choices of the test designers (Tuber, 2014). Indeed a particular picture stimulus may evoke a feeling in us test designers, but how can we know for sure that it will evoke the same in the majority of the population? We believe that deciding, for example, whether a mother or a child seem relaxed or not just by looking at them in a picture, is a highly subjective process that should not be arbitrarily and unilaterally decided by researchers.

It was to honor this principle that we decided to launch an anonymous web-based survey using SurveyGizmo services. The survey randomly introduced each of the 140 preselected pictures, accompanied by sliders that participants could adjust to the right or the left, signaling a judgment about a specific subjective criterion for the theme the picture was meant to evoke. In the case of "attuned couple" (theme 7), for example, one of the available sliders in the survey allowed to score the level of connection perceived in the couple, the extent to which such connection felt rather erotic or tender, and whether the picture felt genuine or scripted (this last slider was present for all pictures). See figure 3 below for an example.

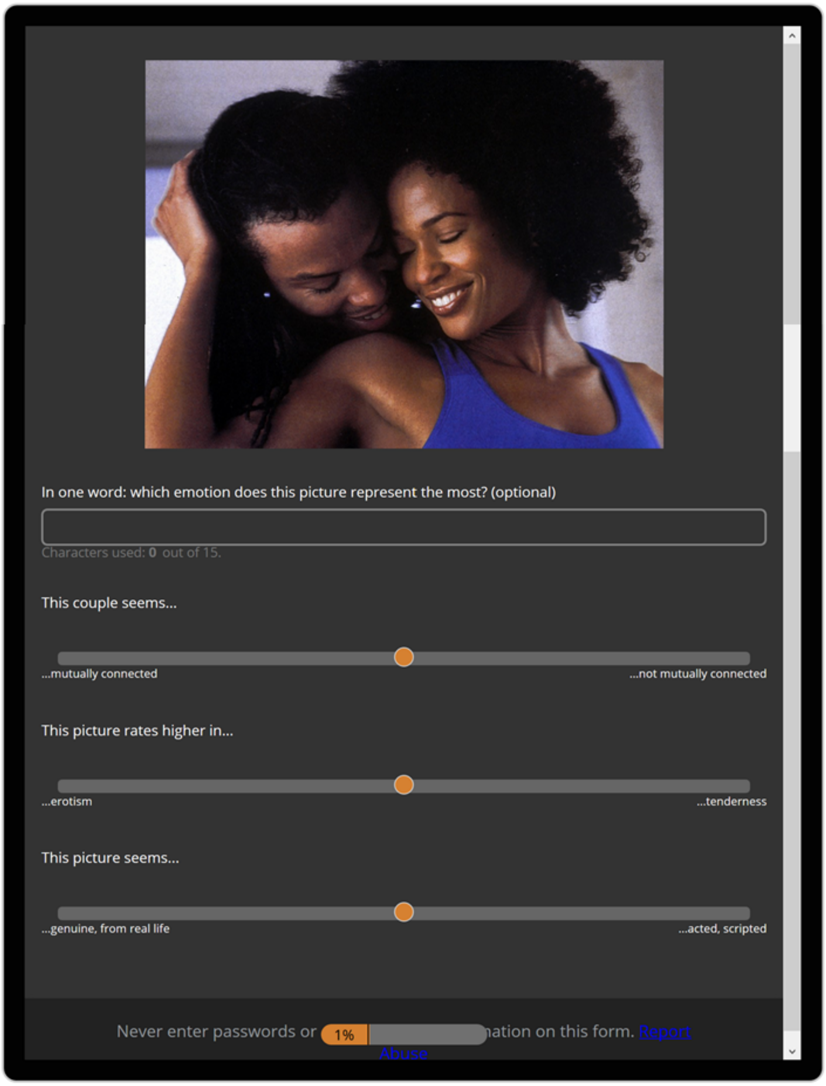


Figure 3: English version of our online survey showing a preselected candidate picture for theme 7 (attuned couple)

By design, we always opposed two traits (e.g., stressed vs. relaxed), randomizing their order of appearance in the screen and starting with the slider in the center among them, to minimize bias.

Our survey was made available in Spanish, English, and French and was distributed through social media and email campaigns in the United States, France, and Argentina. In total, 520 participants (female=72.3%, 376/520, male=27.7%, 144/520), of a variety of ages (mean 37.53, SD 10.87) responded. The survey was kept online for ten days between March 3, 2016, and March 13, 2016. We later created linear composite scores formulas for each of the 14 BAT themes, to combine the subjective criteria measured in the survey.

For the “attuned mother-child dyad” theme (theme 13), for example, the composite score formula added up the perceived level of the genuineness of the picture, plus the perceived attunement between mother and child, and substracted the perceived levels of stress in the child and the mother. A minimum required composite score was finally set for each theme to prevent stimuli that are not evocative enough from being used in the future.

The survey structure was saved, so that any researcher interested in creating a new picture stimuli set for the BAT could use it to rank his or her preselected pictures. That allows for a great deal of flexibility: researchers might want a new stimuli set for ethnic reasons (wanting the subjects in the pictures to resemble more to a specific target culture) or to avoid priming in the case our stimuli set becomes leaked into the public domain. Since the selection method is standardized, different stimuli sets created using such method tend to produce the same results in the BAT (see section 5.1).

4.4.5 Example of selected stimuli for each theme

Theme 1: neutral object (baseline)

Theme 2: child alone

Theme 3: couple in crisis

Theme 4: attuned father-child dyad

Theme 5: child abuse

Theme 6: unattuned mother-child dyad

Theme 7: attuned couple

Theme 8: exploration

Theme 10: baby

Theme 12: death/loss

Theme 13: attuned mother-child dyad

Theme 14: joyful family/brothers

BAT stimuli set sample

4.4.6 Selecting the musical stimuli

The BAT is, to the best of our knowledge, the first psychometric test to implement music as stimuli. Music was included because of its ability to trigger strong emotional feelings and experiences, overcoming psychological defenses (Baumgartner, Lutz, Schmidt, & Jäncke, 2006).

The music selection was easier and did not require a survey process since we benefited from the extraordinary work of Eerola and colleagues at the University of Jyväskylä, Finland (Eerola & Vuoskoski, 2010). Their efforts included the construction of a sample consisting of little known passages from film music, counting 360 short clips. They evaluated this sample using a panel of 12 musician experts and 116 non-experts to assess the samples' emotion-evoking qualities both from a dimensional perspective (e.g., emotional arousal) as well as a discrete perspective (e.g., fear, sadness, joy). The result is a free library made available to researchers around the world (located at https://www.jyu.fi/hytk/fi/laitokset/mutku/en/research/projects2/past-projects/coe/materials/emotion/soundtracks/Index).

Of particular interest to the BAT, Eerola and colleagues theorized that globalization had made films and their soundtracks cross all frontiers, becoming a worldwide cultural phenomenon. Children in remote locations might listen to very different music styles compared to western music, but they will probably also at some point come across a Hollywood movie and become familiar with the music language thereby implemented. According to Eerola et al., this means that film music clips have the power to activate specific emotional states cross-culturally. The downside, they noticed, is that film music clips are often remembered in association with the film itself, and thus might produce retrieval of the memory of the film when listened to. To circumvent this, the team of musical experts from the University of Jyväskylä was careful enough to only choose short clips from unknown parts of the film soundtracks, so that their film source was not recognizable.

As we've seen in a previous section, BAT's themes 9 and 11 ("raw sadness" and "raw fear") are music-only themes. In contrast, BAT's themes 7, 13, and 14 use music as a reinforcement of the picture stimuli's evocative power, a common use of music stimuli (Baumgartner, Lutz, Schmidt, & Jäncke, 2006). Using University of Jyväskylä's own emotional ranking for their music clips, we were able to select 5 sets of clips for each BAT theme that uses music, keeping the 5 higher ranked clips for every given emotion (e.g. for theme 11, "raw fear", we selected the 5 music clips that were the highest rated in terms of evoking fear).

4.4.7 BAT test situation, and software implementations

After having created the first BAT picture and music stimuli sets, we set out to design the exact way in which they would be used during the test situation.

One objective of ours was to create a concise test experience to minimize the burden on participants, in particular, clinical patients, as well as to reduce time load in research experiments.

We decided to expose participants to each theme for 15 seconds.

We also knew that it was essential to get the participants to talk about their experience with the stimuli so that we would be able to analyze the content and delivery of their speech. So we decided that after the 15 seconds of exposure, we would ask the participant to comment on what they experienced for 20 seconds. We decided to use the language of emotion, asking the simple question, “What did you feel?” after each stimulus.

Finally, we chose to offer a 5-seconds pause before moving onto the next stimulus so that the person could recover.

This design made the total duration of the test of 9 minutes and 30 seconds.

The idea was for the whole process to be guided automatically by a computer. That would allow for a participant to take the test in a room all alone, where aloneness can facilitate the activation of the attachment system (Bowlby, 1969) and in that way renders the test situation closer to Ainsworth’s SSP. It also removes possible interference from researchers.

The first BAT automatic administration procedure was constructed using OpenSesame software, version 3.1.2 (Mathôt, Schreij, & Theeuwes, 2011), and required participants to wear Empatica’s E4 research wristband to measure heart rate, body temperature and electrodermal activity (McCarthy, Pradhan, Redpath, & Adler, 2016). This first version of the software could only show the stimuli and record the participants’ responses in video format, which was later analyzed using Matlab.

Much later in 2019, we developed from scratch a stand-alone BAT multi-platform application capable of displaying a randomized combination of BAT stimuli (randomly choosing one stimulus per theme from among five preselected sets) and guiding the person through the whole administration process while automatically analyzing their responses in real-time without the need to record any identifiable data (i.e., video), and capable of extracting heart rate signal from the video without wearing any special equipment (see section 4.5.2). The application was programmed in Microsoft Xamarin (Hermes, 2015), with parts written in Matlab and C++. We called this application BAT v2.

4.5 Automatic analyses of BAT responses

4.5.1 Inspiration for the automatic analysis of the responses

Finding psychophysiological and behavioral markers of psychological conditions is gaining traction within mainstream psychiatry (Cummins et al., 2015; Scherer, Gale, Gratch, Rizzo, & Morency, 2016), as part of a quest to provide more objective and precise clinical assessments to patients.

The American National Institute of Mental Health released a statement in 2013 (URL: https://www.nimh.nih.gov/about/directors/thomas-insel/blog/2013/transforming-diagnosis.shtml) in which it made explicit its desire of moving toward more objective and precise diagnostic methodologies: "[...] Patients with mental disorders deserve better. NIMH has launched the Research Domain Criteria (RDoC) project to transform diagnosis by incorporating genetics, imaging, cognitive science, and other levels of information to lay the foundation for a new classification system [...] It became immediately clear that we cannot design a system based on biomarkers or cognitive performance because we lack the data. In this sense, RDoC is a framework for collecting the data needed for a new nosology".

Several attempts to tackle this problem have arisen from the Computer Sciences. In a recent review, Cummins et al. (Cummins et al., 2015) reviewed the state-of-the-art in the automatic detection of depression and suicidality through the analysis of speech and its paralinguistic acoustic features. Scherer et al. described, in 2013, a set of automatically extracted audiovisual nonverbal behavioral features helpful in the identification of depression, anxiety, and PTSD (Scherer et al., 2013).

The missing link between the objective measure of singular biometric or behavioral markers, and the sought ability to offer more precise diagnoses, lies on the use of machine learning algorithms that can fusion multiple modalities of data at once (Novak, Mihelj, & Munih, 2012). This allows for the uncovering of intricate multimodal data patterns that can serve in the automatic assessment of specific mental conditions.

In recent studies, such multimodal systems have approached human performance in the detection of indicators of PTSD (Lucas, Gratch, Scherer, Boberg, & Stratou, 2015).

In 2012, Sharma and colleagues conducted a survey of available physiological and behavioral measures to detect stress (Sharma & Gedeon, 2012). The ranking represented in Table 1 summarizes their findings and details which of their ranked biometric measures are included in the BAT test. Stress is not attachment, but being able to accurately detect if an attachment-related stimulus increases stress or, induces relaxation is the conundrum at the heart of the BAT.

The BAT implements eight out of the top twelve measures of stress included in Sharma et al.'s ranking.

Rank	Measure	Included in BAT
1	Hear rate variability (HRV)	Yes
2	Electrodermal activity (EDA)	Yes*
3	Electroencephalography (EEG)	-
4	Pupil diameter	-
5	Voice analysis	Yes
6	Eye gaze	Yes
7	Facial expressions	Yes
8	Blood pressure	-
9	Skin temperature	Yes*
10	Blood volume pulse	Yes
11	Eye blinks	Yes
12	Respiration	-

Table 1: Sharma et al. empirical ranking (smaller is better) of measures for stress based on correlation with stress claimed in literature, equipment intrusiveness, and techniques developed for mapping to stress scales, and whether the BAT test includes them (Sharma & Gedeon, 2012). * included in BAT v1 but discarded in BAT v2 (see section 5.2)

We hypothesize that this ranking roughly corresponds to the importance of biometric measures with regards to detecting attachment in adults, with linguistic analysis, which doesn't appear on the list, coming in last. We base our hypothesis on the fact that, as seen in section 4.1, the traces of early attachment experiences lie above all in our autonomic nervous system, amygdala, and procedural memory.

We will now introduce the measures that were retained for the BAT in its latest version, in the order of this ranking. We will use the term sense modality to refer to each of these measures, a semantic choice that invites an analogy with the human senses.

4.5.2 Parasympathetic nervous system: heart rate variability

The autonomic nervous system (ANS) is a section of the peripheral nervous system that supplies smooth muscle and glands, influencing the function of internal organs. A control system acting largely unconsciously, it regulates bodily functions such as the heart rate, digestion, respiratory rate, pupillary response, urination, sexual arousal, coughing, sneezing, swallowing and vomiting, and constitutes the primary mechanism to regulate our fight-or-flight response (A. Schmidt, 1989). The hypothalamus acts as a regulator and an integrator for autonomic functions, receiving ANS regulatory input from the limbic system to do so (Task Force of the European Society of Electrophysiology, 1996).

The autonomic nervous system has two main branches: the sympathetic nervous system and the parasympathetic nervous system. The parasympathetic system is in charge of the stimulation of "rest-and-digest" or "feed and breed" activities that occur when the body is at rest, especially after eating, including sexual arousal, salivation, lacrimation (tears), urination, digestion and defecation (McCorry, 2007). Sympathetic and parasympathetic divisions tend to function in opposition to each other. In terms of psychophysiology, the parasympathetic system is often depicted as promoting a relaxation response through the inhibition of stress and promoting a general sense of calmness (Sakakibara, Takeuchi, & Hayano, 1994; Terathongkum & Pickler, 2004).

One method to reliably measure the activity of the parasympathetic nervous system is to measure heart rate variability. Heart rate variability(HRV) is the physiological phenomenon of a variation in the time interval between heartbeats (see figure 4).

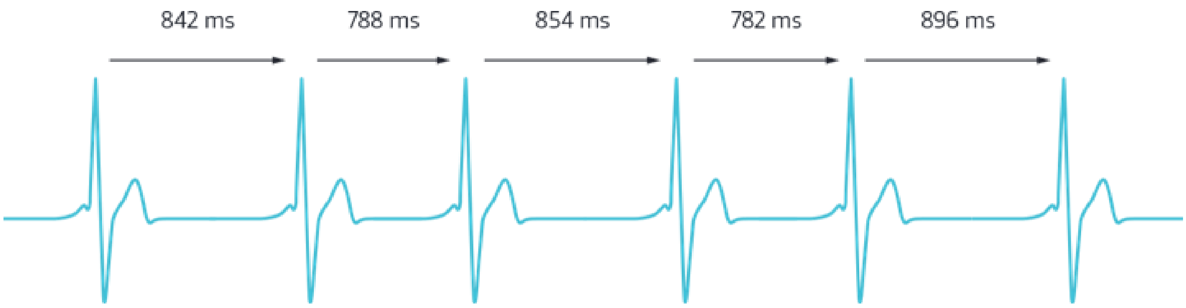


Figure 4: Inter beat intervals, used to calculate heart rate variability (HRV).

In western medicine, HRV interest begun in 1965, when Hon and Lee noted that fetal distress was preceded by alterations in interbeat intervals before any appreciable change occurred in the heart rate itself (Task Force of the European Society of Electrophysiology, 1996).

Once the HRV signal is obtained by measuring the interbeat intervals of a raw heart pulse signal, there are many techniques for processing the HRV signal to obtain features for analysis. Two classic groups of HRV features are the time-domain and frequency-domain features. The former entails different statistical analyses of the HRV signal, such as calculating the square root of its variance (SDNN) or the square root of the mean squared differences (RMSSD). There are myriad such features in the literature, but all of them tend to be highly correlated (Task Force of the European Society of Electrophysiology, 1996). Frequency-domain analyses, on the other hand, entail a power spectral density (PSD) analysis of the signal (usually using a Fourier transform) and provide information on how power distributes as a function of frequency in the signal. Of special interest for the BAT is the high-frequency power feature, or HF, obtained through this method. It is highly associated with the activation level of the parasympathetic nervous system and therefore can be used as a real-time proxy for the body's unconscious relaxation response.

With regard to attachment, Maunder et al. found that attachment avoidance was inversely associated with HF (Maunder, Lancee, Nolan, Hunter, & Tannenbaum, 2006). Farina and colleagues found that dissociative disorder patients exhibited a different pattern of HF-LF ratio during the AAI interview when compared to healthy controls (Farina, Speranza, Imperatori, Quintiliani, & Marca, 2014). Sbarra et al. found that RSA, another time-domain feature of HRV, predicted self-regulation capacity in dismissing participants following the loss of a relationship (Sbarra & Borelli, 2013). Dalsant et al. found that participants with optimal maternal bonding showed a greater calming response to distressful stimuli measured through HRV (Dalsant, Truzzi, Setoh, & Esposito, 2015).

To be able to measure heart rate accurately enough to calculate HF, yet unobtrusively enough to allow for a pleasant testing situation (i.e., without the need for wiring the participant to a machine), we first opted for the use of a research-grade, validated photoplethysmography wrist-band, the Empatica E4 (BAT v1; see figure 5; (McCarthy, Pradhan, Redpath, & Adler, 2016)). Compared to the "gold standard" ECG, the Biopac MP150, the Empatica E4 had an error of 2.1%.

Photoplethysmography (PPG) is a technique that measures variations in the absorption of light by human skin and has acquired enough precision to measure HRV (Schäfer & Vagedes, 2013).

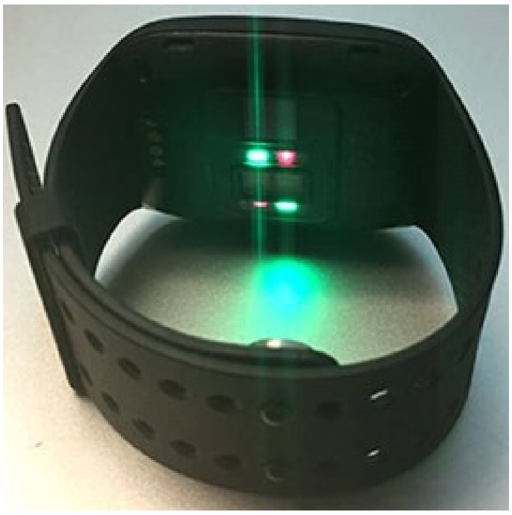


Figure 5: The Empatica E4 implemented in BAT v1

Later on, we came across an even less intrusive technology: RPPG or remote photoplethysmography(BAT v2; (Sun & Thakor, 2016; McDuff, Estepp, Piasecki, & Blackford, 2015)). By analyzing skin tone micro-variations in video images, RPPG does not require any special equipment other than a camera. Since a camera is already used by the BAT to obtain other measures such as facial expressions, and since cameras are available in most consumer computing products (e.g. cell phones, laptops), implementing RPPG makes the BAT v2 a test that can be used without any special equipment at all. An in-depth explanation of the specific RPPG method implemented can be found in section 5.2. The method was developed by a team led by Yannick Benezeth at the University of Bourgogne Franche-Comté (Bobbia, Benezeth, & Dubois, 2016).

To recapitulate: the BAT analyzes the heart pulse, obtained through either a wristband (BAT v1) or through the analyzes of skin tone changes in video (BAT v2). Then, it extracts the interbeat intervals and calculates heart rate variability. Finally, it uses a frequency-domain analysis method to extract high-frequency power, which is a proxy for the involuntary relaxation response. Proceeding this way the BAT has direct access to a level of psychological phenomena that is mainly hidden from view. This allows us to understand whether a particular theme, for instance, theme 13 representing an attuned mother-child dyad, is effective in calming a given participant -and how long it takes them to do so- after confronting theme 12 which evokes the death of a loved one.

If early attachment traces are partly stored as nervous system reaction patterns that are retriggered by intimate relationships, this measure should contribute greatly to the assessment of adult attachment.

4.5.3 Sympathetic nervous system: from electrodermal activity to Bayevsky stress index

The other main branch of the ANS, the sympathetic system, is responsible for the fight-or-flight response and its prolonged activation elicit the release of adrenaline from the adrenal medulla (Shaffer, McCraty, & Zerr, 2014). More generally, its activation dilates the eyes, lungs and blood vessels (increasing blood pressure); increases heart rate, kidney renin secretion and activates sweat secretion. It is responsible for priming our bodies for action, especially in situations perceived as threatening survival (Shaffer, McCraty, & Zerr, 2014).

Although many researchers have attempted to measure HRV's low-frequency power (LF) feature as a proxy for sympathetic activation, the consensus today is that spectral analyses of HRV and the LF feature, in particular, are not a reliable measure of sympathetic activation, because they are mainly determined by parasympathetic activity (del Paso, Langewitz, Mulder, van Roon, & Duschek, 2013).

In the absence of a reliable method to measure sympathetic activity from HRV in the western literature, we first turned towards electrodermal activity (EDA), another physiological measure that has been the "gold standard" for evaluating sympathetic activation since 1878 (BAT v1; (Boucsein, 2012)). EDA, also known as skin conductance and galvanic skin response, is a measure of the continuous variation in the electrical characteristics of the human skin which changes depending on the state of sweat glands in the skin, which are controlled by the sympathetic nervous system.

With regards to attachment, Diamond and colleagues found that individuals with both preoccupied and dismissing attachment exhibit heightened EDA in response to general stressors (Diamond & Fagundes, 2010). Similarly, research studying adults' memories of early attachment experiences found that those with dismissing attachment had heightened EDA responses when evoking past romantic conflict (Roisman, 2007). Interestingly, these studies have not found an increase in the self-reports of stress among dismissing adults, suggesting a form of dissociation. This, in turn, reinforces the BAT's core principle that measuring physiology should be the cornerstone of a complete attachment assessment.

In order to measure EDA in a non-invasive way, compatible with the ergonomic principles of the BAT, we turned to the Empatica's E4 wrist-band mentioned in the previous section, which can also measure EDA (Koskimäki et al., 2017; Garbarino, Lai, Tognetti, Picard, & Bender, 2014).

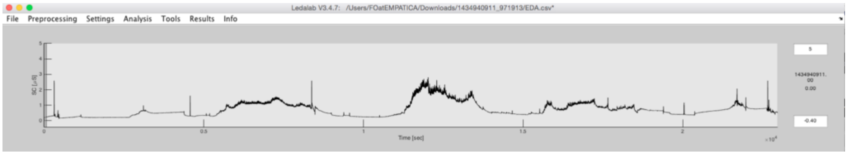


Figure 6: A typical EDA signal inspected using Ledalab software.

In 2018, we became aware of the Soviet literature on HRV. In the USSR, HRV began being studied decades before the west became interested in the subject, thanks to the Soviet spatial program (Baevsky, Bogomolov, Funtova, Slepchenkova, & Chernikova, 2011). Soviet scientists were invested in practical measures for the stress of astronauts in space and developed unique mathematical features for HRV that in many cases do not have an equivalent in the western literature. One such example is the Bayevsky Stress Index, which is capable of measuring sympathetic activity (Mashin, 2007).

By including Bayevsky's feature in our calculations, and using the HRV signal extracted through the RPPG method as described in the previous section (and more in-depth in section 5.2), the BAT v2 can now accurately measure sympathetic activation (i.e., stress response) to our 14 themes and their stimuli without requiring any special equipment other than the camera present in most of today's computers and cell phones, accessing through this modality to the visceral, unconscious responses to attachment-related phenomena.

4.5.4 Paralinguistic characteristics of speech

The study of the human spoken voice and its so-called paralinguistic characteristics (or prosody) in relation to the speaker's mental health have been a subject of interest in psychiatry since the early 20th century. In 1921, for example, Emil Kraepelin noted specific speech characteristics of people with depression, describing them as low pitched, slow, hesitant, monotonous, stuttering, and whispering, whereas pre-suicidal voices were described as hollow and toneless (Kraepelin, 1921).

Speech production implies cognitive planning and complex muscular activity (see figure 7). Planning involves constructing the message one wants to communicate and setting up the phonetic and prosodic information associated with the message (Cummins et al., 2015). This information is then stored in working memory briefly before the speaker executes a series of neuromuscular commands required to produce the speech: the air is forced from the lungs through the vocal folds and the glottis, where the vocal tract act as a filter amplifying and attenuating different frequencies, shaping the glottal waveform (Cummins et al., 2015). Speaking involves the respiratory, laryngeal and articulatory muscles, as well as the articulators: mandible, lip, tongue, velum, jaw and pharyngeal constrictors.

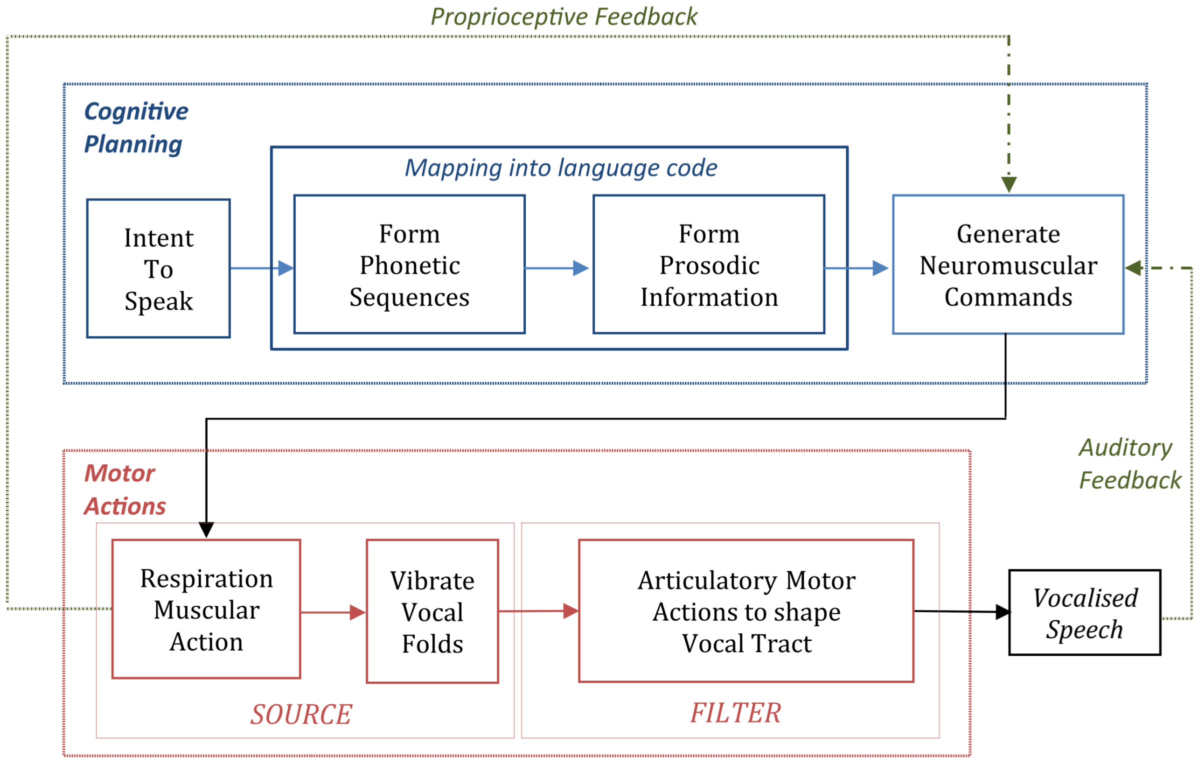


Figure 7: Schematic diagram of speech production, reproduced from (Cummins et al., 2015)

There are many ways in which speech can be influenced by our emotional and psychological states. Serotonin and brain-derived neurotrophic factor (BDNF) are linked to working memory performance (Brooks et al., 2014), affecting the phonological loop causing phonation and articulation errors (Christopher & MacDonald, 2005). Cognitive impairment affects speech planning, degrades neuromuscular motor coordination associated with speech and alters proprioceptive feedback loops necessary for calculating articulator positions (Cummins et al., 2015). Autonomic nervous system changes, such as those seen in people with insecure attachment, can cause disturbances in muscle tension and respiratory rate (Lemche et al., 2006; Cummins et al., 2015). The former can influence vocal fold behavior and change vocal tract dynamics constraining jaw and facial muscles vital for articulation, whereas the latter tends to affects subglottal pressure. Further vocal tract properties variance will be induced by salivation and mucus secretion, both also controlled by the autonomic nervous system.

Prosodic features (or paralinguistic features of speech) are the phoneme-level variations in rhythm, stress, and intonation of speech (Cummins et al., 2015). Examples are the speaking rate, pitch, and energy dynamics. Many of these prosodic features are language-independent, giving the paralinguistic analysis a cross-cultural advantage over other culturally-bound forms of analysis, such as linguistic analysis (Scherer, personal communication, 2016).

Clinically, these prosodic or paralinguistic features have been used to successfully detect depression, PTSD, psychosis, and suicidality (Scherer, Morency, Gratch, & Pestian, 2015; Pestian et al., 2016; Cummins et al., 2015; Wörtwein et al., 2017).

Zeng and colleagues analyzed paralinguistic differences among participants taking the AAI and found that there were useful prosodic markers of adult attachment that should be included in a holistic assessment of the construct (Zeng et al., 2006).

The BAT conducts an automated analysis of the paralinguistic characteristics of the participants' responses to the question "What did you feel?", that is presented after each of the 14 themes stimuli. Each twenty-seconds-long response is scrutinized using the COVAREP framework, an open-source toolbox containing state-of-the-art algorithms for prosodic feature extraction (Degottex, Kane, Drugman, Raitio, & Scherer, 2014). Figure 8 below shows the analysis workflow.

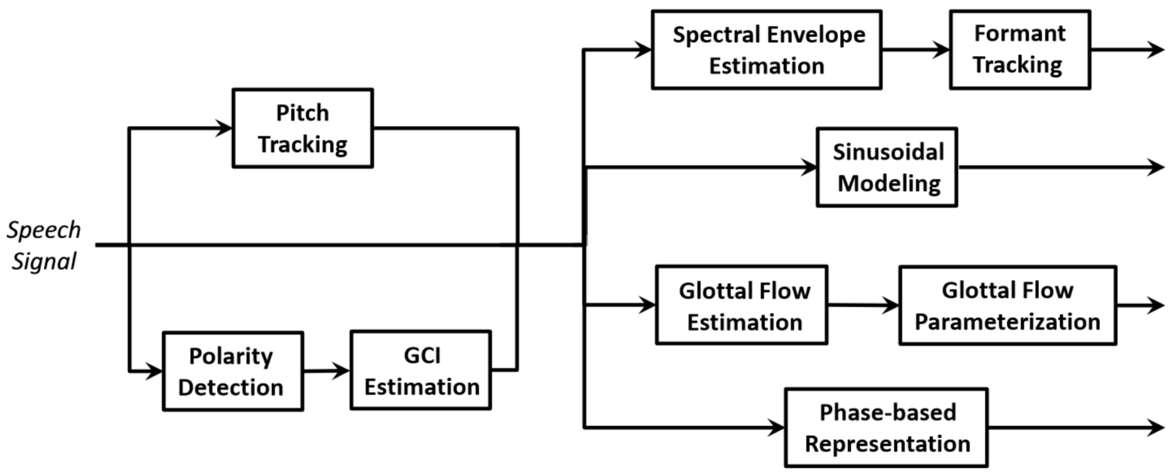


Figure 8: Workflow of the methods implemented in COVAREP (Degottex, Kane, Drugman, Raitio, & Scherer, 2014).

4.5.5 Gaze direction and head movements

Gaze direction signals our focus of social attention: our eyes, and where we look, are important social signals (Carrick, Thompson, Epling, & Puce, 2007). Changes in gaze direction might indicate indifference, attraction, or deceit (Emery, 2000). Tucker and colleagues found that secure partners tend to gaze more at their significant others in couple interactions (Tucker & Anders, 1998). The role of eye-to-eye contact in maternal-infant attachment has been thoroughly studied (Koulomzin et al., 2002; Robson, 1967). And in fact, gaze is so important to attachment that it mediates attachment between dogs and their owners: when dogs gaze at their owners, the latter's urinary oxytocin levels increase after the interaction (Nagasawa et al., 2015).

When it comes to gazing at stimuli, Guastella and colleagues found that participants given oxytocin showed an increased fixation and gaze time towards the eye region of pictures of humans compared to placebo participants (Guastella, Mitchell, & Dadds, 2008). Cecchini and colleagues showed young women picture stimuli depicting mother–newborn dyads gazing at each other, in some of which the mother avoided the newborn's gaze by looking upward. The young women's avoidance attachment dimension was found to be associated with brain activity in response to the depicted avoidant gaze (Cecchini, Iannoni, Pandolfo, Aceto, & Lai, 2015).

Peltola et al. tested whether gaze directed at facial expressions, assessed with an eye-tracking paradigm at 7 months of age, predicted infant-mother attachment in the Strange Situation Procedure at 14 months. Attention to fearful faces at 7 months predicted attachment security, whereas attachment disorganization was linked to an absence of the age-typical attentional bias towards them (Peltola, Forssman, Puura, van IJzendoorn, & Leppänen, 2015).

For gaze assessment, the BAT implements Openface, an open-source framework developed at the MultiComp Lab of Carnegie Mellon University (USA) that implements algorithms that detect gaze direction from video without requiring special equipment such as eye-trackers (Baltrusaitis, Robinson, & Morency, 2016).

The BAT uses the framework to track gaze while participants look at each of the 14 themes: do they look towards, or away, from a mother holding a child tight (theme 13)? How about when theme 5 evokes child abuse?

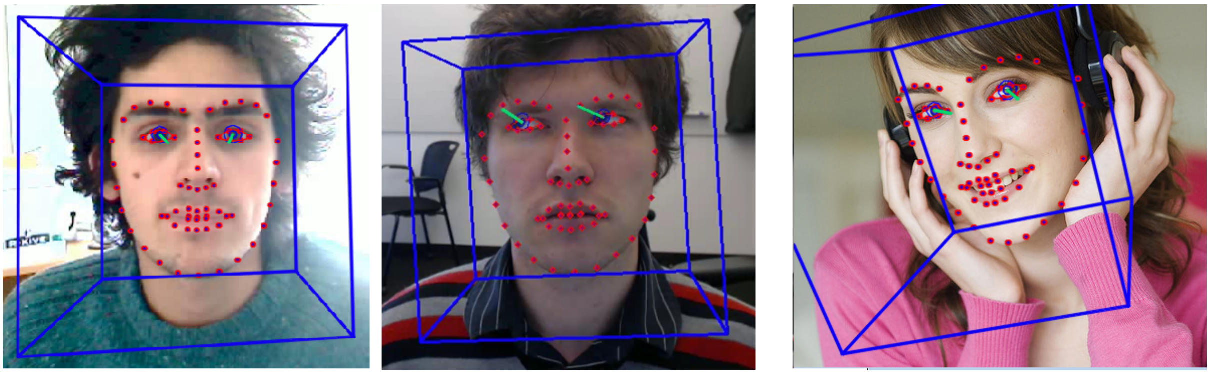


Figure 9: Openface detecting gaze and head movements.

To fully understand the changes in attention focus related to gaze, we must take into account head pose as well. In addition to gaze direction, the Openface framework can detect head movements, creating a 3D representation of facial landmarks and projecting them back into the image using camera projection techniques. Specifically, using this analysis method the BAT obtains head pitch, yaw, and roll information about the participant as they look at the stimuli, complementing the gaze data to more fully understand what the participant was looking at any given time during the test.

Finally, Openface helps the BAT detect whether a participant approaches or distances him or herself from the stimuli, as they move closer or further away from the camera. This data may also signal interest, rejection, or some other behavior or state with regards to what a given theme represents, information that might be important to assess attachment.

4.5.6 Facial expressions

Our interest in facial expressions initiated with the work of Paul Ekman, an American psychologist interested in the intersection of emotions and facial expressions. His research across different literate Western and Eastern cultures supported the Darwinian hypothesis that emotions were evolved traits common to the whole human species (Darwin, 1872). In particular, using his famous test for emotion recognition (Pictures of Facial Affect, POFA, 1976), Ekman found wrath, grossness, scaredness, joy, loneliness, and shock to be universal facial expressions, whereas he also found preliminary evidence for the universality of contempt. Later he and colleagues replicated these findings with preliterate Fore tribesmen in Paua New Guinea that had not been in contact with media and could not have learned western facial expressions that way.

In 1978, Ekman and Friesen published the Facial Action Coding System (FACS) manual (Ekman & Friesen, 1978). FACS is a system for describing observable facial movements corresponding to human emotion. Ekman and Friesen based their system on prior work by Swedish anatomist Carl Herman Hjortsjö.

Each component of facial movement in the FACS is called an action unit (AU), and all facial expressions can be decomposed into several AUs. AUs are independent of any interpretation. Therefore, they can be implemented in higher-order decision-making processes, including the recognition of basic emotions.

FACS defines AUs in terms of the contraction or relaxation of one or more facial muscles.

The system is very subtle: for instance, FACS can distinguish insincere and voluntary "Pan-Am smile" (a contraction of the zygomatic major alone) from sincere and involuntary "Duchenne smile" (a contraction of the zygomatic major and the inferior part of orbicularis oculi) (Giudice & Colle, 2007).

Traditionally, trained FACS coders would watch a videotaped recording of a participant's face while performing a task or participating in an interview, pausing the video every few seconds to manually code each AU, a process that was extremely lengthy. Coding tended to differ somewhat between coders so that some mechanism to reach consensus needed to be agreed upon and implemented.

With the advancements of computer vision since the 2000s, a new possibility begun delineating: the automatic scoring of the FACS powered by machine learning algorithms. These algorithms are trained on hundreds of videos coded by trained FACS human coders so that the algorithms can "learn" how to code. Incredibly, in the last few years, these algorithms have overcome the human raters' performance (Dhamija & Boult, 2018).

The BAT once again uses the Openface framework described in the previous section. Openface implements cutting-edge algorithms for FACS AUs automatic coding (Baltrusaitis, Robinson, & Morency, 2016). Specifically, Openface is able to recognize a subset of Ekman's AUs: 1, 2, 4, 5, 6, 7, 9, 10, 12, 14, 15, 17, 20, 23, 25, 26, 28, and 45 which is eye blinks (see figure 10 below).

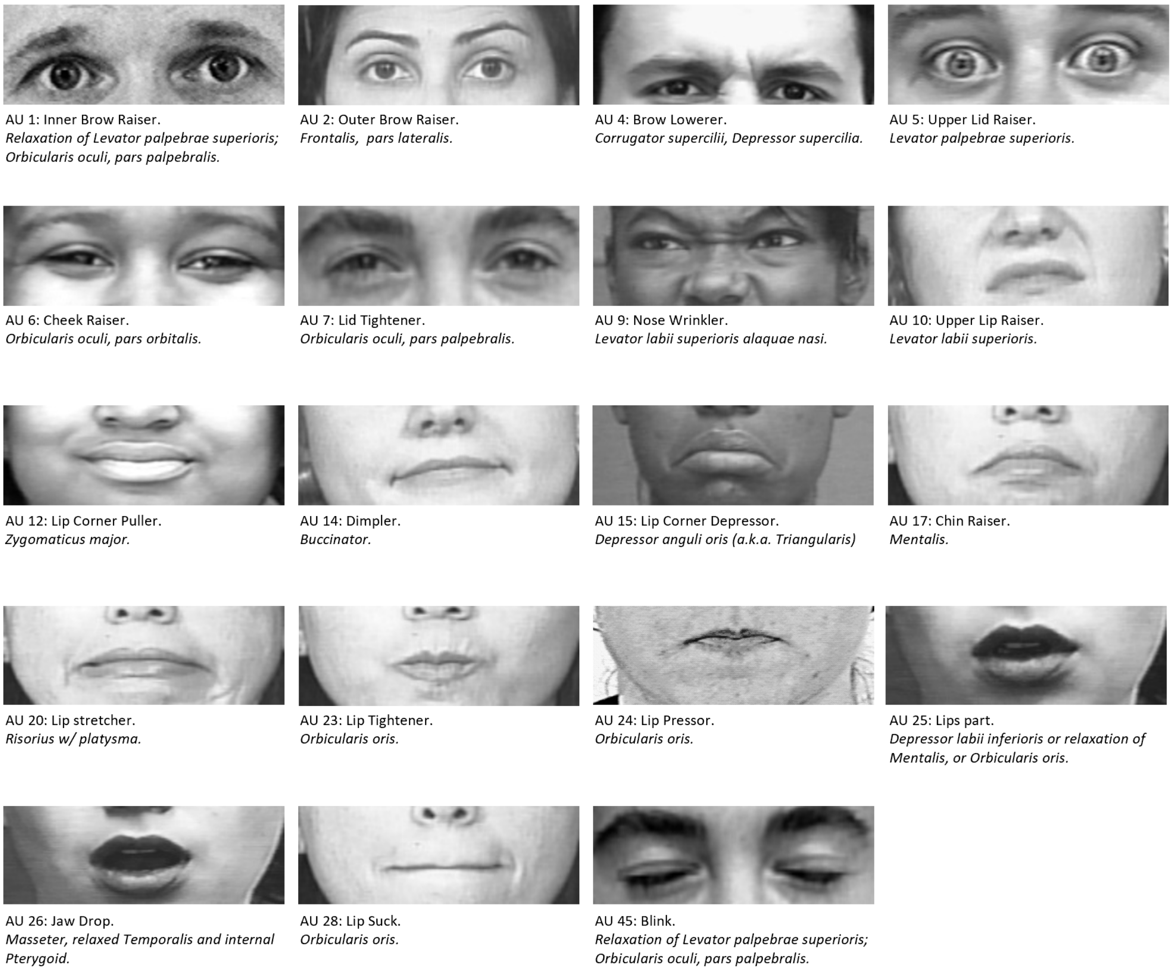


Figure 10: All FACS AUs detected by Openface (AU number, description, facial muscle(s) involved). AU 45 is eye blinks.

Examples of the use of Openface detected FACS AUs in psychology are studies on the automatic detection of suicidal ideation (Laksana, Baltrusaitis, Morency, & Pestian, 2017) and psychosis (Supriya et al., 2016).

Although FACS, manually or automatically coded, has become very popular in psychology, not many studies have been conducted that implement AUs coding in the context of attachment. One exception is research conducted by Zeng and colleagues in 2006, in which participants taking the AAI were video-recorded and the video was automatically coded using the Adaboost algorithm (Zeng et al., 2006). That study concluded with a recommendation for facial expressions to be included as an additional dimension of AAI coding, their results showing that facial expressions revealed precious information not otherwise contained in verbal transcripts.

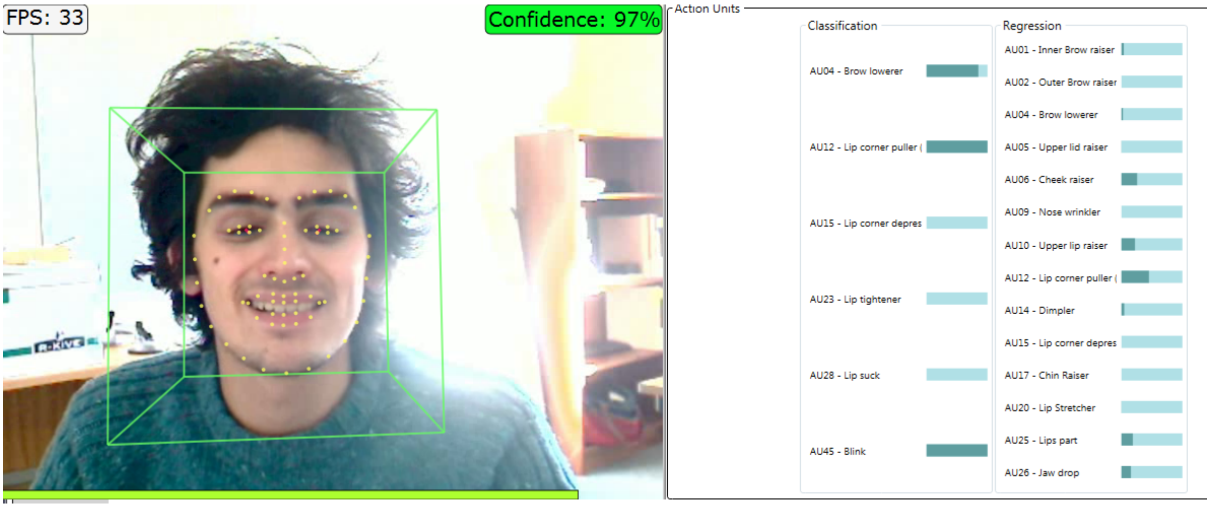


Figure 11: Openface detecting FACS AUs in a participant

During the BAT test, every time a theme stimulus is shown for 15 seconds, the software implements Openface algorithms to capture facial expression changes as a direct reaction to the stimulus.

4.5.7 Linguistic characteristics

Linguistics as a field is a treasure for clinical and developmental psychology. In their systematic review of 10 years of association between childhood language patterns and childhood psychopathology, Toppelberg and colleagues concluded that grammar, semantics, and pragmatics (i.e. the action conveyed through what's being said) overlap significantly with psychiatric disorders, and that language disorders and delays were risk factors for developing psychopathology (Toppelberg & Shapiro, 2000).

The "gold standard" adult attachment test, the AAI, takes advantage of specific patterns and properties of the language adults use to retrospectively describe their early attachment experiences in order to infer attachment characteristics (Buchheim & Mergenthaler, 2000; Cassidy, Sherman, & Jones, 2012). The PACS, a recently developed adult attachment measure, goes further by utilizing pragmatic linguistics analysis to detect phrasing implying a distancing/rejecting action from language implying an approaching action, as a proxy to infer attachment characteristics in any transcribed interview, not only those speaking about attachment-related issues (Talia, Miller-Bottome, Wyner, Lilliengren, & Bate, 2019).

When developing the BAT, it was evident to us that we must include the linguistic dimension in our analysis. But to do so, important obstacles needed to be surmounted.

First, the AAI, the PACS, the AAP, and all other interview-based attachment tests rely on human transcriptions of interviews that are later scored. Obtaining these transcriptions is not trivial, and often times that procedure itself takes up to seven times the duration of the interview, even for trained professional transcribers. As we mentioned in section 4.5.1, a design principle of the BAT is that all analyses shall be immediate, automatic, and computer-based. How, then, to conduct linguistic analysis without human transcriptions of the participants' responses to the stimuli?

Enter Automatic Speech Recognition or ASR. ASRs are automatic transcription systems based on machine learning algorithms trained on massive sets of previously transcribed speech and their human annotator transcriptions. The most powerful ASR to date is Google's Cloud Speech model (Halpern et al., 2016). Ziman and colleagues examined its performance by comparing it to human annotator transcripts and found them to be matched at a high degree, exhibiting similar statistical properties. The study concluded that ASR's were ready to be implemented as a "[...]cheap, reliable, and rapid means of automatically transcribing speech data in psychological experiments" (Ziman, Heusser, Fitzpatrick, Field, & Manning, 2018).

In addition to these advantages, which are essential for the BAT, ASRs such as Google's are usually trained in a variety of languages. In the specific case of Google Cloud Speech, it can transcribe speech in different languages, and even in different dialects, adding up to a total of 120 language-dialect combinations (e.g. American English; Hong Kong Chinese Cantonese Traditional; Canadian French; Puerto Rican Spanish), all with similar transcription performance (Halpern et al., 2016).

The second challenge is related to the error rate. Even the best ASR available to date still makes transcription errors: the Google Brain team's research demonstrated a state-of-the-art word error rate of 18% for clinical conversations (Kodish-Wachs, Agassi, Kenny, & Overhage, 2018). In our empirical experimentations, we quickly realized that complex linguistic analyses, the likes of the pragmatic analysis implemented by the PACS, would require full-phrase investigation in order to uncover the subtle actions intended with a given utterance. The success rate of such analysis would be seriously handicapped by that 18% word error rate: one or two words misunderstood in a 10-words-long phrase could change the full meaning of the entire phrase.

However, other powerful linguistic analyses were still possible using vocabulary instead of full phrases as the object of analysis.

Pennebaker's Linguistic Inquiry and Word Count approach is based on specially tailored word dictionaries that separate words into 64 psychologically meaningful categories (for a full list and some word examples, see Table 2). Empirical results using LIWC have demonstrated its capacity to differentiate attentional focus, emotionality (positive vs. negative), social relationships, status, dominance, and social hierarchy, social coordination and group processes, honesty and deception, close relationships, and thinking styles (Tausczik & Pennebaker, 2009).

Another advantage of LIWC dictionaries for the BAT is that they have been professionally translated and validated in many different languages, including Spanish, English, French, Brazilian and Portugal Portuguese, Dutch, German, Italian, Russian, simplified and traditional Chinese. Together with Google Cloud Speech's capacity for polyglotism, this empowers the BAT to function in very different languages all the while conserving equivalent linguistic features.

CATEGORY PSYCHOLOGICAL PROCESSES	EXAMPLES	CATEGORY STANDARD LINGUISTIC DIMENSIONS	EXAMPLES
Social Processes	talk, us, friend	Pronouns	I, them, itself
Friends	pal, buddy, coworker	Articles	a, an, the
Family	mom, brother, cousin	Past tense	walked, were, had
Humans	boy, woman, group	Present tense	Is, does, hear
Affective Processes	happy, ugly, bitter	Future tense	will, gonna
Positive Emotions	happy, pretty, good	Prepositions	with, above
Negative Emotions	hate, worthless, enemy	Negations	no, never, not
Anxiety	nervous, afraid, tense	Numbers	one, thirty, million
Anger	hate, kill, pissed	Swear words	fuck, shit, asshole
Sadness	grief, cry, sad	PERSONAL CONCERNS	
Cognitive Processes	cause, know, ought		
Insight	think, know, consider	Work	work, class, boss
Causation	because, effect, hence	Achievement	try, goal, win
Discrepancy	should, would, could	Leisure	house, TV, music
Tentative	maybe, perhaps, guess	Home	house, kitchen, lawn
Certainty	always, never	Money	audit, cash, owe
Inhibition	block, constrain	Religion	altar, church, mosque
Inclusive	with, and, include	Death	bury, coffin, kill
Exclusive	but, except, without	SPOKEN CATEGORIES	
Perceptual Processes	see, touch, listen		
Seeing	view, saw, look	Assent	agree, OK, yes
Hearing	heard, listen, sound	Nonfluencies	uh, rr
Feeling	touch, hold, felt	Fillers	blah, you know, I mean
Biological Processes	eat, blood, pain		
Body	ache, heart, cough		
Sexuality	horny, love, incest		
Relativity	area, bend, exit, stop		
Motion	walk, move, go		
Space	Down, in, thin		
Time	hour, day, oclock		

Table 2: LIWC categories and sample words.

LIWC has been used in numerous psychological studies examining language patterns correlates to clinical conditions (e.g. (Junghaenel, Smyth, & Santner, 2008)).

Of particular interest for the BAT is the study by Cassidy et al. that analyzed AAI transcripts using LIWC, finding that participants with secure, dismissing, and preoccupied AAI classifications significantly differed in their use of 14 of the 44 LIWC categories examined. The study also found that 10 LIWC categories were significantly correlated with AAI coherence of mind, and most strikingly, it discovered that AAI group assignment based on automatically-calculated LIWC linguistic profiles yielded 71% agreement with AAI traditional codings (Cassidy, Sherman, & Jones, 2012).

The BAT uses Google Cloud Speech to automatically transcribe the participants' verbal responses to the question "What did you feel?" that is shown after each of the 14 BAT stimuli. Once the response is transcribed, linguistic analysis is automatically performed on the transcript using LIWC. The second version of the BAT (v2) is able to conduct this analysis with French, English, and Portuguese speech. This could be easily extended to any other available LIWC dictionary.

4.6 Scoring the BAT: from human coding to multimodal fusion

In a clinical setting, attachment is often assessed through the therapeutic relationship. The therapist tends to play the role of a temporary attachment figure for the patient and the latter's attachment anxiety and avoidance tend to arise and play out in the relationship (Brown & Elliott, 2016). Like the expert evaluators in the SSP, the clinician directly examines complex patterns of behavior: a phrase being repeated (linguistic analysis) in a specific tone of voice (paralinguistic features), the physical distance the patient chooses for their chair with respect to the therapist's (distance from stimulus), the fact that they keep looking away the whole time (head pitch/yaw), the combination of all these behavioral features might hit the therapist as their patient relationally "pushing them away" (i.e., attachment avoidance). Clinician assessments offer an undeniable holistic advantage over unidimensional tests like the AAI that are limited to the evaluation of transcripts of speech, but clinician assessments suffer from subjectivity, and different clinicians with different sensibilities may judge the attachment characteristics of a client differently.

As we have just seen, the BAT mimics this clinical holistic advantage, analyzing many of the same sensory modalities, but in a more objective (i.e. replicable) manner. In addition, clinician assessments, good as they might be, miss the very earliest of attachment-related reactions: autonomic nervous system responses. Several studies suggest that physiological responses are sometimes the only way of differentiating among attachment characteristics, for example for separating a secure response from a dismissing (i.e. high in avoidance) one. Both a secure and a dismissing person might show a relatively calm facial expression while being reminded of attachment-related content (e.g. last separation), and yet the dismissing person will tend to additionally show an elevated physiological level of distress (Powers, Pietromonaco, Gunlicks, & Sayer, 2006). The BAT obtains precise physiological information to analyze in conjunction with all the other extracted behavioral features to produce a multifaceted assessment.

But the question remains: once all those analyzes are conducted and their results are available, what exactly should the BAT do with them? How exactly should it interpret such wide-ranging results coming from such different modalities, in response to 14 different themes? How are the number of pronouns, the timing of blinks, the high-frequency band of HRV or the acoustic formants of a phrase in the response to a stimulus of an unattuned mother-child dyad to be interpreted?

At the onset of our work, we thought of conducting an exhaustive literature review on each feature of each modality in relation to attachment related themes and based on prior findings, to "guesstimate" expected composite scores for each theme depending on our experimental objectives (see section 4.4.2). Such an approach quickly revealed too colossal to undertake. For one, each modality (e.g. facial expressions) is composed of a plethora of features (e.g. AUs 1 to 45), some of which on their own don't mean much: what hypothesis could be made about the expected Dimpler face muscle action as a response to a theme evoking couple separation? In isolation, many or even most of these features don't mean much and little can be thus be hypothesized about their expected behavior. Additionally and even if there would be enough prior work to build hypotheses for all features, the sheer number of them would render this all but impossible to accomplish: how long would it take to develop reasonable hypotheses for the expected behavior of 19 facial action units, 4 dimensions of head movements plus 2 dimensions of gaze direction, 16 paralinguistic features, 64 linguistic categories, and dozens of HRV and EDA features in response to each of the 14 stimulus, for each of the adult attachment patterns? And finally, what if we wanted the BAT to output continuous dimensional scores? In that case, our hypotheses would not only need to discern among attachment patterns but should provide expected equations that would integrate all the aforementioned features in a way that the equation's output corresponds to adult attachment dimensions.

Enter multimodal fusion. Multimodal fusion is a machine learning technique that enables the creation of non-linear models for the optimal integration of data coming from different sensors or modalities in order to predict a given outcome (Poria, Cambria, Bajpai, & Hussain, 2017). It is said to imitate the way the human brain works, by combining different sources of information about the same phenomenon in such a way that each channel of information adds detail and different perspectives, allowing for a better judgment about the phenomenon than what could be made based on information from one channel alone.

For instance, an isolated phrase might be suggestive of anger, but uttered in a certain tone and accompanied by a certain facial expression and body posture it might suggest lightness and humor.

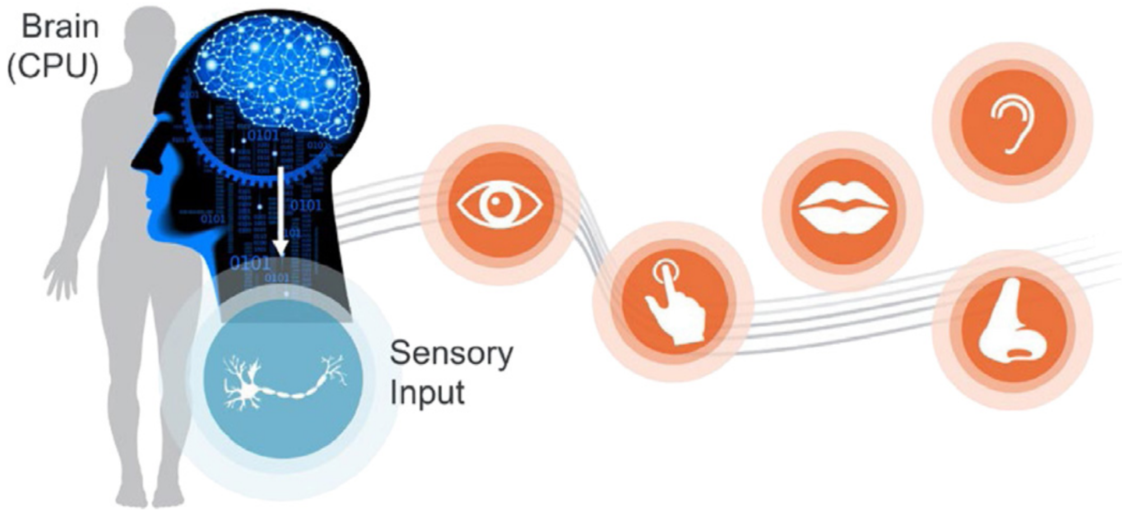


Figure 12: Reproduced from (Poria, Cambria, Bajpai, & Hussain, 2017). Multimodal fusion imitates the human brain, which considers information from several senses simultaneously for decision making.

As an experiment, one can alternate between using smell, taste, sight, and touch to recognize different fruits in a basket, and then replicate the experiment using only smell, and compare the recognition accuracy of both attempts. Each sense modality adds information to our fruit model and allows us to differentiate among them more accurately. The same is true for artificial intelligence: when algorithms base their decisions on several channels of information, they can be more performant.

As we'll see later in our second article, an original multimodal fusion procedure had to be invented for the integration of all the channels of information the BAT implements.

While the subject might seem too technical at first, the processes involved in both human perception and multimodal fusion highly overlap, and thus studying or modeling the latter is a remarkable source of insight about the former.

There is an additional advantage of using machine learning and multimodal fusion as the BAT's last step to learn to combine all its features and produce final scores: flexibility. The BAT multimodal fusion algorithm can be retrained to produce scores based on any existing attachment measure. Like a human student learning to score an interview-based assessment, mastering the linguistic criteria that better describe each attachment pattern, etc., using machine learning the BAT "learns" how to better combine its available modalities to better approximate a provided score. Specifically, testing a group of people with both the BAT and another validated measure and offering this data to the multimodal fusion algorithm, it will create a model for the combination of all BAT features that better predicts the scores found by the target test. That can be repeated with different validated attachment measures, and as a result, the BAT, a 9 minutes test, can be able to produce different scores that should be convergent with each of them, so that a participant taking the BAT can have access to several complementary but different results. This concept will be better clarified in our third article in the next section.

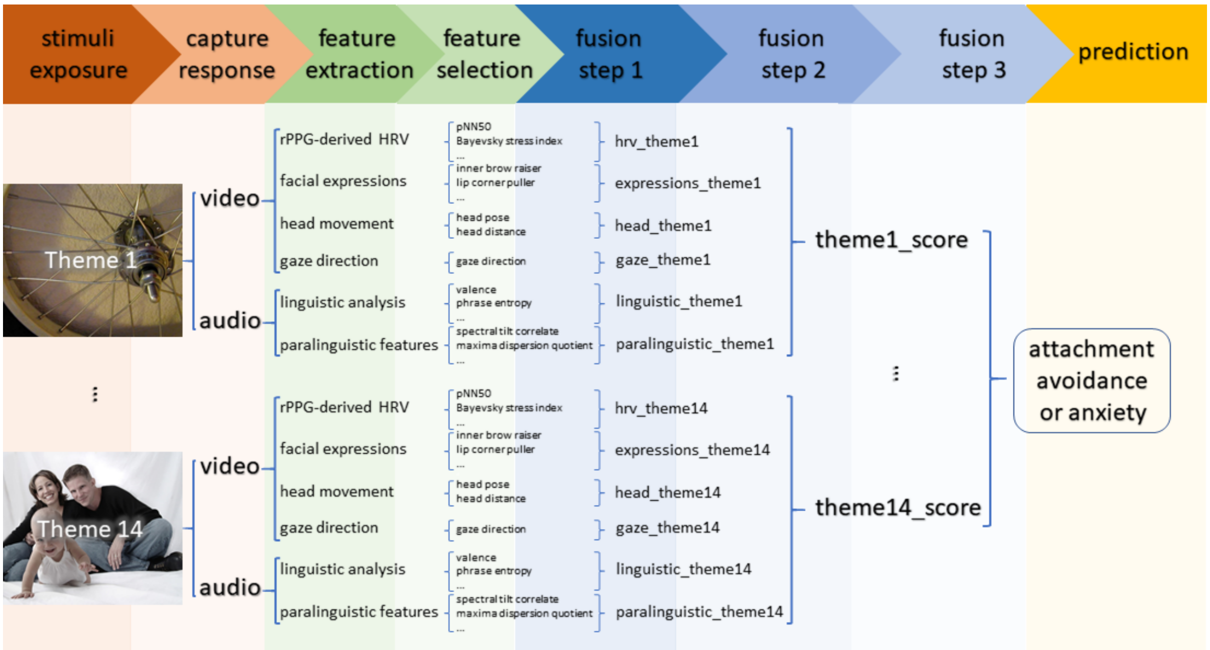


Figure 13: Our original three-step multimodal fusion procedure (see section 5.2)

4.7 Cross-modal plasticity and the BAT: towards BAT v3

Cross-modal plasticity is the term coined to describe the reorganization of neural networks that take place in humans and some animals to adapt to sensory deprivation, such as vision loss (Bavelier & Neville, 2002). More recently, it has been shown that this process occurs even during short-term sensory deprivation, the likes of being blind-folded during a task (Lazzouni, Voss, & Lepore, 2012). When humans are asked to identify an object without using their sight, the brain increases the sensibility and integration of all other available senses, such as touch, hearing, and smell, improving the chances of correctly identifying the object.

During the last few months of work that led to our third article (see section 5.3), we were faced with a dilemma: the BAT v2 scoring algorithm was incapable of scoring about 50% of our participants due to missing or erratic sense modalities data for one or a few of the BAT themes. An example of a missing modality was a lack of paralinguistic features for a given theme due to the person being distracted and silent during that particular theme. An example of erratic data was impossible (i.e., not human) heart rate values for a given theme, the outcome of a sudden change in lighting conditions in the room affecting the RPPG algorithm. In either case, whether this happened during one theme or many, the BAT v2 scoring algorithm would not be able to produce a score. If missing/erratic values were replaced by zeros, the algorithm would succeed in producing a score, but those scores would not be convergent with the RSQ, our reference measure for our third study (see section 5.3).

Inspired by cross-modal plasticity, we decided to evolve our three-step multimodal fusion algorithm (this algorithm is introduced in-depth in our second article, section 5.2) to try and imitate the way the human brain deals with missing sense modalities.

Without getting into computer science technical details that would not make sense here, we would like to describe the barebones of the mechanism responsible for the BAT's success during our third study.

The process begins at the second step of our three-step multimodal fusion procedure, which is in charge of combining different modalities (e.g., paralinguistic characteristics, HRV) per theme, into a single score per theme that integrates them all.

Our new code begins by identifying, on the training sample (i.e., the sample it will use to learn how to score according to a given measure), all the sense modalities per theme per participant that either are missing or have erratic data. For example, this analysis might find that for participant number 72, during theme 4, for some reason the heart rate data became erratic and thus is not trustworthy.

Once all such cases have been marked, the algorithm separates out participants for which there were no errors or missing data at all. Using only those participants, the program trains a "best-case scenario" scoring model.

Then, it roams through each modality per theme, this time grouping all participants that had an error or missing data on the same modality and theme, and that did not have errors on any other modality in the same theme. For example, it might group together 15 participants for which the facial expression data in theme 5 is missing or erratic, but for which all other modalities in theme 5 are fine. The program will proceed to train a different scoring model for every single modality that might be erratic or missing in each theme.

The end product is not a unique scoring model. Rather, it's more like a tree of scoring models, branching out into different possible models that are better for addressing each particular error case. We call this group of models a cross-modal scoring model.

When this cross-modal scoring model is applied to a new participant's data, it will first identify, per theme, if there are missing or erratic modalities. It will then automatically assign to each theme a scoring model (a branch from the tree) that has been trained to predict attachment without the missing or erratic modalities.

In the current code, this process is limited to one missing or erratic sense modality per theme. For example, if for theme 8 a participant is missing both facial expressions and linguistic data, the program would not be able to score that theme.

However, we implemented a similar procedure within the third step of our three-step multimodal fusion procedure. The third step is where the BAT learns how to integrate scores from all different themes into a final score. Once again, we train a "best-case scenario" scoring model using participants who miss no theme scores. Then, we train "branch" models with participants for which one theme is missing until all themes are covered. As a result, the BAT disposes of different scoring models to face the situation where one theme's score is not available. The limitation of this procedure is that it cannot deal with the case in which more than one theme is missing - but we decided this is also a good limit on how much error the BAT should tolerate. If more than one theme is missing, it means that more than one modality was missing in more than one theme. It might be better, in that case, to administer the BAT again.

This qualitative leap in our algorithm's functionality defines the third version of the BAT.

When cross-modal scoring models are implemented in both steps two and three of the three-step multimodal fusion procedure, the BAT becomes able to deal with common error factors such as human distraction, abrupt changes in lighting conditions that might affect the RPPG algorithm or computer bugs.

Coming back to our third study, the BAT v3 was able to produce scores for almost all (96.6%) of our cases, and the predicted scores showed convergent validity and test-retest reliability, just like the scores produced for participants with no errors in their data (this technique of cross-modal scoring is called "resilience mechanism" in our third article, which could lead to confusion).

It is worth mentioning that this cross-modal mechanism is not simply taking a linear equation and eliminating from it its missing factors. The artificial neural networks that are implemented in both steps 2 and 3 of the multimodal fusion procedure in the BAT create different artificial neuron connections when a modality is absent or when a theme is not available, in a similar fashion to the process previously described for humans and animals.

5 Experiments (published and submitted articles)

5.1 The Multimodal Assessment of Adult Attachment Security: Developing the Biometric Attachment Test

Our first published article about the BAT, this paper responded to a call for papers from the Journal of Medical Internet Research (impact factor 4.94), for a special edition focusing on Computing and Mental Health, which was curated by the affective computing pioneer Rafael A. Calvo (Calvo, D'Mello, Gratch, & Kappas, 2015). The journal is known for its interdisciplinary emphasis. Upon publication, the paper was selected among a few to be presented orally and through a poster at the Computer and Human Interaction 2017 congress, in Denver, Colorado. We reproduce here the article, as well as the poster.

In this paper, we explored the relationships between the raw scores of all of the BAT's different modalities (e.g., facial expressions, EDA), and the attachment security score obtained by three validated measures of adult attachment, two of them interview-based. We also tested the hypothesis that different stimuli sets designed using our standardized procedure and survey would produce similar responses in participants.

Original Paper

The Multimodal Assessment of Adult Attachment Security: Developing the Biometric Attachment Test

Federico Parra[1,2], MA; Raphaële Miljkovitch[2], PhD; Gwenaelle Persiaux[3], MA; Michelle Morales[1,4], MS; Stefan Scherer[1], PhD

[1]Institute for Creative Technologies, University of Southern California, Los Angeles, CA, United States

[2]Paragraphe Laboratory, Paris VIII University, Saint-Denis, France

[3]University Hospital Center Sainte-Étienne, Sainte-Étienne, France

[4]The Graduate Center, City University of New York, New York, NY, United States

Corresponding Author:
Federico Parra, MA
Institute for Creative Technologies
University of Southern California
12015 E Waterfront Dr
Los Angeles, CA, 90094
United States
Phone: 33 782132695
Fax: 1 (310) 574 5725
Email: federico.parra@hotmail.com

Abstract

Background: Attachment theory has been proven essential for mental health, including psychopathology, development, and interpersonal relationships. Validated psychometric instruments to measure attachment abound but suffer from shortcomings common to traditional psychometrics. Recent developments in multimodal fusion and machine learning pave the way for new automated and objective psychometric instruments for adult attachment that combine psychophysiological, linguistic, and behavioral analyses in the assessment of the construct.

Objective: The aim of this study was to present a new exposure-based, automatic, and objective adult-attachment assessment, the Biometric Attachment Test (BAT), which exposes participants to a short standardized set of visual and music stimuli, whereas their immediate reactions and verbal responses, captured by several computer sense modalities, are automatically analyzed for scoring and classification. We also aimed to empirically validate two of its assumptions: its capacity to measure attachment security and the viability of using themes as placeholders for rotating stimuli.

Methods: A total of 59 French participants from the general population were assessed using the Adult Attachment Questionnaire (AAQ), the Adult Attachment Projective Picture System (AAP), and the Attachment Multiple Model Interview (AMMI) as ground truth for attachment security. They were then exposed to three different BAT stimuli sets, whereas their faces, voices, heart rate (HR), and electrodermal activity (EDA) were recorded. Psychophysiological features, such as skin-conductance response (SCR) and Bayevsky stress index; behavioral features, such as gaze and facial expressions; as well as linguistic and paralinguistic features, were automatically extracted. An exploratory analysis was conducted using correlation matrices to uncover the features that are most associated with attachment security. A confirmatory analysis was conducted by creating a single composite effects index and by testing it for correlations with attachment security. The stability of the theory-consistent features across three different stimuli sets was explored using repeated measures analysis of variances (ANOVAs).

Results: In total, 46 theory-consistent correlations were found during the exploration (out of 65 total significant correlations). For example, attachment security as measured by the AAP was correlated with positive facial expressions (r=.36, P=.01). AMMI's security with the father was inversely correlated with the low frequency (LF) of HRV (r=−.87, P=.03). Attachment security to partners as measured by the AAQ was inversely correlated with anger facial expression (r=−.43, P=.001). The confirmatory analysis showed that the composite effects index was significantly correlated to security in the AAP (r=.26, P=.05) and the AAQ (r=.30, P=.04) but not in the AMMI. Repeated measures ANOVAs conducted individually on each of the theory-consistent features revealed that only 7 of the 46 (15%) features had significantly different values among responses to three different stimuli sets.

XSL•FO
RenderX

Conclusions: We were able to validate two of the instrument's core assumptions: its capacity to measure attachment security and the viability of using themes as placeholders for rotating stimuli. Future validation of other of its dimensions, as well as the ongoing development of its scoring and classification algorithms is discussed.

(J Med Internet Res 2017;19(4):e100) doi: 10.2196/jmir.6898

KEYWORDS

psychometrics; linguistics; heart rate; facial expression; psychophysiology; psychopathology; COVAREP; attachment

Introduction

The Relevance of Adult Attachment in Mental Health

Attachment theory originated with the work of a British psychiatrist, John Bowlby [1]. Inspired by ethological observations and evolution theory, he theorized that the chance for survival of human genes had increased by the natural selection of behaviors that augmented proximity and bonding between infants and their caregivers, leading to a greater probability of protection for the children [1,2]. Attachment theory posits an innate psychobiological behavioral system, the *attachment system*, which activates specially in times of perceived threat, inciting the child to seek the proximity and care of their caregivers, the *attachment figures*. The system deactivates once a felt sense of security and safety is reestablished [1,3]. Despite the universality of attachment proximity-seeking behaviors in children [4,5], the security and care sought are only found when the attachment figures are capable of responding promptly and adequately [6]. The quality and outcome of these repeated early attachment interactions leave an enduring mark in the developing person [7-10]. The nature of this mark is threefold: it is cognitive, since dynamic representational models of the attachment figures and the relationship with them develop [11,12], contributing in adulthood to appraisals of the self as worthy of care and of others as capable of providing care [13]; behavioral, because our innate attachment behaviors accommodate to the environment [6], for example, in case of continuous unavailability of caregivers children might stop proximity-seeking behaviors entirely and act as if they were totally independent, a pattern that is then carried into adulthood [14]; and psychobiological, because negative early attachment experiences can lead to differences in the response of the bilateral amygdala and left ventral striatum during stressful situations, and to an overall higher sympathetic activation baseline [15-17].

The different adult attachment patterns have been extensively described in the literature using both dimensional and categorical models [2,18]. In the dimensional approach, the single most important dimension is *attachment security* [12].

Attachment theory sparked some of the largest and more rigorous longitudinal studies in psychology to date [8,19], proving itself essential in three overlapping research domains of mental health: the study of psychopathology, the study of psychological development, and the study of the psychology of adult interpersonal relationships.

In terms of clinical research, longitudinal studies have shown that negative early attachment interactions in childhood predict childhood attachment security [9], which in turn partially predict adult psychopathology [7,20], whereas cross-sectional studies have consistently linked adult *attachment insecurity* to several psychopathologies [21,22], such as depression [23], post-traumatic stress disorder (PTSD) [24], or borderline personality disorder [25]. Positive attachment experiences in adulthood, whether naturally occurring or the outcome of therapeutic interventions, can help increase attachment security, which in turn improves mental health [26-28].

In terms of developmental psychology, studies show that developmental competencies that are essential to sustain mental health and to cope with mental health disorders, such as emotional regulation, social skills, or cognitive ability, are associated and interdependent with attachment across the lifespan [7,29,30].

Finally, adult attachment is key in the psychology of interpersonal relationships, including long-term romantic relationships [31,32], which tend to function as a buffer in coping with psychopathology and stress [33,34]. Attachment insecurity has been associated with having more interpersonal problems in general [35], and these problems explain insecure persons' self-reported loneliness, social isolation, low relationship satisfaction, more frequent relationship breakups, greater physiological stress reaction to interpersonal conflict, and more frequent conflicts and violence [21,36-38]. Secure attachment, in relationship with social support, has been acknowledged as a protective factor for psychological stress [34], with perceived social support mediating the relationship between attachment security and depressive symptoms [33].

Current Limitations in the Assessment of Adult Attachment

Since 1985 (Findings by George C, Kaplan N, and Main M, unpublished data, 1985), various validated instruments for the assessment of adult attachment developed concomitantly within the fields of social psychology and developmental psychology (for a review, see [39]). Social psychology has spurred the development of several questionnaires, such as the Adult Attachment Questionnaire (AAQ [40]) or the Adult Attachment Scale (AAS [41]). Developmental psychology, on the other hand, has relied on a variety of broadly defined semistructured interview methods, beginning with the Adult Attachment Interview (AAI; Findings by George C, Kaplan N, and Main M, unpublished data, 1985) which is considered the "gold standard" [39]. Both approaches suffer from several limitations that affect both research and clinical assessments, and that are reflective of the current state of psychometrics.

Questionnaire-based assessments are self-report measures. As such, they are prone to self-report biases that have been well

XSL•FO
RenderX

described in the literature [42]. In terms of construct validity, there has been no longitudinal association demonstrated between attachment in childhood as measured for example with the Strange Situation Procedure (SSP [43]) and adult attachment as measured with questionnaires [19]. Furthermore, almost no concurrent validity has been found between questionnaires and interview-based assessments of adult attachment, adding to the construct validity controversy [39]. On the other hand, questionnaires of adult attachment are easy, economic, and fast to both administrate and score. Administration can be done remotely, and automatic scoring is possible. These positive practical psychometric characteristics may explain the surge of studies that have chosen questionnaires of adult attachment as their measure [39].

Interview-based assessments of adult attachment rely on some form of semistructured interview, which is later transcribed and scored by a trained judge, that has undergone substantial training in a specific standardized scoring tradition. In a way, this form of assessment is closer to child assessments which also rely on third-party experts for scoring and classification. However, in childhood-attachment assessments such as the SSP, the scoring experts observe behavior in general, whereas in interview methods only transcribed language is analyzed during scoring, thus limiting the scope of dimensions evaluated in this process. The Attachment Multiple Model Interview (AMMI [12]) circumvents this limitation in part, by including in the interview specific probes to gather self-reported information about behaviors.

In terms of construct validity, the AAI has consistently shown a link between parents and their children's attachment patterns, which is considered strong evidence of its validity [2]. Moreover, a substantial longitudinal link has been found using the AMMI [12], further supporting the consensus that interview methods based on expert judgment can produce results with higher construct validity than self-report measures.

But despite this consensus, interview methods are not without their own limitations. In contrast to their questionnaire-based counterparts, interview methods are difficult, costly, and lengthy to both administer and score. They add additional layers to the process, that is, the manual transcription and coding of the interview. There is a training load required for both administrating and scoring. This process is costly.

Interview methods are impacted by an additional problem: the subjectivity inherent to an expert judge [44]. This limitation might decrease the replicability of attachment studies, adding to psychology's current "replicability crisis" [45].

Finally and contrary to questionnaires, interview methods cannot be administered remotely, limiting their application, for instance, in Internet-based research.

Advances in Multimodal Analysis and Automatic Detection of Psychological Markers

Finding psychophysiological and behavioral markers of psychological conditions is gaining traction within mainstream psychiatry [46,47], as part of a quest to provide more objective and precise clinical assessments to patients. The American National Institute of Mental Health released a statement in 2013 [48] in which it made explicit its desire of moving toward more objective and precise diagnostic methodologies. Several attempts to tackle this problem have arisen from the Computer Sciences. In a recent review, Cummins et al [46] reviewed the state-of-the-art in the automatic detection of depression and suicidality through the analysis of speech and its paralinguistic acoustic features. Scherer et al [49] described, in 2013, a set of automatically extracted audiovisual nonverbal behavioral features helpful in the identification of depression, anxiety, and PTSD [49]. The link between the objective measure of singular biometric or behavioral markers, and the sought ability to offer more precise diagnoses, relies on the use of machine learning algorithms that can fusion multiple modalities of data at once [50]. This allows for the uncovering of complex multimodal data patterns that can serve in the automatic assessment of specific mental conditions. In recent studies, such multimodal systems have approached human performance in the detection of indicators of PTSD [51]. Since several studies on the specific psychophysiological [2] and linguistic [44] traits of adult attachment already exist and show promise, we decided it was time to use this new technology in the assessment of adult attachment.

The Biometric Attachment Test

The Biometric Attachment Test (BAT) was created with the objective of automatically and objectively measuring attachment in adults. At its core, the BAT is an exposure-based test, which means that the participant being tested is exposed to a short (9 min) standardized set of visual and music stimuli, whereas their immediate reactions and verbal responses, captured by several computer sense modalities, are automatically analyzed for scoring and classification.

There are two aspects of the development of the BAT that require separate attention: the instrument itself, meaning its assumptions, stimuli selection, and administration protocols, which will be articulated in this work; and the test's automatic classification and scoring algorithms, a work-in-progress that we will briefly touch upon in "Discussion" section.

Construction of the Biometric Attachment Test (BAT)

The BAT was strongly influenced by three previous instruments: Bowlby's first Separation Anxiety Test (SAT [52]), the previously mentioned SSP [43], and the AAP [13].

The SAT (1976 version [52]) is a projective attachment test for children aged 4-7 years consisting of a set of 6 pictures depicting situations in which a child, separated from their family, must cope on their own without help from their parents. The tested child is asked to interpret the protagonist's feelings and predict their behavior, and their transcribed responses are later scored and classified.

The SSP is a structured observation protocol for assessing attachment in children aged between 12 and 24 months. During 20 min, the child undergoes a series of separations and reunions from their caregiver, while they are also exposed to the arrival and presence of a stranger. The child's behavior is videotaped and then analyzed for attachment scoring and classification.

XSL•FO
RenderX

The AAP is an adult attachment test based on a set of black and white drawings, some of which are ambiguous, depicting more diverse situations that activate the attachment system: separation, loss, solitude, and physical threat [13]. Participants are asked to tell a short story about the pictures, which are transcribed and analyzed, and an attachment classification and continuous scores are obtained [13].

Like the SAT, our BAT uses photos, of real people, in explicit situations. Like the AAP, our stimuli depict a variety of attachment-sensitive situations. Like the SSP, the BAT is meant to produce an *alternating activation and deactivation* of the attachment system, with stimuli representing themes such as loss, death, or separation alternating with stimuli representing themes such as intimate connection, soothing, or protection.

Unlike other exposure-based and projective tests, the BAT uses music stimuli in addition to visual stimuli, both on its own and concomitantly with visual stimuli. Music was included because of its ability to trigger strong emotional feelings and experiences [53].

Like the SSP, scoring and classification in the BAT take into consideration observed behaviors. In fact, unlike other tests in which verbatim transcripts of verbal responses are analyzed, the BAT captures the participants' reactions and responses in a variety of modalities: physiological (heart rate [HR] and electrodermal response [EDA]) from which psychophysiological features can be derived (eg, Bayevsky stress index [54]), behavioral (facial expressions, gaze, face distance from stimuli, paralinguistic speech characteristics), and verbal.

The Concept of Themes in the Biometric Attachment Test (BAT)

Exposure-based and projective psychometric tests typically rely on a fixed set of stimuli selected by the authors [13,55]. We pose the following critiques to this approach: first, stimuli can eventually leak into the public domain, such as in the case of the Rorschach [56], and this might undermine a test's effectiveness due to priming effects. Second, longitudinal studies such as clinical trials require participants to be tested several times using the same instruments, and if stimuli are always the same this might also lead to priming effects. Finally, we believe ideally stimuli should be selected based on input from the general population toward which it is destined.

Our BAT innovates introducing the concept of *themes*: placeholders for actual stimuli. A theme is a narrative that describes a specific situation to be evoked by a stimulus, with a specific objective. For example, in terms of adult attachment, a theme could be "the loss of a close one," its objective being to activate the attachment system (ie, to cause attachment-related distress).

Themes thus can solve the aforementioned problems with fixed-stimuli test designs: since themes are placeholders for stimuli as opposed to fixed stimuli, there is no risk if a stimuli set becomes widely known. All the contrary: stimuli in the BAT can—and should—be replaced from time-to-time and from context-to-context. In the case of clinical trials, stimuli sets in the BAT could rotate between assessments. Finally, the process for stimuli selection in the BAT is standardized and crowdsourced, as we will see briefly.

About the themes' objectives, each is meant to evoke a reaction in the participant depending on the participant's attachment patterns. The themes were inspired by the SAT, the AAP, the SSP, and attachment theory core principles. In total, 14 themes resulted from this work (see Figure 1).

Theme 1 ("baseline") was designed to measure the participants' reactions to being in the test situation, where they are still not being confronted with any attachment-specific stimulus. This provides proper baselines for all biometric and behavioral measures.

Themes 2, 8, and 10 were designed to elicit specific reactions depending on the underlying attachment pattern of the test participant, to help in classification.

All other BAT themes have per objective to either activate (ie, stress) or deactivate (ie, calm) the attachment system. We would like to clarify that throughout this paper we use the terms "attachment activation" and "attachment deactivation" in their literal sense, that is, the way in which the attachment system is activated when under specific relational stress and how it becomes deactivated when that relational stress is sufficiently addressed. This is not to be confused with "avoidant deactivation," a "Minimizing strategy (...) conceived by Main (1990) as a shift of attention away from conditions normally eliciting attachment behavior, leading to the apparent absence of attachment behaviors in such circumstances" [12].

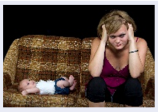
Stimuli Selection in the Biometric Attachment Test (BAT): A Standardized Process

A set of objective and subjective criteria were developed for each of the BAT's themes. The objective criteria were directly derived from the themes' narratives: for example, for a stimulus to be appropriate to represent the "attuned mother-child" theme, there should be a mother and a child in the picture. Subjective criteria are notions that require more complex judgments: for example, for a stimulus to be appropriate to depict the "attuned mother-child" theme, the child and the mother must seem attuned to each other and, thanks to said attunement, they should both seem relatively relaxed. To decide whether a mother or a child seem relaxed or not just by looking at them in a picture is a subjective process that should not be arbitrarily decided by researchers.

We used the straightforward objective criteria to preselect stimuli: three large picture databases conceived for the study of emotion were used: the Nencki Affective Picture System (NAPS [57]), the International Affective Picture System (IAPS [58]), and the Geneva Affective Picture Database (GAPED [59]). In some cases, none of these databases had enough pictures for some of the themes, so we turned to a stock picture service, iStockPhoto. We ended up with 126 preselected pictures.

XSL•FO
RenderX

Figure 1. Biometric Attachment Test (BAT) themes, goals, and stimuli set sample.

Theme	Theme name	Objective	Stimuli sample
1	Baseline	-	
2	Child solitude	Depends on attachment pattern	
3	Couple separation	Attachment activating or stressing	
4	Attuned father-child	Attachment deactivating or soothing	
5	Violent father	Attachment activating or stressing	
6	Unresponsive mother	Attachment activating or stressing	
7	Attuned couple	Attachment deactivating or soothing	
8	Exploration	Depends on attachment pattern	
9	Raw sadness	Attachment activating or stressing	Sad music
10	Vulnerable baby	Depends on attachment pattern	
11	Raw fear	Attachment activating or stressing	Tense music
12	Death or loss	Attachment activating or stressing	
13	Attuned mother-child	Attachment deactivating or soothing	
14	Family connection	Attachment deactivating or soothing	

The Web-Based Survey

We then created an anonymous web-based survey using SurveyGizmo services. The survey randomly introduced each of the preselected pictures, accompanied by sliders that participants could adjust to the right or to the left, signaling a judgment about a specific criterion. In the case of the "attuned mother-child" theme, for example, one of the available sliders allowed the participant to judge the perceived level of stress of the child in the picture. We always opposed two traits (eg,

XSL•FO
RenderX

stressed vs relaxed), randomizing their order and starting with the slider in the center among them.

Our survey was made available in Spanish, English, and French and was distributed through social media and email campaigns in the United States, France, and Argentina. A total of 520 participants (female=72.3%, 376/520, male=27.7%, 144/520), of a variety of ages (mean 37.53, SD 10.87) responded. The survey was kept online for a period of 10 days between March 3, 2016 and March 13, 2016. Results where then cleaned-up using standard survey results cleaning best practices [60].

We created composite scores formulas for each of the 14 BAT themes, allowing to combine the subjective criteria measured in the survey. For the "attuned mother-child" theme, for example, the composite score formula was composed by the perceived level of genuineness of the picture, plus the perceived attunement between mother and child in the picture, minus the perceived levels of stress in the child and in the mother in the picture, individually.

A minimum required composite score was set for each theme to prevent stimulus that are not evocative enough from being used in the future.

The list of themes, objective and subjective criteria, as well as the survey design, are available for other researchers to generate new stimuli sets for the BAT in the future (contact corresponding author).

Music selection was easier and did not require a survey process. Music themes were conceived to convey basic raw emotions (eg, theme 9, "raw sadness"). A total of 25 short music clips were selected from a music set conceived to elicit emotion and that already provides scores on discrete perceived emotions [61]. We simply chose the music clips with higher scores in the required emotion per theme.

Biometric Attachment Test (BAT) Administration Procedure

The BAT automatic administration procedure was constructed using OpenSesame software, version 3.1.2 [62]. The full test duration is of 9 min.

Before beginning, the participant is instructed to observe the visual or listen to the music stimuli as long as it is visible or audible, and then to describe aloud what they felt about it.

During the test, each theme stimulus is automatically presented for 15 s, followed by a black screen displaying the phrase "What did you feel?" which shows for 25 s, whereas a countdown slider displays the available time to respond. It then shows the phrase "Thank you. Here is the following stimulus...," for 5 s, followed by the following stimulus, and so forth.

The test is administered with the person being alone with the computer in a room; aloneness can facilitate the activation of the attachment system [1] and renders the test situation closer to Ainsworth's SSP [6]. It also removes possible interference from researchers.

Hypotheses of This Study

This study was designed to empirically evaluate two core assumptions of the BAT:

H1: Those adults with higher attachment security will more successfully use the BAT's attachment-deactivating themes to reassure and soothe themselves, and this will be in turn reflected in specific psychophysiological, behavioral and linguistic markers. Theme 4 ("attuned father-child"), 7 ("attuned couple"), and 13 ("attuned mother-child") are evocative of the availability of attachment figures and will be used to test this hypothesis.

H2: Different stimuli sets, selected through our standardized process, are interchangeable and cause very similar responses or reactions in participants. Specifically, the features most associated with attachment security will remain consistent across three different stimuli sets.

Methods

Sample

The sample consisted of 59 French francophone participants (45 females, 14 males) that were interviewed between March and May, 2016. The sample was formed from multiple sources in different regions of France: 9 psychiatric patients recruited at University Hospital Center Sainte-Étienne and 7 recruited at the Ville Evrard Center of Psychotherapy in Saint Denis; 29 volunteers enrolled in Mornant, Paris, and Rouen; and 14 college students enrolled at Paris 8 University in Saint Denis. It was intended for the sample to be as diverse as possible in terms of age (mean 35.7, SD 12.2), occupational status (10% unemployed, 6/59, 32% employed, 19/59, 33% students, 20/59, 23% other, 14/59), as well as relationship status (37% in a relationship, 22/59, 23% married, 14/59, 11% separated or divorced, 7/59, 25% single, 14/59, 3% unknown, 2/59) and psychopathology (27%, 16/59 were patients). Since questions about ethnicity or race are not allowed in French research, we don't have information to report about the ethnic diversity of our sample. All participants signed informed consent forms in accordance to best practices in French Universities.

Measures

Adult Attachment Questionnaire

Fifty of our participants completed the AAQ before the interview, as a web-based questionnaire. The AAQ is a 17-item measure that asks individuals to indicate how they relate to romantic partners in general. It yields a continuous measure of attachment security with regards to romantic partners [40].

Adult Attachment Projective Picture System

All our participants completed the AAP test, which was introduced earlier. Transcripts of the AAP were scored by a trained member of our team blind to all information about the participants. Interjudge reliability was obtained for 5 cases that were double-coded by one of the AAP's creators, with 80% of interrater agreement for both classifications and scores. The AAP outputs a continuous attachment security score, called "agency of self," which per George [13] has both an inward and outward aspects. For this study, we'll focus on the latter, which

XSL•FO
RenderX

evaluates the degree to which an individual seeks for, and trust, attachment figures to provide for them a haven of safety in times of stress [13].

Attachment Multiple Model Interview

The AMMI is a validated semistructured interview that investigates participants' reactions in attachment-related situations. By analyzing and scoring transcriptions of the interviews, AMMI provides scores for three different attachment relationships: attachment to mother, father, and partner [12]. Since each relationship requires a specific amount of interview time, not all participants were able to complete all the interviews: attachment to the mother was evaluated for 27 participants, attachment to the father for 23, and attachment to the partner for 17. Transcripts of the AMMI were scored by a trained member of our team. Six cases were double-coded by the AMMI's creator, and interrater reliability was satisfying (83% of agreement).

The aforementioned measures have been validated in several languages including English. Their French version was used during this study.

Biometric Attachment Test

In order to evaluate our second hypothesis, we produced three BAT stimuli sets for this study: two fixed ones (ie, that show the same stimuli each time they are used) and a randomized one (ie, that shows different stimuli each time it is used). We have used the results from the French subsample of the survey respondents (n=194) to select the best pictures for a French population. The higher ranked pictures for each theme were put together in a stimuli set; the second higher ranked pictures were put together in a second stimuli set; and the pictures ranked third, fourth, and fifth were used to create a third set that randomly chooses one of those pictures each time it is played.

All 59 participants were exposed to the first stimuli set, 41 of them were also exposed to the second set, and 50 to the third rotating-stimuli set. Sets were presented one after the other.

Physiological Measures

HR was measured using the photoplethysmography sensor of an Empatica E4 wristband device. The sensor's reliability has been established [63]. Like all heart sensors, the E4 is subject to artifacts produced by movement. Quality readings were obtained for 29 participants during the first BAT set, 19 during the second BAT set, and 9 during the third. Electrodermal activity (EDA), with a specific interest in skin conductance response (SCR), was measured using the EDA sensor of the Empatica E4 wristband device. Quality readings were obtained for all participants during all BATs. The EDA sensor's reliability has been tested by the manufacturer [64]. Deliberately choosing to use a wireless wristband to measure physiological signals allowed our participants a more natural experience during the test.

Video and Audio Recording

Video of the participants' faces was obtained through the computer's webcam (Microsoft Surface Pro 4) and their speech was recorded using a USB Microphone (Samson GoMic). Since the BAT stimuli were presented using the same computer, gazing toward the stimuli was almost equivalent to gazing in the direction of the camera, facilitating gaze tracking.

Feature Extraction

We conducted extensive feature extraction from each of the sense modalities captured during the BAT. All feature extraction procedures described below, including noise filtering processes, were performed programmatically without the need for human supervision.

The interbeat interval (IBI) was automatically calculated from the HR data by proprietary algorithms of the Empatica E4 research wristband [65]. The IBI files were cleaned of artifacts using Artiifact software, version 209 [66]. The same software was used for the extraction of heart rate variability (HRV) features (for a review of most standard HRV features, see [67]). We created a function in Microsoft Excel's Visual Basic for Applications version 7.1 to automatically calculate Bayevsky stress index [54] from the IBI files.

From the EDA data, SCR, phasic maximal activity, and tonic skin conductance features were extracted using LedaLab software version 349 [68].

From the video data, facial expressions (such as anger and contempt, as well as the composites negative, neutral, and positive) were extracted using FACET's Emotient [69]. A face size measure was extracted by the same software, which permits to establish the movement toward or away from the camera and thus the stimuli. Gaze direction was extracted using OpenFace [70].

From the audio recordings of the participants' responses, paralinguistic acoustic features were extracted using the Cooperative Voice Analysis Repository for Speech Technologies (COVAREP) version 1.2 [71]. They help identify a breathy, relaxed voice from a tense voice.

We used Python and the French language model of Google's Cloud Speech API to generate automated transcripts of all responses. We then processed the transcripts using Python and Linguistic Inquiry and Word Count (LIWC) French dictionary [72]. This dictionary is organized in 64 psychologically meaningful word categories. The frequency of each word category in the response to each theme was calculated, to be used as linguistic features.

Extracted features per theme were then treated in two different ways:

Subtracted baseline: results on the first theme (baseline theme) were subtracted from all other themes' results. In theory, the resultant score should be more individualized to each person's individual characteristics (eg, their specific mean HR baseline).

Subtracted previous theme: results on each theme were subtracted from the following one. In theory, the resultant score would isolate results from the exposure to the theme under analysis from the cumulative score due to exposure to all precedent themes (eg, the specific augmentation or decrease in mean HR when exposed to theme 7).

For many features, we further separated the reaction during exposure to the stimuli from the reaction during the verbal

XSL•FO
RenderX

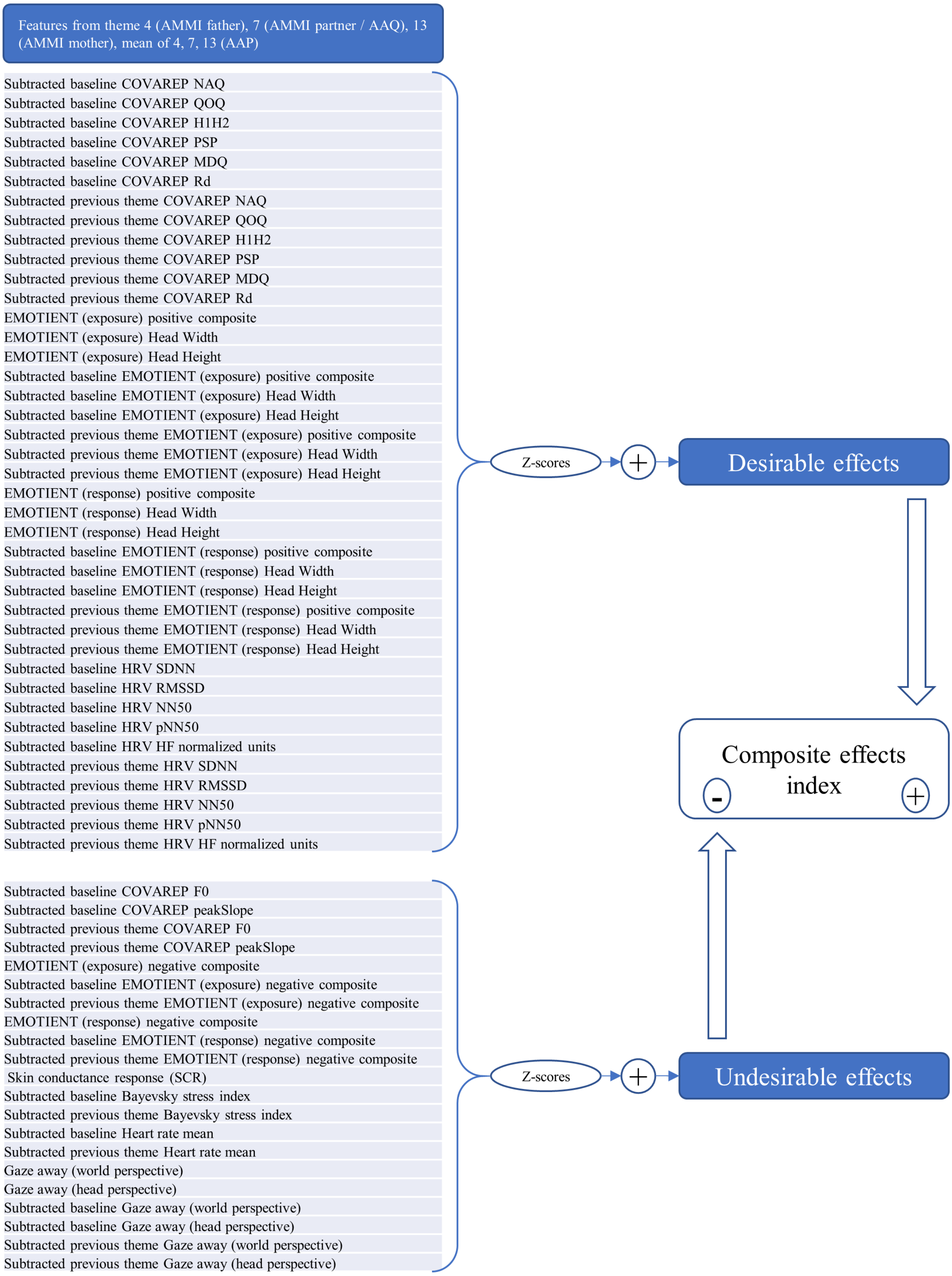
response to the stimuli, for example, facial expressions during exposure versus response.

Due to the high number of features extracted, the number of BAT themes, the two treatments we just described, and the separation between exposure and response, feature extraction led to a total of 4264 features per participant per stimuli set.

In this study, we will focus on specific themes instead of the entire stimuli set, and each theme has 202 features (see Figure 2). In total, 2436 features pertain to the entire stimuli set as opposed to any single theme (eg, total stimuli-set-wise mean HR).

XSL•FO
RenderX

Figure 2. Development of the composite effects index.



Analyses

A first exploratory analysis, conducted in MathWorks Matlab version R2016A, consisted of performing correlation matrices to uncover associations between attachment security as measured by the AAQ, the AAP, and the AMMI, and the features extracted from BAT responses. For the AAQ and the AAP, we used Pearson correlations, whereas Spearman rank was used for the AMMI due to the small number of assessed participants. Since AAQ measures attachment to romantic partners, it was evaluated

XSL•FO
RenderX

vis-à-vis theme 7 ("attuned couple"). AAP attachment security is concerned with attachment figures in general, thus we evaluated it with regards to a composite formed by the mean of responses to theme 4 ("attuned father-child"), theme 7 ("attuned couple"), and theme 13 ("attuned mother-child"). As per the AMMI, since it yields security scores for mother, father, and partner separately, we explored each in relationship to the corresponding BAT theme (themes 13, 4, and 7, respectively).

A second, confirmatory analysis was conducted using IBM SPSS Statistics version 23 to verify if our exploratory findings were not a mere spurious byproduct of multiple hypotheses testing [73]. We proceeded with a stringent approach consisting of producing a single "composite effects index" out of all available features (weighted in the same direction), then testing such index for a Pearson correlation vis-à-vis the variable of interest [74,75]. This approach circumvents the problem of type I errors often encountered in exploratory analyses. It also accounts for the problem of type II errors, which are likely when statistical correction procedures to control for family wise error rate (eg, Bonferroni correction) or false discovery rate (eg, Benjamini-Hochberg correction) are performed in studies with small samples, an elevated number of features, or small effect sizes such as ours [76-78].

For this analysis, our features' scores were first transformed into *z* scores. Next, they were added to either an undesirable effects group or a desirable effects group. The decision was based on the literature available on each of the set of features, with the following results: the high frequency (HF) component of HRV (associated with parasympathetic "relaxing" activation), HRV's SDNN, RMSSD, NN50, and pNN50 features (all of which convey slightly different aspects of the same, desirable construct: HRV), COVAREP features associated with a "breathy" relaxed voice, Emotient's "positive emotions" composite, as well as the head size (proximity of participant to stimuli source), were all summed up within a desirable effects group. On the other hand, Bayevsky stress index, HR, gazing away from the stimuli, COVAREP features associated with a "tense" voice, Emotient's "negative emotions" composite, as well as SCR levels were all summed up within an undesirable effects group. For each of the aforementioned features, scores extracted from the exposure phase and those from the response phase of the BAT were summed up (when available). Score treatments (subtracted baseline, subtracted previous theme) described earlier were also summed up, when available. Finally, a single composite effects index was created by subtracting the total score of the undesirable effects group from that of the desirable effects group. This index therefore is weighed in such a way that a higher score means more desirable effects and vice versa. Figure 2 illustrates this analysis.

Unfortunately, we could not include LIWC (linguistic features) in the analysis because they cannot easily be distributed among simple desirable or undesirable effects groups (eg, features such as "frequency of future tense verbs"). Finally, a few mathematically redundant (ie, equal information) features were omitted from this analysis, namely, the LF component of HRV in normalized units, as well as the HF/LF ratio of HRV (because their information is mathematically redundant with respect to the HF component in normalized units, see [79]); the percentage (%) and absolute power versions of the HF component feature (because the normalized units version of the feature controls for the very low frequency (VLF) component of HRV and thus is a more realistic measure of the same construct); the mean and median R-R features of HRV (because they are redundant with respect to HR). Specific EMOTIENT emotion features (eg, sadness) were not included separately since they are all included in two composites already produced by the software, one for negative expressions and the other for positive expressions. The Phasic maximal activity feature of EDA was not included for being redundant with respect to EDA's SCR. The tonic skin conductance feature was not included because it requires longer measuring durations to be meaningful (they were calculated for future analyses focusing on the totality of the BAT instead of just isolated themes). The total amount of features per theme that ended being added up in the composite effects index is of 61 (see Figure 2).

A third analysis, conducted using IBM SPSS Statistics version 23, consisted in performing repeated measures analysis of variances (ANOVAs) on the BAT responses extracted features that were revealed as both statistically significant in their correlation to attachment security as well as theory consistent with attachment theory. The objective was to evaluate if those features yielded different results across different BAT stimuli sets or if they were consistently similar.

Results

Correlation Exploratory Analyses

Adult Attachment Questionnaire (AAQ; Pearson Correlations)

In the responses to BAT's theme 7 ("attuned couple"), AAQ romantic attachment security was negatively correlated with negative facial expressions in general during exposure (r=−.32, P=.02) and anger in particular during response (r=−.43, P=.001) and exposure (r=−.38, P=.006). AAQ attachment security was also negatively correlated with the inhibition (r=−.38, P=.008), tentative (r=−.34, P=.02), and feeling (r=−.41, P=.004) categories of LIWC.

Adult Attachment Projective Picture System (AAP; Pearson Correlations)

In the responses to BAT's theme 4 ("attuned father-child"), theme 7 ("attuned couple"), and theme 13 ("attuned mother-child"), using the mean of the responses to the three themes as a composite score, AAP attachment security was correlated with the NN50 after subtracting baseline (r=0.48, P=.007) and pNN50 after subtracting baseline (r=.38, P=.04), features of HRV, while it was negatively correlated with Bayevsky's stress index after subtracting baseline (r=−.45, P=.01). AAP security was also correlated with the H1-H2 ratio of COVAREP after subtracting baseline (r=.30, P=.02). It was correlated as well with the hearing (r=.31, P=.02), we (r=.45, P<.001), leisure (r=.28, P=.04), and they (r=.40, P=.002) categories of LIWC. It was also correlated with positive facial expressions in general during exposure (r=.32, P=.01) and response (r=.36, P=.005) after subtracting previous theme, and joy in particular during response (r=.38, P=.003) after

XSL•FO
RenderX

subtracting previous theme, as well as disgust in both exposure (r=.31, *P*=.02) and response (r=.33, *P*=.01) after subtracting previous theme. It was also negatively correlated with surprise on both exposure (r=−.29, *P*=.03) and response (r=−.32, *P*=.01), confusion during both exposure (*r*=−.27, *P*=.04) and response (*r*=−.27, *P*=.04), confusion during response (*r*=−.28, *P*=.03) after subtracting previous theme, anger during both exposure (*r*=−.37, *P*=.004) and response (*r*=−.31, *P*=.02) after subtracting previous theme, sadness during response (*r*=−.26, *P*=.04) after subtracting previous theme, neutral expressions during response (*r*=−.29, *P*=.02) after subtracting previous theme, fear during exposure after subtracting baseline (*r*=−.26, *P*=.05), and contempt during response (*r*=−.26, *P*=.05). It was negatively correlated with head size during exposure (*r*=−.32, *P*=.01) after subtracting previous theme and with head size during both exposure (*r*=−.37, *P*=.003) and response (*r*=−.35, *P*=.01) after subtracting baseline.

Attachment Multiple Model Interview (AMMI; Spearman Rank Correlations)

In the responses to BAT's theme 4 ("attuned father-child"), AMMI father attachment security was correlated with the past tense (*r*=.47, *P*=.02), and negatively correlated with the discrepancy (*r*=−.46, *P*=.03) categories of LIWC. It was also negatively correlated with the LF feature of HRV after subtracting baseline (*r*=−.87, *P*=.03), and with SDNN feature of HRV after subtracting baseline (*r*=−.87, *P*=.03). It was also negatively correlated with head size in both exposure (*r*=−.43, *P*=.04) and response (*r*=−.45, *P*=.03). Finally, it was negatively correlated with facial expression of frustration during response (*r*=−.67, *P*<.001) and frustration after subtracting baseline during exposure (*r*=−.44, *P*=.04), as well as expressions of frustration (r=−.53, *P*=.01) and surprise (r=−.43, *P*=.04) after subtracting previous theme.

In the responses to BAT's theme 13 ("attuned mother-child"), AMMI mother attachment security was correlated with SCR (*r*=.46, *P*=.02). It was correlated with facial expressions of sadness during exposure after subtracting baseline (*r*=.42, *P*=.03). It was negatively correlated with facial expression of sadness (r=−.51, *P*=.01) and with gazing away from the stimuli during exposure (r=−.41, *P*=.03) after subtracting previous theme.

In the responses to BAT's theme 7 ("attuned couple"), AMMI partner attachment security was correlated with the high frequency feature of HRV in both normalized units (r=1, *P*=.02) and percentage (r=1, *P*=.02) after subtracting previous theme, and it was negatively correlated with the low frequency feature of HRV in both normalized units (r=−1, *P*=.02) and percentage (r=−1, *P*=.02) after subtracting previous theme, negatively correlated with the ratio of low versus high frequency of HRV (r=−1, *P*=.02) after subtracting previous theme, as well as negatively correlated with the mean heart rate (r=−1, *P*=.02). It was correlated with SCR (r=.61, *P*=.01). It was also correlated with face closeness to screen after subtracting baseline (*r*=.56, *P*=.02). It was negatively correlated with COVAREP Rd feature after subtracting baseline (*r*=−.60, *P*=.01). It was correlated with facial expressions of surprise after subtracting baseline (*r*=.59, *P*=.01). Finally, it was correlated with LIWC *exclusion* category (*r*=.49, *P*=.05).

Confirmatory Analysis

Our composite effects index was significantly correlated to attachment security in the Adult Attachment Projective Picture System (*r*=.26, *P*=.05) by using the mean score from features of themes 4, 7, and 13 (like in the previous analysis), and significantly correlated to attachment security in the AAQ (*r*=.30, *P*=.04) by using scores from features of theme 7. Security with father, mother, and partner in the AMMI were unrelated to the composite effects index by using scores from features of themes 4, 13, and 7, respectively.

Analysis of Variances (ANOVAs)

Repeated measures ANOVAs were conducted on each of the 46 features that were both statistically significant in their correlation to attachment security as well as theory consistent across the responses to the 3 BAT stimuli sets during the aforementioned specific themes. They revealed that only 7 (15%) of the 46 features had significantly different values depending on the stimuli set. Those features were the tentative category of LIWC during theme 7, $F_{1.56,49.92}$=4.81, *P*=.02 (after a Greenhouse-Geisser correction); gazing away from the stimuli (after subtracting previous theme) during exposure to theme 13, $F_{2,76}$=5.75, *P*=.005; and during themes 4, 7, and 13 (mean of the three), the hearing category of LIWC, $F_{2,64}$=4.37, *P*=.02; the leisure category of LIWC, $F_{2,64}$=4.63, *P*=.01; confusion facial expressions during exposure, $F_{1.7,66}$=5, *P*=.01 (after a Greenhouse-Geisser correction) and during response, $F_{1.7,63}$=6.3, *P*=.005 (after a Greenhouse-Geisser correction), as well as anger facial expressions during response after subtracting previous theme, $F_{1.7,62}$=3.7, *P*=.04 (after a Greenhouse-Geisser correction).

Discussion

Principal Findings

Earlier in this work, we discussed the relevance that attachment theory has earned in mental health research, and we commented on the current limitations of psychometric instruments for assessing adult attachment.

We presented the BAT, a new adult-attachment assessment instrument, explicating its sources as well as its rationales and assumptions.

We then set to empirically evaluate two of the BAT's core assumptions: that its themes can help measure attachment security as assessed by validated measures such as the AAQ, the AAP, and the AMMI; and that rotating the stimuli sets in the BAT would not alter the participants' responses to the test.

Regarding the first hypothesis H1, during our exploratory analysis we were able to find physiological, behavioral, and linguistic markers of attachment security, both in general, and specific to romantic partners, mother, and father. These markers were elicited by the BAT's specifically designed attachment-deactivating themes, which counts as preliminary evidence for the instrument's internal and construct validity.

XSL•FO
RenderX

There was an important level of coherence, as well as theory consistency within our findings: in the presence of attachment-deactivating, reassuring stimuli, the more securely attached individuals experienced parasympathetic activation and sympathetic deactivation, a relaxation response revealed by increase in the HF and decrease in the LF of HRV [67], which also produced a decrease in overall stress as shown by the Bayevsky stress index [54]. During the verbal responses, the more securely attached participants' voices became *breathier*, as revealed by COVAREP, indicating relaxation [71], and used more words that can convey attunement, like words related to hearing (eg, listen, heard) and the we pronoun, and conversely less words that can convey relational distress, like words related to inhibition (eg, block, constrain) and to tentativeness (eg, maybe, guess). Positive facial expressions, including joy, were related to attachment security, and their negative counterparts were mostly negatively correlated, as the theory would suggest. The more secure participants tended to not gaze away from the reassuring stimuli but, instead, got physically closer to them. Findings were not all theory consistent, however, as we'll see below.

Since our exploratory analysis was based on multiple hypotheses testing, a statistical concern arose: could these findings be just the product of chance? But when all the available features, including the many that were not *significantly* correlated with attachment in the exploratory analyses, were summed up in a single composite effects index, said index was significantly correlated with two of our three attachment security "ground truth" measures, attesting to the robustness of the findings. This analysis might also suggest that some of the features not showing a statistically significant association with attachment security might not achieve so because of a small sample size or small effect sizes. Composites help increase the effect size of features weighting in the same direction, statistically revealing their direction [74].

We believe this sort of multimodal automatic appraisal of "the whole picture" that is an attachment-deactivating reaction, from a behavioral, psychophysiological, and linguistic standpoint, is a taste of what is becoming possible for psychometrics. Moreover, the fact that our data was obtained outside of a lab setting, using a consumer tablet and its webcam, a consumer-grade USB microphone, and a wireless wristband, attests to the pace at which sensing technology is advancing, offering a glimpse at how effortlessly these measures could be obtained in a close future.

Regarding the second hypothesis H2, 84% (39/46) of the features revealed as associated with attachment security and that were theory consistent were stable across three different BAT stimuli sets. This is especially remarkable given that one of those stimuli sets (always presented second in order) was randomized every time the test was administrated, which means that no person, among our sample, saw or listened to the exact same second stimuli set. This finding suggests that during the BAT, participants react and respond mainly to the themes (ie, the attachment narratives) which are being evoked by the stimuli, and not so much to the stimulus details themselves (eg, the color of a person's hair or the specific background). This also suggests that, as long as stimuli are selected using the standardized procedure described in this work and our minimum fitness scores are respected, new stimuli sets could be developed for the BAT without affecting its capacity to evoke and measure attachment. H2 results suggest, of course, that we should stop including the 7 features that did change across stimuli sets in further developments of the BAT's scoring algorithms, as they seem to be less reliable when stimuli sets are varied.

We have chosen not to perform multiple comparison corrections (eg, Bonferroni) in our repeated measures ANOVAs to, counterintuitively, increase the rigor of the analysis. This is because within our ANOVA analysis, we compared 46 features across 3 different stimuli sets, for a total of 138 F tests. A Bonferroni correction would imply that the alpha level is divided by the number of comparisons (.05/138) for a corrected alpha level of .0003. It would be very difficult for any difference to be found under this alpha level with our sample size. This would be a convenient result, but probably a false one.

Limitations

Out of the 65 significant correlations revealed by our exploratory analysis, 19 (29%) seemed to go against what would be expected from an attachment theory perspective. Some of the most striking examples were the increase in sadness expressions correlated with AMMI mother model security, or the increase of disgust expressions correlated with AAP security. Thus when developing our scoring and classification algorithms in the future, it will be important to discard such features, unless we can find theoretical underpinnings for them.

The confirmatory analysis we performed, based on creating a composite effects index, was designed to prove that *overall* the many features studied when put together weighted in the right direction in correlation with attachment security. We argue that this is confirmatory evidence for the construct validity of the BAT, namely that the test activates people in a way that can be captured by multiple modalities and that is correlated with attachment security; but we do not present this as confirmatory evidence for the relationship of any specific feature and attachment security. For example, our study suggests, but does not confirm, that a breathier voice can be a watermark of a more securely attached person during an attachment-soothing situation. In this sense, specially associations of LIWC features and attachment security should be seen as merely exploratory since LIWC was not included in our confirmatory analysis (for a rigorous confirmatory study about LIWC features and attachment, see [80]). As for security in the AMMI, which was not significantly correlated with our composite effects index (albeit it was with several individual features), it is important to restate that only a fraction of our sample took that test (n=27 for the mother model, n=23 for the father model, n=17 for the partner model) and since the effect sizes of most features are small, this might explain the lack of association.

In our ANOVA analysis, we decided to include only features that were found to be both statistically significant in their associations to attachment security as well as theory consistent in that association. Why? We compared incredibly different variables in these ANOVAs; from word count on a variety of categories to facial expressions to HRV, and so on, as a reaction to very different image and music combinations. All these

XSL•FO
RenderX

features were calculated without human supervision out of the raw input data, introducing some degree of random error that should favor a finding of difference between answers to different stimuli sets. As a result, odds were stacked in favor of finding differences, and avoiding a multiple comparison correction made it more so as explained above. Including more features (eg, theory-inconsistent features) without controlling for multiple comparisons would generate just too many type I errors for the analysis to be useful.

Future Directions

The BAT was designed to test far more than the attachment security dimension. The different themes in the BAT were designed to also test for the other three main attachment dimensions: attachment anxiety, attachment avoidance [40], and attachment disorganization [12]. They were also designed to measure attachment defenses, such as deactivation, cognitive disconnection, and segregated systems [13]. Different themes heighten differences in the reactions of the four classic adult attachment groups (dismissing, preoccupied, secure, unresolved) to help in classification of attachment. Finally, some themes in the BAT were designed to measure emotional regulation, as measured by instruments such as the Difficulties in Emotion Regulation Scale (DERS [81]), and relational trauma, as measured by the trauma system of the AAP [13]. Empirical validation of the BAT's fitness to measure these constructs is, of course, warranted.

An important area of our work with the BAT is the development of algorithms to automatically score and classify attachment based on extracted features from responses to the test, like the ones highlighted by this study.

This endeavor is complex. It entails finding the right fusion formula for the different BAT features so that the emerging multimodal pattern can accurately predict attachment continuous scores and classifications. It also entails extensive cross-validation to verify the generalizability of the prediction capability to new cohorts. We are underway in this work, and in fact we have developed preliminary regression and classification algorithms capable of predicting ground truth attachment continuous scores and classifications better than chance, cross-validating our results to prevent over-fitting and to warrant generalizability.

Conclusions

Overall, this study brings us one step closer to our goal of developing an automatic and objective adult attachment test. In the future, a 9-min BAT test could be deployed through the Internet to participants or patients residing in remote areas. The test could be scored instantaneously and automatically, with the results becoming available to the researcher or clinician just minutes later. We hope that this could unleash a new wave of attachment research as well as favor clinical attachment testing, in turn benefiting patients by offering them more cost-effective and efficient mental health assessments and treatments.

Acknowledgments

We wish to thank Ville Evrard Clinical Research Unit's Director, Dr Dominique Januel, for her support in the form of financial help to purchase material for this study, as well as granting us access to patients in the Ville Evrard Psychotherapy Center. The work depicted here is sponsored by the US Army Research Laboratory (ARL) under contract number W911NF-14-D-0005. Statements and opinions expressed and content included do not necessarily reflect the position or the policy of the US Government and no official endorsement should be inferred.

Conflicts of Interest

None declared.

References

1. Bowlby J. Attachment and loss. New York: Basic Books; 1969.
2. Cassidy J, Shaver P. editors. Handbook of attachment: Theory, research, and clinical applications. 3rd edition. New York: Guilford Press; 2016.
3. Sroufe LA, Waters E. Attachment as an Organizational Construct. Child Development 1977 Dec;48(4):1184. [doi: 10.2307/1128475]
4. Mesman J, van IJzendoorn MH, Sagi-Schwartz A. Cross-cultural patterns of attachment: universal and contextual dimensions. In: Cassidy J, Shaver PR, editors. Handbook of Attachment: Theory, Research, and Clinical Applications. New York: Guilford Press; 2016.
5. Grossmann K, Grossmann K, Keppler A. Universal and Culture-Specific Aspects of Human Behavior: The Case of Attachment. In: Friedlmeier W, Chakkarath P, Schwarz B, editors. Culture and Human Development: The Importance of Cross-Cultural Research to the Social Sciences. Hove: Psychology Press; 2005.
6. Ainsworth MD. Attachments across the life span. Bull N Y Acad Med 1985 Nov;61(9):792-812 [FREE Full text] [Medline: 3864511]
7. Sroufe LA, Coffino B, Carlson EA. Conceptualizing the role of early experience: lessons from the minnesota longitudinal study. Dev Rev 2010 Mar 1;30(1):36-51 [FREE Full text] [doi: 10.1016/j.dr.2009.12.002] [Medline: 20419077]
8. Fraley C, Roisman G. Do early caregiving experiences leave an enduring or transient mark on developmental adaptation? Curr Opin Psychol 2015 Feb;1:101-106. [doi: 10.1016/j.copsyc.2014.11.007]

XSL•FO
RenderX

9. De Wolff MS, van IJzendoorn MH. Sensitivity and attachment: a meta-analysis on parental antecedents of infant attachment. Child Development 1997 Aug;68(4):571-591. [doi: 10.1111/j.1467-8624.1997.tb04218.x]
10. Lorber MF, Egeland B. Infancy parenting and externalizing psychopathology from childhood through adulthood: developmental trends. Dev Psychol 2009 Jul;45(4):909-912 [FREE Full text] [doi: 10.1037/a0015675] [Medline: 19586169]
11. Miljkovitch R, Moran G, Roy C, Jaunin L, Forcada-Guex M, Pierrehumbert B, et al. Maternal interactive behaviour as a predictor of preschoolers' attachment representations among full term and premature samples. Early Hum Dev 2013 May;89(5):349-354. [doi: 10.1016/j.earlhumdev.2012.11.006] [Medline: 23265254]
12. Miljkovitch R, Moss E, Bernier A, Pascuzzo K, Sander E. Refining the assessment of internal working models: the Attachment Multiple Model Interview. Attach Hum Dev 2015;17(5):492-521. [doi: 10.1080/14616734.2015.1075561] [Medline: 26325611]
13. George C, West M. The adult attachment projective picture system: attachment theory and assessment in adults. New York: Guilford Press; 2012.
14. Mikulincer M, Dolev T, Shaver PR. Attachment-related strategies during thought suppression: ironic rebounds and vulnerable self-representations. J Pers Soc Psychol 2004 Dec;87(6):940-956. [doi: 10.1037/0022-3514.87.6.940] [Medline: 15598116]
15. Coan J. Towards a neuroscience of attachment. In: Cassidy J, Shaver PR, editors. Handbook of Attachment: Theory, Research, and Clinical Applications. New York: Guilford Press; 2016.
16. Hane A, Fox N. Studying the biology of human attachment. In: Cassidy J, Shaver PR, editors. Handbook of Attachment: Theory, Research, and Cinical Applications. New York: Guilford Press; 2016.
17. Lemche E, Giampietro VP, Surguladze SA, Amaro EJ, Andrew CM, Williams SC, et al. Human attachment security is mediated by the amygdala: evidence from combined fMRI and psychophysiological measures. Hum Brain Mapp 2006 Aug;27(8):623-635. [doi: 10.1002/hbm.20206] [Medline: 16284946]
18. Fraley RC, Spieker SJ. Are infant attachment patterns continuously or categorically distributed? a taxometric analysis of strange situation behavior. Dev Psychol 2003;39(3):387-404. [doi: 10.1037/0012-1649.39.3.387]
19. Grossmann KE, Grossmann K, Waters E. Attachment from Infancy to Adulthood: The Major Longitudinal Studies. New York: The Guilford Press; 2005.
20. Carlson EA. A prospective longitudinal study of attachment disorganization/disorientation. Child Development 1998;69(4):1107-1128. [doi: 10.1111/j.1467-8624.1998.tb06163.x]
21. Mikulincer M, Shaver PR. An attachment perspective on psychopathology. World Psychiatry 2012 Feb;11(1):11-15. [doi: 10.1016/j.wpsyc.2012.01.003]
22. Bakermans-Kranenburg MJ, van IJzendoorn MH. The first 10,000 adult attachment interviews: distributions of adult attachment representations in clinical and non-clinical groups. Attach Hum Dev 2009 May;11(3):223-263. [doi: 10.1080/14616730902814762] [Medline: 19455453]
23. Cantazaro A, Wei M. Adult attachment, dependence, self-criticism, and depressive symptoms: a test of a mediational model. J Pers 2010 Aug 01;78(4):1135-1162. [doi: 10.1111/j.1467-6494.2010.00645.x] [Medline: 20545820]
24. Ein-Dor T, Doron G, Solomon Z, Mikulincer M, Shaver PR. Together in pain: attachment-related dyadic processes and posttraumatic stress disorder. J Couns Psychol 2010;57(3):317-327. [doi: 10.1037/a0019500]
25. Levy KN, Meehan KB, Weber M, Reynoso J, Clarkin JF. Attachment and borderline personality disorder: implications for psychotherapy. Psychopathology 2005;38(2):64-74. [doi: 10.1159/000084813] [Medline: 15802944]
26. Brown DP, Elliott DS. Attachment disturbances in adults: Treatment for comprehensive repair. New York: W. W. Norton & Company; 2016.
27. Mikulincer M, Shaver PR. The psychological effects of the contextual activation of security-enhancing mental representations in adulthood. Curr Opin Psychol 2015 Feb;1:18-21. [doi: 10.1016/j.copsyc.2015.01.008]
28. Parnell L, Siegel DJ. Attachment-focused EMDR: Healing relational trauma. New York: W. W. Norton & Company; 2015.
29. Mikulincer M, Shaver P. Adult attachment and emotion regulation. In: Cassidy J, Shaver PR, editors. Handbook of Attachment: Theory, Research, and Clinical Applications. New York: Guilford Press; 2016.
30. Bergman K, Sarkar P, Glover V, O'Connor TG. Maternal prenatal cortisol and infant cognitive development: moderation by infant-mother attachment. Biol Psychiatry 2010 Jun 1;67(11):1026-1032 [FREE Full text] [doi: 10.1016/j.biopsych.2010.01.002] [Medline: 20188350]
31. Miljkovitch R. Les fondations du lien amoureux. Paris: Presses universitaires de France; 2009.
32. Feeney J. Adult romantic attachment: developments in the study of couple relationships. In: Cassidy J, Shaver PR, editors. Handbook of Attachment: Theory, Research, and Clinical Applications. New York: Guilford Press; 2016.
33. Rodin G, Walsh A, Zimmermann C, Gagliese L, Jones J, Shepherd FA, et al. The contribution of attachment security and social support to depressive symptoms in patients with metastatic cancer. Psychooncology 2007 Dec;16(12):1080-1091. [doi: 10.1002/pon.1186] [Medline: 17464942]
34. Ditzen B, Schmidt S, Strauss B, Nater UM, Ehlert U, Heinrichs M. Adult attachment and social support interact to reduce psychological but not cortisol responses to stress. J Psychosom Res 2008 May;64(5):479-486. [doi: 10.1016/j.jpsychores.2007.11.011] [Medline: 18440400]
35. Bartholomew K, Horowitz LM. Attachment styles among young adults: a test of a four-category model. J Pers Soc Psychol 1991;61(2):226-244. [doi: 10.1037/0022-3514.61.2.226]

XSL•FO
RenderX

36. Larose S, Bernier A. Social support processes: mediators of attachment state of mind and adjustment in late adolescence. Attach Hum Dev 2001 Apr;3(1):96-120. [doi: 10.1080/14616730010024762] [Medline: 11708386]
37. Powers SI, Pietromonaco PR, Gunlicks M, Sayer A. Dating couples' attachment styles and patterns of cortisol reactivity and recovery in response to a relationship conflict. J Pers Soc Psychol 2006 Apr;90(4):613-628. [doi: 10.1037/0022-3514.90.4.613] [Medline: 16649858]
38. Wallace JL, Vaux A. Social support network orientation: the role of adult attachment style. J Soc Clin Psychol 1993 Sep;12(3):354-365. [doi: 10.1521/jscp.1993.12.3.354]
39. Ravitz P, Maunder R, Hunter J, Sthankiya B, Lancee W. Adult attachment measures: a 25-year review. J Psychosom Res 2010 Oct;69(4):419-432. [doi: 10.1016/j.jpsychores.2009.08.006] [Medline: 20846544]
40. Simpson JA, Rholes WS, Phillips D. Conflict in close relationships: An attachment perspective. J Pers Soc Psychol 1996;71(5):899-914. [doi: 10.1037/0022-3514.71.5.899]
41. Collins NL, Read SJ. Adult attachment, working models, and relationship quality in dating couples. J Pers Soc Psychol 1990;58(4):644-663. [doi: 10.1037/0022-3514.58.4.644]
42. Podsakoff PM, MacKenzie SB, Lee J, Podsakoff NP. Common method biases in behavioral research: a critical review of the literature and recommended remedies. J Appl Psychol 2003 Oct;88(5):879-903. [doi: 10.1037/0021-9010.88.5.879] [Medline: 14516251]
43. Ainsworth MDS, Bell SM. Attachment, exploration, and separation: illustrated by the behavior of one-year-olds in a strange situation. Child Development 1970 Mar;41(1):49. [doi: 10.2307/1127388]
44. Cassidy J, Sherman LJ, Jones JD. What's in a word? linguistic characteristics of adult attachment interviews. Attach Hum Dev 2012 Jan;14(1):11-32. [doi: 10.1080/14616734.2012.636649] [Medline: 22191604]
45. Maxwell SE, Lau MY, Howard GS. Is psychology suffering from a replication crisis? what does “failure to replicate” really mean? Am Psychol 2015 Sep;70(6):487-498. [doi: 10.1037/a0039400] [Medline: 26348332]
46. Cummins N, Scherer S, Krajewski J, Schnieder S, Epps J, Quatieri TF. A review of depression and suicide risk assessment using speech analysis. Speech Commun 2015 Jul;71:10-49. [doi: 10.1016/j.specom.2015.03.004]
47. Scherer S, Gale L, Gratch J, Rizzo A, Morency LP. Self-reported symptoms of depression and PTSD are associated with reduced vowel space in screening interviews. IEEE Trans Affective Comput 2016 Jan 1;7(1):59-73. [doi: 10.1109/TAFFC.2015.2440264]
48. NIMH. 2013. Director’s blog: transforming diagnosis URL:https://www.nimh.nih.gov/about/directors/thomas-insel/blog/2013/transforming-diagnosis.shtml [accessed 2017-03-14] [WebCite Cache ID 6oxwh0qdc]
49. Scherer S, Stratou G, Lucas G, Mahmoud M, Boberg J, Gratch J, et al. Automatic audiovisual behavior descriptors for psychological disorder analysis. Image Vis Comput 2014 Oct;32(10):648-658. [doi: 10.1016/j.imavis.2014.06.001]
50. Novak D, Mihelj M, Munih M. A survey of methods for data fusion and system adaptation using autonomic nervous system responses in physiological computing. Interact Comput 2012 May;24(3):154-172. [doi: 10.1016/j.intcom.2012.04.003]
51. Lucas G, Gratch J, Scherer S, Boberg J, Stratou G. Towards an affective interface for assessment of psychological distress. 2015 Presented at: International Conference on Affective Computing and Intelligent Interaction (ACII); 2015; Washington, DC p. 539-545.
52. Klagsbrun M, Bowlby J. Responses to separation from parents: a clinical test for young children. British Journal of Projective Psychology & Personality Study 1976;21(2):7-27.
53. Baumgartner T, Lutz K, Schmidt CF, Jäncke L. The emotional power of music: how music enhances the feeling of affective pictures. Brain Res 2006 Feb 23;1075(1):151-164. [doi: 10.1016/j.brainres.2005.12.065] [Medline: 16458860]
54. Bayevsky RM. Analysis of heart rate variability in space medicine. Hum Physiol 2002;28(2):202-213. [doi: 10.1023/A:1014866501535]
55. Tuber S. Understanding personality through projective testing. Lanham: Rowman & Littlefield Publishers; Mar 2014.
56. Schultz DS, Loving JL. Challenges since Wikipedia: the availability of Rorschach information online and internet users' reactions to online media coverage of the Rorschach-Wikipedia debate. J Pers Assess 2012;94(1):73-81. [doi: 10.1080/00223891.2011.627963] [Medline: 22176268]
57. Marchewka A, Zurawski L, Jednoróg K, Grabowska A. The Nencki Affective Picture System (NAPS): introduction to a novel, standardized, wide-range, high-quality, realistic picture database. Behav Res Methods 2014 Jun;46(2):596-610 [FREE Full text] [doi: 10.3758/s13428-013-0379-1] [Medline: 23996831]
58. Lang P, Bradley M, Cuthbert B. International affective picture system (IAPS): Technical manual and affective ratings. Gainesville: University of Florida, Center for Research in Psychophysiology; 1999.
59. Dan-Glauser ES, Scherer KR. The Geneva affective picture database (GAPED): a new 730-picture database focusing on valence and normative significance. Behav Res Methods 2011 Jun;43(2):468-477. [doi: 10.3758/s13428-011-0064-1] [Medline: 21431997]
60. Zapin M. SurveyGizmo. Why and how to clean your online survey data URL:https://www.surveygizmo.com/survey-blog/clean-your-online-survey-data/ [accessed 2016-10-27] [WebCite Cache ID 6lYu19NYf]
61. Eerola T, Vuoskoski JK. A comparison of the discrete and dimensional models of emotion in music. Psychol Music 2010 Aug 27;39(1):18-49. [doi: 10.1177/0305735610362821]

XSL•FO
RenderX

62. Mathôt S, Schreij D, Theeuwes J. OpenSesame: an open-source, graphical experiment builder for the social sciences. Behav Res Methods 2012 Jun;44(2):314-324 [FREE Full text] [doi: 10.3758/s13428-011-0168-7] [Medline: 22083660]
63. McCarthy C, Pradhan N, Redpath C, Adler A. Validation of the Empatica E4 wristband. 2016 Presented at: InIEEE EMBS International Student Conference (ISC); 2016; Ottawa.
64. Empatica. 2015. Comparison Procomp Vs Empatica E3 Skin Conductance Signal URL:https://dl.dropboxusercontent.com/u/1421169/ProcompVsE3_final20150309.pdf [accessed 2017-03-14] [WebCite Cache ID 6oxxj7DaT]
65. Garbarino M, Lai M, Bender D. Empatica E3 A wearable wireless multi-sensor device for real-time computerized biofeedback and data acquisition. Gent, Belgium: ICST; 2014 Nov Presented at: 4th International Conference on Wireless Mobile Communication and Healthcare - Transforming healthcare through innovations in mobile and wireless technologies (MOBIHEALTH); 2014; Athens, Greece p. 03.
66. Kaufmann T, Sütterlin S, Schulz SM, Vögele C. ARTiiFACT: a tool for heart rate artifact processing and heart rate variability analysis. Behav Res Methods 2011 Dec;43(4):1161-1170. [doi: 10.3758/s13428-011-0107-7] [Medline: 21573720]
67. Task Force of the European Society of Cardiology, The North American Society of Pacing and Electrophysiology. Heart rate variability: standards of measurement, physiological interpretation and clinical use. Circulation 1996;93(5):1043-1065.
68. Benedek M, Kaernbach C. A continuous measure of phasic electrodermal activity. J Neurosci Methods 2010 Jun 30;190(1):80-91 [FREE Full text] [doi: 10.1016/j.jneumeth.2010.04.028] [Medline: 20451556]
69. Littlewort G, Whitehill J, Wu T, Fasel I, Frank M, Movellan J, et al. The computer expression recognition toolbox (CERT). Gesture Recognition 2011:298-305.
70. Amos B, Ludwiczuk B, Satyanarayanan M. ADM. 2016. Openface: A general-purpose face recognition library with mobile applications URL:http://reports-archive.adm.cs.cmu.edu/anon/anon/2016/CMU-CS-16-118.pdf [accessed 2017-03-14] [WebCite Cache ID 6oxyEWbGa]
71. Degottex G, Kane J, Drugman T, Raitio T, Scherer S. COVAREP: A collaborative voice analysis repository for speech technologies. 2014 Presented at: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); 2014; Florence.
72. Piolat A, Booth R, Chung C, Davids M, Pennebaker J. La version française du dictionnaire pour le LIWC : modalités de construction et exemples d'utilisation. Psychologie Française 2011 Sep;56(3):145-159. [doi: 10.1016/j.psfr.2011.07.002]
73. Bender R, Lange S. Adjusting for multiple testing-when and how? J Clin Epidemiol 2001 Apr;54(4):343-349. [doi: 10.1016/S0895-4356(00)00314-0]
74. Anderson ML. Multiple inference and gender differences in the effects of early intervention: a reevaluation of the abecedarian, perry preschool, and early training projects. J Am Stat Assoc 2008 Dec;103(484):1481-1495. [doi: 10.1198/016214508000000841]
75. Kling JR, Liebman JB, Katz LF, Sanbonmatsu L. Princeton. 2004. Moving to opportunity and tranquility: neighborhood effects on adult economic self-sufficiency and health from a randomized housing voucher experiment URL:http://dataspace.princeton.edu/jspui/handle/88435/dsp01qz20ss50t [accessed 2017-03-14] [WebCite Cache ID 6oxyPqYa3]
76. Rothman KJ. Six persistent research misconceptions. J Gen Intern Med 2014 Jul;29(7):1060-1064 [FREE Full text] [doi: 10.1007/s11606-013-2755-z] [Medline: 24452418]
77. Pascovici D, Handler DC, Wu JX, Haynes PA. Multiple testing corrections in quantitative proteomics: a useful but blunt tool. Proteomics 2016 Sep;16(18):2448-2453. [doi: 10.1002/pmic.201600044] [Medline: 27461997]
78. Feise RJ. Do multiple outcome measures require p-value adjustment? BMC Med Res Methodol 2002 Jun 17;2(1):1450. [doi: 10.1186/1471-2288-2-8]
79. Burr RL. Interpretation of normalized spectral heart rate variability indices in sleep research: a critical review. Sleep 2007;30(7):913-919.
80. Waters TE, Steele RD, Roisman GI, Haydon KC, Booth-LaForce C. A linguistic inquiry and word count analysis of the adult attachment interview in two large corpora. Can J Behav Sci 2016 Jan 01;48(1):78-88 [FREE Full text] [doi: 10.1037/cbs0000035] [Medline: 27065477]
81. Gratz KL, Roemer L. Multidimensional assessment of emotion regulation and dysregulation: development, factor structure, and initial validation of the difficulties in emotion regulation scale. J Psychopathol Behav Assess 2004 Mar;26(1):41-54. [doi: 10.1023/B:JOBA.0000007455.08539.94]

Abbreviations

AAI: Adult Attachment Interview
AAP: Adult Attachment Projective Picture System
AAQ: Adult Attachment Questionnaire
AAS: Adult Attachment Scale
AMMI: Attachment Multiple Model Interview
ANOVA: analysis of variance
ARL: Army Research Laboratory
BAT: Biometric Attachment Test

XSL•FO
RenderX

COVAREP: Cooperative Voice Analysis Repository for Speech Technologies
DERS: Difficulties in Emotion Regulation Sale
EDA: electrodermal activity
GAPED: Geneva Affective Picture Database
HF: high frequency
HR: heart rate
HRV: heart rate variability
IAPS: International Affective Picture System
IBI: interbeat interval
LF: low frequency
LIWC: Linguistic Inquiry and Word Count
NAPS: Nencki Affective Picture System
PTSD: posttraumatic stress disorder
SAT: Separation Anxiety Test
SCR: skin conductance response
SPT: Standard Penetration Test
SSP: Strange Situation Procedure
VLF: very low frequency

Edited by R Calvo; submitted 27.10.16; peer-reviewed by A Jaroszewski, D Messinger; comments to author 28.11.16; revised version received 27.01.17; accepted 05.03.17; published 06.04.17

Please cite as:
Parra F, Miljkovitch R, Persiaux G, Morales M, Scherer S
The Multimodal Assessment of Adult Attachment Security: Developing the Biometric Attachment Test
J Med Internet Res 2017;19(4):e100
URL: http://www.jmir.org/2017/4/e100/
doi: 10.2196/jmir.6898
PMID: 28385683

©Federico Parra, Raphaële Miljkovitch, Gwenaelle Persiaux, Michelle Morales, Stefan Scherer. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 06.04.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

The Multimodal Assessment of Adult Attachment Security: Developing the Biometric Attachment Test

Federico Parra[1,2], MA; Raphaële Miljkovitch[2], PhD; Gwenaelle Persiaux[3], MA; Michelle Morales[1,4], MS; Stefan Scherer[1], PhD

USC Institute for Creative Technologies

[1]Institute for Creative Technologies, University of Southern California, Los Angeles, CA, United States
[2]Paragraphe Laboratory, Paris VIII University, Saint-Denis, France
[3]University Hospital Center Sainte-Étienne, Sainte-Étienne, France
[4]The Graduate Center, City University of New York, New York, NY, United States

UNIVERSITÉ PARIS 8 VINCENNES-SAINT-DENIS

INTRODUCTION

Attachment theory has been proven essential for mental health, including psychopathology, development, and interpersonal relationships.

Validated psychometric instruments to measure attachment abound but suffer from shortcomings common to traditional psychometrics: bias in the case of self report measures, and cost in the case of interview-based measures.

Recent developments in computer vision, multimodal fusion and machine learning pave the way for new automated and objective psychometric instruments for adult attachment that combine psychophysiological, linguistic, and behavioral analyses in the assessment of the construct.

THE BIOMETRIC ATTACHMENT TEST (BAT)

- **We introduce a new exposure-based, automatic, and objective adult-attachment assessment, the Biometric Attachment Test (BAT)**
- It exposes participants to a 9-minute set of visual and music stimuli, based on 14 themes (see figure) that depict situations of separation and loss, on the one hand, and of human attachment and connection, on the other
- According to attachment theory, adults with secure attachment would experience some soothing and relief when human connection is evoked during a stressful situation, whereas adults with insecure attachment would not
- The participants' immediate reactions and verbal responses to the stimuli are captured in several modalities: electro dermal activity (EDA), heart rate (HR), video and audio

Theme	Theme name	Objective	Stimuli sample
1	Baseline	-	
2	Child solitude	Depends on attachment pattern	
3	Couple separation	Attachment activating or stressing	
4	Attuned father-child	Attachment deactivating or soothing	
5	Violent father	Attachment activating or stressing	
6	Unresponsive mother	Attachment activating or stressing	
7	Attuned couple	Attachment deactivating or soothing	
8	Exploration	Depends on attachment pattern	
9	Raw sadness	Attachment activating or stressing	Sad music
10	Vulnerable baby	Depends on attachment pattern	
11	Raw fear	Attachment activating or stressing	Tense music
12	Death or loss	Attachment activating or stressing	
13	Attuned mother-child	Attachment deactivating or soothing	
14	Family connection	Attachment deactivating or soothing	

- Features are extracted from the raw data: heart rate variability (HRV) using Artiifact; Bayevsky stress index, skin conductance response using Ledalab, facial expressions and head distance using Facet Emotient, gaze using Open Face, tenseness of the voice using COVAREP, and frequency of linguistic characteristics of speech using Google Cloud Speech and LIWC

THE PRESENT STUDY

We aimed to empirically validate two assumptions of the BAT:

- its capacity to measure attachment security
- the notion that using different stimuli sets evoking the same themes would produce similar responses

As ground truth we administered three validated attachment measures:

- A projective measure: the Adult Attachment Projective Picture System (AAP)
- An interview-based measure: the Attachment Multiple Model Interview (AMMI)
- A self-report measure: the Adult Attachment Questionnaire (AAQ)

Hypotheses

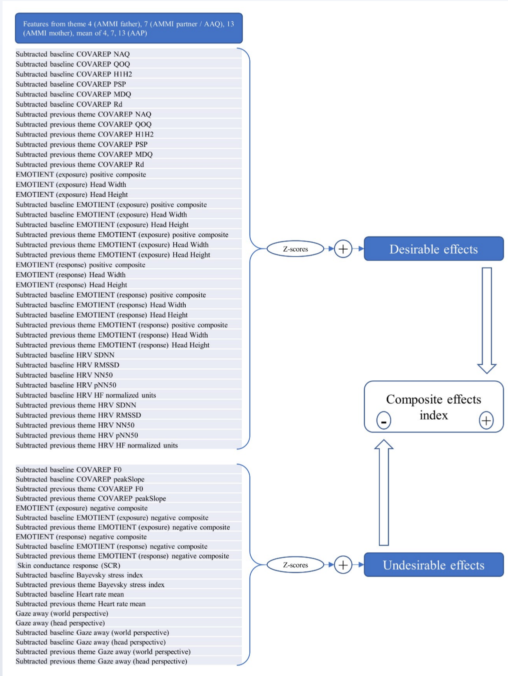
- H1: We'll find a coherent association between the scores of features of BAT responses (e.g. HRV during theme 13) and ground truth attachment security scores (e.g. AAP attachment security). When combined together in a composite effect index, the composite score of all features of BAT responses will be associated with ground truth attachment security as well
- H2: We'll not find major differences between the responses of same subjects to different stimuli sets

Sample:

- The sample consisted of 59 French francophone participants (45 females, 14 males).
- The sample was formed from multiple sources in different regions of France.
- It was intended for the sample to be as diverse as possible in terms of age (mean 35.7, SD 12.2), occupational status (10% unemployed, 32% employed, 33% students, 23% other), as well as relationship status (37% in a relationship, 23% married, 11% separated or divorced, 25% single, 3% unknown) and psychopathology (27% were patients).

Analyses:

- An exploratory analysis was conducted using Pearson and Spearman Rank correlation matrices to uncover the extracted features that are most associated with ground truth attachment security.
- A confirmatory analysis was conducted by creating a single composite effects index (see figure) and by testing it for correlations with ground truth attachment security.
- The stability of the theory-consistent features across three different stimuli sets was explored using repeated measures analysis of variances (ANOVAs).



PRINCIPAL FINDINGS

Regarding the first hypothesis H1: in the presence of attachment-deactivating, reassuring stimuli (themes 4, 7 and 13), the more securely attached individuals experienced parasympathetic activation and sympathetic deactivation, a relaxation response revealed by increase in the high frequency and decrease in the low frequency of HRV, which also produced a decrease in overall stress as shown by the Bayevsky stress index. During the verbal responses, the more securely attached participants' voices became breathier, as revealed by COVAREP, indicating relaxation, and used more words that can convey attunement, like words related to hearing (eg, listen, heard) and the we pronoun, and conversely less words that can convey relational distress, like words related to inhibition (eg, block, constrain) and to tentativeness (eg, maybe, guess). Positive facial expressions, including joy, were related to attachment security, and their negative counterparts were mostly negatively correlated, as the theory would suggest. The more secure participants tended to not gaze away from the reassuring stimuli but, instead, got physically closer to them.

When all the available features were summed up in a single composite effects index, said index was significantly correlated with two of our three attachment security "ground truth" measures, attesting to the robustness of the findings.

Regarding the second hypothesis H2, 84% (39/46) of the features revealed as associated with attachment security and that were theory consistent were stable across three different BAT stimuli sets. This finding suggests that during the BAT, participants react and respond mainly to the attachment narratives themes which are being evoked by the stimuli, and not so much to the stimulus details themselves (eg, the color of a person's hair or the specific background).

CONCLUSIONS

This study brings us one step closer to our goal of developing an automatic and objective adult attachment test. In the future, a 9-min BAT test could be deployed through the Internet to participants or patients residing in remote areas. The test could be scored instantaneously and automatically, with the results becoming available to the researcher or clinician just minutes later. We hope that this could unleash a new wave of attachment research as well as favor clinical attachment testing, in turn benefiting patients by offering them more cost-effective and efficient mental health assessments and treatments.

ACKNOWLEDGEMENTS

We wish to thank the Ville Evrard Clinical Research Unit in France for its support. The work depicted here is sponsored by the US Army Research Laboratory (ARL) under contract number W911NF-14-D-0005. Statements and opinions expressed and content included do not necessarily reflect the position or the policy of the US Government and no official endorsement should be inferred.

CONTACT

Corresponding Author: Federico Parra, MA
Institute for Creative Technologies, University of Southern California
Correspondence: 45 Rue de Rochechouart, 75009, Paris
Phone: +33 782132695 | Email: federico.parra@hotmail.com

RESEARCH POSTER PRESENTATION DESIGN © 2015
www.PosterPresentations.com

5.2 Development and cross-cultural evaluation of a scoring algorithm for the Biometric Attachment Test: Overcoming the challenges of multimodal fusion with “small data”

This paper was published in 2019 on the Transactions on Affective Computing journal (impact factor 6.28), an interdisciplinary journal dedicated to the intersection of psychology, particularly the study of human emotion, and computer science.

Our most technical paper to date, it describes in detail an original combination of machine learning methods we implemented in order for the BAT to be able to learn how to score adult attachment. The paper addressed a common problem affecting the use of machine learning with psychology samples, namely how to apply algorithms designed to deal with massive amounts of data to the small samples common in psychology research.

But the paper also offers original psychological interpretations of these machine learning algorithms and methods, making it a good introduction to the “guts” of the BAT for psychologists without a background in computer science.

Finally, this paper is also the first to present psychometric evidence of the BAT’s convergent and multi-cultural validity, using samples from France and the United States.

(revised May 2019) Development and cross-cultural evaluation of a scoring algorithm for the Biometric Attachment Test: Overcoming the challenges of multimodal fusion with "small data"

Federico Parra, Stefan Scherer, Yannick Benezeth, Plamena Tsvetanova, Susana Tereno

***Abstract*—The Biometric Attachment Test (BAT) is a recently developed psychometric assessment that exposes adults to standardized picture and music stimuli-sets while simultaneously capturing their linguistic, behavioral and physiological responses, with the goal of objectively measuring their psychological attachment characteristics. Within this work, (I) we describe a new version of the BAT (v2) that implements a remote photoplethysmography method to obtain physiological measures from video alone. (II) We discuss the specific challenges we found when trying to develop an automatic scoring algorithm for the BAT v2 using machine learning: practicing multimodal fusion over a high-dimensional feature space with a small learning sample. We propose and evaluate an original combination of methods, including a three-step hybrid multimodal fusion procedure, that overcomes these challenges. (III) Using the proposed methodology, we train a scoring algorithm for the BAT v2 on a francophone sample, using the Adult Attachment Questionnaire as ground-truth. (IV) We then validate the scoring algorithm cross-culturally, testing its performance on an independent anglophone sample, showing low error and high correlation and serving as the BAT v2's first convergent validity evidence. We believe this work constitutes a breakthrough in the development of the first objective and automatic measure for adult attachment, and we hope that our "small data" learning methodology could be useful for other machine learning projects involving small samples coming from psychological research.**

Index terms

multimodal, fusion, hybrid, Bayesian, regularization, global optima, learning objective, machine learning, artificial intelligence, psychometrics, attachment, biometric, small data, global minima, RPPG, remote photophlysmography, xamarin, high-dimensional

Federico Parra is with LPPS (EA 4057) Institut de Psychologie, Université Paris Descartes, Sorbonne Paris Cité, France

Stefan Scherer is with Institute for Creative Technologies, University of Southern California

Yannick Benezeth is with Université Bourgogne Franche-Comté, LE2I EA 7508, Dijon, France

Plamena Tsvetanova is with Sofia University "St. Kliment Ohridski", Sofia, Bulgaria

Susana Tereno is with LPPS (EA 4057) Institut de Psychologie, Université Paris Descartes, Sorbonne Paris Cité, France

1 Introduction

1.1 What is attachment?

Within social psychology, adult attachment refers to a specific kind of intimate bond characterized by four qualities: the attached partners attempt to be physically and emotionally close ("proximity seeking"), they experience separation stress when involuntarily separated, they use the other person as a source of reassurance when in distress ("haven of safety") and they use the other person as a source of confidence and motivation for inner and outer exploration ("secure base") [1]. Attachment also refers to the capacity to form and to maintain such a bond. This capacity has been modeled by Bartholomew and Horowitz as a two-dimensional space (see Fig. 1 below) consisting of attachment anxiety on one axis and attachment avoidance on the other [2] (see Fig. 1).

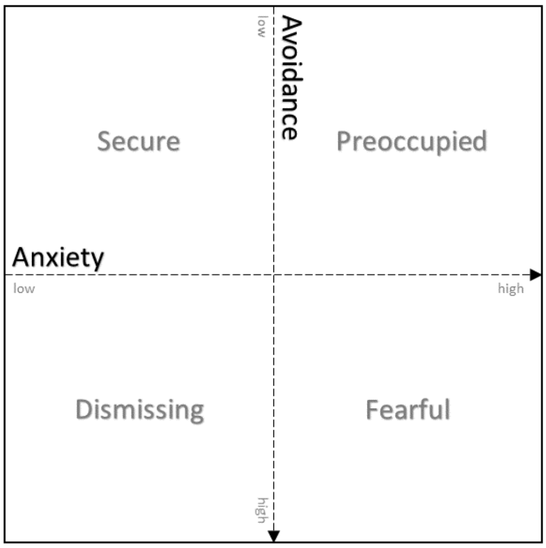


Figure 1. Bartholomew and Horowitz's two-dimensional adult attachment model.

Attachment anxiety refers to the degree to which an individual gets over-involved in and over-stressed by attachment relationships, worried about losing their attachment figure or stressed about not having one. Attachment avoidance, on the other hand, reflects the extent to which an individual keeps relationships at arm's length, actively rejecting intimate attachment bonds or being convinced he or she doesn't need that. When a person scores low on both dimensions, we say they have a *secure attachment* style. Scoring high (i.e. more than 50%) on either anxiety or avoidance means one has a *preoccupied attachment* or *dismissing attachment* style, respectively. Those scoring high on both dimensions have a *fearful attachment* style (known as unresolve*d* in developmental psychology). Preoccupied, dismissing, and fearful attachment styles are sometimes simply described as *insecure attachment*.

The importance of attachment in psychology cannot be emphasized enough. Insecure attachment has been linked, in cross-sectional studies [3] and longitudinal studies [4], to numerous psychopathologies; whereas secure attachment is seen as a protective factor against them [5], [6]. Attachment is vital for healthy development, being interdependent with emotion regulation [7], self-esteem [5], self-efficacy [8], social ability [9] and even cognitive ability [10]. Finally, research abounds on the vital role attachment plays in the establishment and sustainability of intimate relationships [11].

Several authors have explored associations between attachment, psychophysiology [12] and neuroimaging [13]. Attachment developmental effects on the brain and the autonomic nervous system have been documented [14]. Globally, these studies found that negative early attachment experiences can lead to differences in the response of the bilateral amygdala and left ventral striatum during stressful situations, and to an overall higher sympathetic activation baseline.

1.2 Our previous study: introducing the Biometric Attachment Test

Within the attachment field, psychometrics plays a major role as the ground-truth for all research and the entry point for clinical treatments. In [15] we introduced the Biometric Attachment Test (BAT), a psychometric assessment designed to elicit attachment-related responses in adults through exposure to standardized sets of pictures and music. The stimuli evoke attachment-related themes such as "loss" or "separation". The test cycles participants between 15-second periods of exposure to stimuli and 20-second periods to respond using their voices to the on-screen queue "what did you feel?". Each stimulus-response cycle is followed by a 5-second blank separator. The full duration of the test is of 9.3 minutes. In our 2017 study, our input set consisted of a tablet's camera, a consumer-grade USB microphone, and a research-grade photophlysmography (PPG) and electro-dermal activity (EDA) bracelet. We used a variety of feature extraction procedures to derive facial expressions, head pose, distance from stimuli, gaze direction, paralinguistic characteristics of speech, psychologically meaningful words, heart rate variability frequency-domain and time-domain features, and skin conductance response. We showed that BAT's stimuli were capable of evoking quantifiable responses through all the aforementioned sensory modalities (i.e. channels of information) that were correlated with attachment security as measured by three validated attachment measures: the Adult Attachment Questionnaire [16], the Adult Attachment Projective Picture System [17], and the Attachment Multiple Model Interview [18]. In that paper, we discussed the challenges of building an automatic scoring algorithm for the BAT, capable of integrating all those different sensory modalities to produce single scores predictive of specific attachment dimensions. The construction of such a scoring algorithm **(III)**, as well as its evaluation and validation **(IV)**, are essential aims of the present paper. We will also present a new version of the BAT (v2), which incorporates a few new features and notably introduces a remote photoplethysmography method to detect heart rate variability from video **(I)**.

1.3 The challenge of small data in psychology samples (II)

During the last decade, we have witnessed the unlikely cross-pollination of two seemingly incompatible gardens. Computer science, through the exponential development of affective computing, crossed-over to a territory of expertise once psychology's own - human emotions and mental processes. Psychology has similarly invited itself into the heart of computing, its insights becoming central to fields such as computer vision, computer-human interaction, and multimodal fusion.

Despite a common object of study, however, a significant methodological difference exists between computer science and psychology with respect to sample acquisition: while the former generally relies on large corpora of already available data (so-called *big data*), the latter's research tends to happen at a slower pace through "bite-sized" experiments [19], the results of which are weaved together by periodic reviews and the occasional meta-analysis.

Reasons for this difference in sample sizes are manifold. They include: 1) a more stringent deontology for medical research on human subjects that most often applies to psychology but not so much to computer science; 2) geographical limitations, related to the need for physical proximity with subjects in clinical outcome studies as opposed to online crowdsourcing favored by computer science [20]; and 3) inclusion-related limitations (i.e., working with specifically diagnosed participants as opposed to general population samples).

Computer scientists might find psychology samples overly small, often relying on less than 100 subjects [21], which at first glance seems incompatible with machine learning methods. Indeed, three problems might jeopardize projects using machine learning curve-fitting (also known as regression) algorithms on very small samples: 1) rank deficiency, or the lack of sufficient number of subjects for an algorithm to converge [22]; 2) over-fitting, or the way in which models tend to adopt the form of unimportant fortuitous data features for lack of enough examples to identify such features as noise [23]; and 3) "the curse of dimensionality", or the way in which algorithms' performance worsens as the feature space increases while the sample size remains steady [24] (feature space refers to the n-dimensions where variables live).

We confronted all three of these hindrances during the development of a scoring algorithm for the BAT v2. Our samples are quite small, with 57 subjects in the learning/validation sample and 19 subjects in the independent test sample. Our feature space is comparatively huge, 1932 features before feature selection. Therefore, traditional regression algorithms failed to converge due to rank deficiency; all varieties of curve-fitting algorithms heavily over-fitted the training set, producing poor performance during validation; and both phenomena further deteriorated the more features were added.

As part of our **(II)** aim, we will describe and evaluate an original combination of techniques that helped us navigate this conundrum of small data learning.

1.4 A cross-cultural perspective on attachment and psychometric validation (IV)

Attachment as a construct has been widely studied in several different cultures, as shown by the fact that the theory was first investigated by Mary Ainsworth in Uganda [25]. Several studies show that attachment needs and their very early expressions are innate and thus arise cross-culturally [26]. Attachment security or insecurity first develops between six months and two years of age, mainly as an outcome of early interactions with our caregivers [25]. What we retain from these early events are the felt experiences and behavioral coping mechanisms, as well as psycho-physiological traits [5], [27]. These early outcomes of attachment security or insecurity are rather uninfluenced by culture. For example, children in Uganda and children in the United States both display very similar behaviors of avoidance during controlled experiments [25], and the same seems to be true for psycho-physiological traits. These early developed characteristics are carried into adulthood. As we age, however, other more complex attachment-related behaviors develop, some of which are heavily influenced by the surrounding culture. A salient such case is language.

In psychometrics, to test the validity of a new psychological assessment, a vital step is to test its *convergent validity*. A sub-type of *construct validity*, this step is to assure that the construct being assessed by the new measure is indeed the same that previously validated measures of the same construct assess. If it is the same, then testing a group of participants with both measures should produce correlated results [28].

In general, psychological assessments are validated on a per-culture basis, creating similar but nonidentical versions of the measure adjusted for different countries or regions. The reason for this is the assumption that most psychological constructs are shaped by culture and that, *ad minimum*, the distributions - and thus the cutting scores - might change from one culture to another. To test convergent validity, therefore, the common practice is to compare a newly developed assessment, that has been created with a specific culture in mind, against another assessment for the same construct that has been previously validated on that same culture.

With all this in mind we asked ourselves: what convergence level shall be expected between the BAT v2, initially created for assessing attachment in France, and the English version of the Adult Attachment Questionnaire [16], a measure validated for the United States? *A priori*, we hypothesize two different sets of results. On the one hand, scores associated with sensory modalities that reveal innate or early developed behaviors, such as psychophysiological responses or facial expressions, should shield convergent results. That is because, as we have seen, these "low-level" characteristics are rather unaffected by culture [26], [25]. On the other hand, scores from sensory modalities that reveal more complex behaviors that arise later in life, such as language, should not be convergent with scores from a measure specifically adjusted to a different culture.

Since the BAT v2 combines scores from all these sensory modalities in the assessment of adult attachment, the present work implies a remarkable challenge both from a psychology standpoint as well as from a computer science one. To the former, success in validating convergence of an attachment measure cross-culturally would add more evidence to the universality of attachment, and perhaps expand the range of attachment-related behaviors that are believed to be universal. To the latter, designing a learning algorithm capable of training on a dataset from a given culture and of successfully performing on an independent cross-cultural dataset would represent a remarkable generalizability feat.

1.5 Study aims

The purpose of the present work is manifold. **(I)** First, we will present BAT v2 and describe a remote photoplethysmography method we now include with it that enables obtaining physiological measures from video alone. **(II)** Second, we will propose and evaluate a methodology to succeed in applying machine learning to small samples. **(III)** Third, we will use said methodology to learn from a francophone sample how to create an automatic scoring algorithm for the BAT v2. **(IV)** Fourth, we will validate such an algorithm cross-culturally, testing its performance on an independent anglophone sample, which in turn serves as a convergent validity test for the BAT v2.

2 Methods

2.1 Participants

2.1.1 Our francophone learning / validation sample: This sample consisted of 57 francophone participants (45 females, 12 males) that were interviewed between March and May 2016. The sample was formed from multiple sources in different regions of France: 9 psychiatric patients recruited at University Hospital Center Sainte-Étienne and 5 recruited at the Ville-Evrard Center of Psychotherapy and Psychotrauma in Saint-Denis; 29 volunteers enrolled in Mornant, Paris, and Rouen; and 14 college students enrolled at Paris 8 University in Saint-Denis. It was intended for the sample to be as diverse as possible in terms of age (minimum 20, maximum 64, mean 35.7, SD 12.2), occupational status (10.5% unemployed, n=6/57, 33% employed, n=19/57, 35% students, n=20/57, 21% other, n=12/57), as well as relationship status (38.6% in a relationship, n=22/57, 24.5% married, n=14/57, 12.3% separated or divorced, n=7/57, 24.5% single, n=14/57) and psychopathology (24.5%, n=14/57 were patients). Since questions about ethnicity or race are not allowed in French research, we don't have information to report about the ethnic diversity of our sample. All participants signed informed consent forms and a declaration with the "Commission Nationale de l'Informatique et des Libertés" (CNIL) was filled (filling number 2159625 v 0), in accordance to best practices in French research.

2.1.2 Our anglophone test sample: This sample consisted of 19 anglophone speakers (10 females, 9 males) recruited over the summer of 2016 at the USC Institute for Creative Technologies in Playa Vista, California. Mean age was 31.2 (minimum 19, maximum 57, SD 10.2). Occupational status was 36.8% employed, n=7/19, 57.8% students, n=11/19, 5.2% other, n=1/19. Relationship status was 36.8% in a relationship, n=7/19, 21% married, n=4/19, 5.2% separated or divorced, n=1/19, 31.5% single, n=6/19. We did not recruit any clinical participants for this sample. We don't have information to report about the ethnic diversity in our sample. All participants signed informed consent forms and permission was obtained from the USC Institutional Review Board (IRB) under the identifier UP-16-00447.

2.2 Measures

2.2.1 The Adult Attachment Questionnaire: The Adult Attachment Questionnaire (AAQ) was used in the present study, the French version [29] as ground-truth for training our algorithm, the original English version [16] as the target for convergent validity. One of the first dimensional measures for attachment in adults, the AAQ its composed mainly of sentences from Hazan and Shaver's prototypical descriptions of romantic relationships from the perspective of different attachment styles [30]. It comes in the form of a 17-item questionnaire answered on a 7-point Likert-type scale (1 = strongly disagree to 7 = strongly agree). Similar to other self-report measures based on Hazan and Shaver's scale (see [31] for reviews), two orthogonal dimensions underlie the AAQ (see fig. 1), as revealed by a principal-axis factor analysis of the 17 AAQ items followed by a varimax rotation [16], [31]. The first dimension reflects the extent to which individuals exhibit attachment avoidance. The avoidance index was shown as internally consistent (Cronbach alphas = .70 for men and .74 for women). The second dimension taps individuals' level of attachment anxiety. The internal consistency of this index was .72 for men and .76 for women. According to this measurement model, prototypically secure individuals tend to score low on both dimensions. Several papers have offered additional psychometric support for the AAQ, including reliability and validity [32], [33], [34]. It possesses adequate test–retest and inter-rater reliability, as well as excellent convergent, discriminant, and predictive validity [35], [36].

Our 57 francophone participants completed the French version of the AAQ [37], and our 19 anglophone participants the English version of the AAQ [16], as a web-based questionnaire, several days prior to being tested with the BAT, to prevent priming effects. Obtained AAQ scores were scaled to the 1-100 range to facilitate interpretation.

In order to provide additional external validation for the AAQ measure, all our participants completed three additional questionnaires that evaluate theoretically interdependent constructs: the Difficulties in Emotion Regulation Scale (DERS, [38]), the Rosenberg Self-Esteem Scale (RSES, [39]), and the General Self-Efficacy Scale (GSE, [40]).

2.2.2 The Biometric Attachment Test, second version (v2): The version of the BAT used in the present study (v2) is similar to the one presented at length in [15] and briefly described in the introduction above, but includes the following important differences: First, we have added two new linguistic analysis features at the feature-extraction stage: phrase length and phrase entropy, extracted using ReaderBench (version 3.0, Mihai Dascalu, Apache License, [41]). Second, BAT v2 uses a remote photoplethysmography (rPPG) technique to extract heart rate from the video, so we have discarded the E4 bracelet that was previously required [15]. This also means we have discarded EDA. EDA was for us a useful physiological marker to measure sympathetic nervous system arousal, but since we can still accurately measure such arousal using a specific feature derived from HRV (Baevsky Stress Index, [42], [43]), EDA becomes somewhat redundant. Additionally, EDA requires special equipment like the E4 bracelet to be measured, reducing the portability of the solution. Details on the implemented rPPG technique follow.

In terms of software implementations, BAT v2 uses Openface (version 2.1.0, Carnegie Mellon University, academic or non-profit organization noncommercial research use only, [44]) to extract facial expressions, head pose, head distance and gaze direction; Linguistic Inquiry and Word Count (LIWC) dictionaries (French version [45] and English version [46], version 2007, Pennebaker Conglomerates, commercial license) for linguistic analyses; and the Cooperative Voice Analysis Repository for Speech Technologies (COVAREP, version 1.2, repository/no single license, [47]) for extracting paralinguistic features of speech.

2.3 Procedures

2.3.1 Introducing remote photoplethysmography in BAT v2 (I): Photoplethysmography (PPG) is a technique that measures variations in the absorption of light by human skin. PPG has been ubiquitously used for heart rate measurements since it is easy to use, low-cost and non-invasive. However, even more non-invasive is the technique of remote photoplethysmography (rPPG), which aims at measuring the same parameters as PPG remotely, i.e. using only the ambient light and a video camera. In the previous version of the BAT, physiological signals were recorded using a bracelet-based PPG device. However, the use of this specific device was limiting and cumbersome. In this work, we have discarded bracelet-based PPG measures in favor of the more convenient camera-based rPPG technique. Therefore, only the internal webcam of any consumer-grade device is necessary to estimate the heart rate variability-based features used in BAT v2 (see fig. 2).

Several approaches have been proposed for extracting the PPG signal from image sequences. The interested reader may refer to recent reviews on rPPG [48], [49]. In this work, we use a pipeline-based framework where regions of interest are first detected and tracked over frames, RGB channels are then combined to estimate the pulse signal, which is filtered to extract the rPPG signal [50], [51]. For each video frame, face detection is first performed using the well-known Viola-Jones face detector [52]. In order to avoid spurious movements of the detected face, a simple linear Kalman filter is used to smooth the temporal trajectories of the face location and size. Skin detection, as formulated by Conaire et al. [53], is then performed to select the skin pixels, which subsequently are spatially averaged to obtain a triplet of RGB values per frame and concatenated to obtain the RGB temporal trace. The RGB temporal traces are then preprocessed to remove the DC component of the signals dividing samples by their mean over a temporal interval. DC-normalized RGB signals are then band-pass filtered using Butterworth filter (with cut-off frequencies of 0.7 and 3.5Hz) and projected onto a plane orthogonal to the unit vector $(1, 1, 1)^T$ to remove intensity variations caused by motion noise [54]. The hypothesis is that motion-induced intensity modulations are equal for all channels. The last step is to tune an exact projection direction within the bounded region defined by the 2 axes of the plane using the alpha-tuning procedure introduced in [55]. This simple method maintains an interesting balance between complexity and efficiency. Peaks in the extracted rPPG signal are finally detected to calculate heart rate variability (HRV).

Figure 2. Taking the BAT v2 test on a Microsoft Surface Pro 4 tablet. Using only the integrated webcam and microphone, BAT v2 is capable of extracting reliable HRV features, facial expressions, linguistic and paralinguistic features of speech, gaze, and head movements.

:

2.3.2 Feature selection: As we explained in section 2.2.2, five sensory modalities were extracted from the video and audio responses to each of the BAT v2's 14 stimuli. Since each modality consists of dozens of features, the total extracted feature space of the BAT v2 is of 1932 features. Therefore, a feature selection procedure was mandatory to reduce the feature space size.

We first selected features based on the exploratory results from our previous study [15]. We then manually incorporated BAT v2 new features (see section 2.3.1). Probably because of the different nature of attachment avoidance and attachment anxiety, the features that were retained for each were slightly different. In Table I below we detail the different features that were selected for the attachment anxiety and the attachment avoidance scoring models. Please note that the purpose of this table is to improve reproducibility of our work. A description of these features is beyond the scope of this paper. To learn more about them, please refer to the references for each of the software implementations in section 2.2.2.

Through the use of feature selection our feature space was reduced to 238 features for attachment anxiety (17 features x 14 responses) and 210 features for attachment avoidance (15 features x 14 responses).

2.3.3 Developing a scoring algorithm for the BAT v2: methodology implemented to overcome specific problems associated with small data learning (II and III): To create an automatic scoring algorithm for the BAT v2, we had to implement an original combination of machine learning techniques to automatically learn how to combine all its extracted features to produce scores equivalent to the AAQ ones. To accomplish this, we only had at our disposal a small learning/validation dataset consisting of 57 francophone subjects, and we previously discussed how such small samples, common in psychology research, can hinder the performance of machine learning curve-fitting algorithms. We describe here five ways in which we've circumvented this limitation, unleashing efficient small data learning to train our scoring algorithm.

2.3.3.1 Our three-step hybrid multimodal fusion procedure

First, we developed a three-step hybrid multimodal fusion procedure (see Fig. 3). The term "hybrid" here refers to the concomitant use of feature-level fusion and decision-level fusion (for a review of multimodal fusion approaches, see [56]).

At a conceptual level, our three-step approach integrates human intelligence and machine learning in a symbiosis that has been described before as beneficial [57]. Three-step fusion is our response to "the curse of dimensionality" [58] and is analogous to spoon-feeding the data to the algorithm. No algorithm, as of today, can figure out alone how to make sense of a large set of features such as ours having so few cases

Table I. FEATURES RETAINED AFTER FEATURE SELECTION

Input source	Modality	Implementation	Attachment anxiety features	Attachment avoidance features
Video	Facial expressions	Openface	inner brow raiser	inner brow raiser
			lip corner puller	lip corner puller
			blink	blink
	Head pose		pitch	pitch
			yaw	yaw
	HRV features	rPPG	pNN50	Baevsky stress index
			HF	
Audio	Linguistic analysis	LIWC	valence	present tense
		ReaderBench	phrase length	phrase length
			phrase entropy	phrase entropy
	Paralinguistic features	COVAREP	Normalised Amplitude Quotient	Normalised Amplitude Quotient
			Quasi-open Quotient	Quasi-open Quotient
			H1-H2 (*)	H1-H2 (*)
			Parabolic Spectral Parameter	Parabolic Spectral Parameter
			Maxima Dispersion Quotient	Maxima Dispersion Quotient
			Spectral Tilt Correlate	Spectral Tilt Correlate
			Vowel space	

* Difference in amplitude of the first two harmonics of the differentiated glottal source spectrum.

to learn from. Human intelligence can help by providing a theory-based framework that simplifies the work of learning algorithms, such as our three-step model.

In the first step, we used a feature-level, early-fusion approach. This step is about learning the best solution to combine together the different features coming from the various sensory modalities within each of the responses to the 14 BAT v2 stimuli. For example, here the algorithm learns how to integrate different HRV-derived features that were extracted from the response to theme 3, producing a single "theme 3 HRV" score which better predicts our ground-truth scores. This process drastically reduces the feature space (from the initial 238/210 to a more palatable 70) outputting one score per modality per theme (5x14). This first fusion step has to deal with wildly different scales as well as outliers and measuring errors. To accomplish this, we implemented a linear regression [59] with a customized learning objective and optimization function. Specifically, in section 2.3.3.2 we will explain how we found Pearson linear correlation coefficient maximization to be the best learning objective for this step. And in section 2.3.3.3, we will explain how we implemented the Generalized Pattern Search (GPS) optimization algorithm to make sure we could find the best possible combination of our features (the best coefficients for our regression) circumventing rank deficiency. Regularization is not required at this fusion step since the chosen learning objective is not prone to over-fitting.

During the second step of our fusion procedure, a feature-level early-fusion is applied, where modality-scores are fused at the heart of each theme to produce 14 per-theme scores. For instance, here the algorithm learns how to combine the facial expressions, head pose, HRV, linguistic and paralinguistic scores for theme 3 produced by the previous fusion step, in a way that can better predict the ground-truth attachment scores. We've implemented for this step a shallow artificial neural-network, "a computational model based on biological neural networks [...] that consists of an interconnected group of artificial neurons and processes information using a connectionist approach to computation" [60]. Our neural network's structure included an input layer recruiting the same amount of neurons as there were sensory modalities to be processed (5); a hidden layer half the size of the input layer (rounded to 3); and an output layer possessing only one neuron, which outputs the theme score. The learning objective for the network was the more traditional mean squared error (MSE) minimization. In section 2.3.3.4, we will describe how we implemented Bayesian regularization on this neural network, to allow the algorithm to find the best hyperparameters for regularization as a byproduct of the learning process, instead of having to split the already small learning sample to create a separate validation set for that.

In the third and last step of our fusion procedure, a decision-level, late-fusion is applied. This step is about finding the right combination of the 14 themes scores generated in the previous step that can better predict the attachment anxiety and attachment avoidance ground-truth scores from the AAQ. Here we use a smaller, shallow artificial neural network, again with MSE minimization as our learning objective, recruiting 14 neurons in the input layer, 7 in the hidden layer and 1 in the output layer, and we implemented Bayesian regularization again.

Since each step of this fusion procedure was designed to strongly prevent overfitting, we did not have to implement a cross-validation scheme.

2.3.3.2 We maximized Pearson linear correlation coefficient as the learning objective of our linear regression

An important aspect to consider for small data learning is our learning objective itself. Common curve-fitting algorithms by default rely on error minimizing objectives, implementing one of the few popular error metrics such as mean squared error (MSE) and attempting to reduce it in each iteration. While this works well on larger samples, we've found it to be rather inefficient working with small data due in part to a tendency to reify outliers. While outliers can be manually removed, our intention was to create an automatic scoring algorithm capable of dealing with all kinds of data, so that solution was discarded. Another disadvantage of error metrics is that they are scale-dependent. Z-scoring can help with the issue of unequal scaling, but we've found empirically that z-scoring our inputs led to a decrease in performance, probably due to z-scoring's assumption that the distribution of the

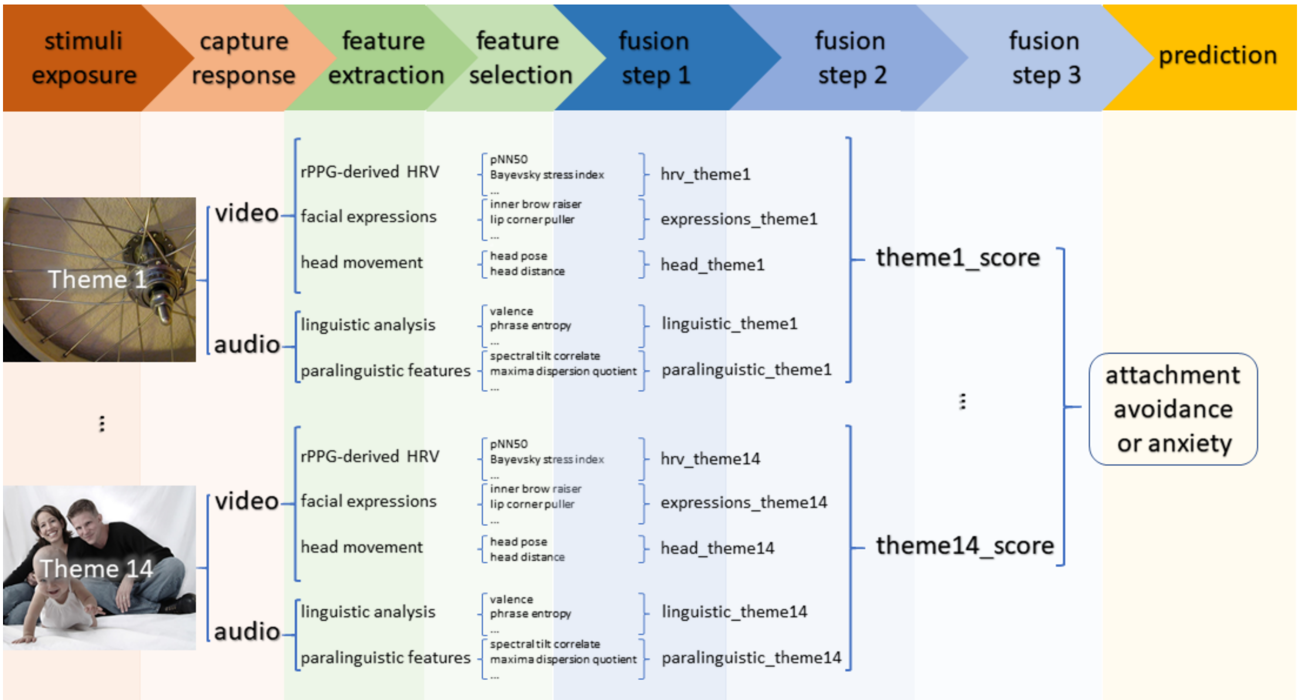


Figure 3. BAT v2 procedure, from exposure to three-step hybrid multimodal fusion. Sensory modalities are extracted from video and audio signals, with several features per modality. After selecting the best performing features, step 1 of our fusion integrates features of each modality; step 2 integrates sensory modalities into theme scores, and step 3 integrates theme scores into attachment anxiety/avoidance scores.

sample is representative of the population, clearly not the case in such small samples. We tried to find an alternative to error-metrics-based learning objectives; one focused on the *general relationship* between predictions and ground-truth as opposed to the *individual distances* among discrete predictions and their corresponding ground-truth data-points. A learning objective with a tendency to bypass outliers, thus preventing over-fitting, and one unbothered by different feature scales. We found that maximizing the Pearson linear correlation coefficient was the perfect learning objective to fulfill these requirements. Defined as:

$$\rho(a,b) = \frac{\sum_{i=1}^{n}\left(X_{a,i}-\overline{X}_a\right)\left(Y_{b,i}-\overline{Y}_b\right)}{\left\{\sum_{i=1}^{n}\left(X_{a,i}-\overline{X}_a\right)^2\sum_{j=1}^{n}\left(Y_{b,j}-\overline{Y}_b\right)^2\right\}^{\frac{1}{2}}}$$

Formula 1 Pearson linear correlation coefficient

Where n is the number of observations, values of the Pearson linear correlation coefficient can range from -1 to $+1$. A value of -1 indicates perfect negative correlation, while a value of $+1$ indicates perfect positive correlation. A value of 0 indicates no correlation between predictions and ground-truth. We've implemented this learning objective in the first step of our three-step hybrid multimodal fusion procedure described above.

2.3.3.3 We used Generalized Pattern Search for optimization of our linear regression

When performing feature extraction on small data, the number of features quickly overcomes the number of observations. This, in turn, produces rank deficiency [22], rendering impossible for many algorithms to estimate the parameters of the model. Some optimization algorithms can still find coefficients but their estimations are local optima, i.e. possible solutions, not the best possible solution. In our empirical experience this proved problematic, for as long as the best coefficient for each added feature could not be found, adding more features beyond the number of observations only amounted to a decrease in prediction performance.

Enter the Generalized Pattern Search (GPS) algorithm [61]: GPS is a method for solving optimization problems that does not require any information about the gradient or higher derivatives of the objective function. At each optimization step, GPS searches a set of points, so-called a *mesh*, around the point computed at the previous step of the algorithm. The mesh is formed by adding the current point to a scalar multiple of a set of vectors called a *pattern*. If the pattern search algorithm finds a point in the mesh that improves the objective function at the current point, the new point becomes the current point at the next step of the algorithm. The GPS algorithm uses fixed direction vectors. This algorithm can work even with objective functions which are not convex as well as for nonlinear problems.

Using the GPS optimization algorithm at multiple initial points $x0$, we can find the *best possible* solution (i.e., the global optimum) for the fusion of our large set of features without requiring our sample to be large. We've used Matlab's implementation of the GPS pattern search algorithm, which is robust concerning objective function failures (i.e., evaluations resulting in NaN, Inf, or complex values), during the first phase of our three-step fusion procedure, described above [62], [61].

2.3.3.4 We applied Bayesian regularization to our neural networks

Dealing with over-fitting is first and foremost the mission of regularization, which entails introducing bias (i.e., compromise) during a learning procedure, forcing the algorithm to under-perform [63]. A regularized regression algorithm tends to fit gross data features while bypassing more peculiar ones, resulting in a less data-accurate, simpler model. According to "Occam's razor" principle, simpler models should be preferred to unnecessarily complex ones [64], since they tend to be more stable and more generalizable.

For regularization to be effective, its hyperparameters need to be tuned either manually or programmatically as part of a separate learning procedure done on an independent validation sample. The main regularization hyperparameter is lambda and sets the amount of bias imposed on the algorithm. The problem with standard regularization in small data learning is that we don't have the luxury of cases to spare to set aside a dedicated validation set. We might be tempted to adjust lambda based on prediction performance on the learning set, but this leads to an overoptimistic performance and thus to low generalizability.

Our response to this problem is implementing Bayesian regularization [65]. Under Bayes' rule, complex models are automatically penalized. That is why Bayesian methods are said to embody Occam's razor principle [64]. Bayesian regularization takes place within the Levenberg-Marquardt algorithm [66] used by both of our neural networks during step 2 and 3 of our three-step fusion procedure. Backpropagation, a method used to iterative and recursively train artificial neural networks following a gradient descent approach, is used to calculate the Jacobian jX (matrix of first-order partial derivatives of a vector-valued function) of performance with respect to the weight and bias variables X. Each variable is adjusted according to Levenberg-Marquardt. The adaptive Marquardt adjustment parameter mu is increased by a chosen factor until the change results in reduced performance value. The change is then applied to the network, and mu is decreased by a specified factor. In our implementation training automatically stops when 5000 iterations are completed.

From a psychological standpoint, a byproduct of regularization is incentivizing our algorithm to focus on less specific, more generally-occurring behavioral traits, which theoretically should favor universal behaviors such as attachment behaviors.

2.3.3.5 We amplified variety in the inputs

To learn from a small number of people, one approach is to observe them under almost identical circumstances for a number of trials. This provides an opportunity to differentiate behaviors that occurred in only one trial from those that happened iteratively. The former can be interpreted as fortuitous whereas the latter can safely be assumed to have a deeper meaning.

In the case of the BAT v2 we accomplished this by exposing our participants to several different stimuli-sets. In [15] we described how our stimuli-set design procedure is standardized and we presented evidence that participants tend to have similar reactions to the different stimuli-sets which are a product of that standardized procedure. Each BAT v2 stimuli-set depicts the same range of attachment-related themes in the same order, but uses different music and picture stimuli to do so. Participants in our francophone training/validation sample were exposed to two different stimuli-sets, and, in addition, were shown a third randomized stimuli-set, consisting of an on-the-fly computer selection among previously validated stimuli, so that no participant experienced the exact same third stimuli-set. The intention was to facilitate our algorithm's dismissal of unintended stimuli-related behaviors (e.g. reactions to a specific color in a given picture) as mere noise while favoring recognition of theme-related behaviors (e.g. reactions to the situation being depicted) as meaningful data features.

Once our participants were exposed to the three stimuli sets, their responses to each were treated during learning as if they were coming from different people altogether. Implementing this technique we were able to exploit our original N=57 learning/validation sample to produce 148 different responses, spanning a more diverse range of behaviors and allowing our algorithm to more efficiently separate noise from important data features (it would have been 171 responses, but not all participants were able to stay long enough to take the BAT v2 three times).

Another way to introduce variety is to focus on diversity during the construction of the training/validation sample itself. In particular, we shall be vigilant to find participants that differ considerably between them on dimensions associated with the ground-truth variable we attempt to predict. In our case for example, attachment can potentially change over the life span [67], it can be affected by our marital situation, and it can be significantly different among psychotherapists [68]. Therefore, during recruitment we intended to have a large age span, to include people in a variety of marital situations, and to prevent excessive recruitment of psychologists. These precautions helped our algorithm single out real attachment-related behaviors from more idiosyncratic behaviors related to age, marital status or the sensibilities associated with psychotherapists.

2.4. Analyses and metrics

In the following section, we will detail our results. First, in order to justify our choice of the AAQ as ground-truth for attachment, we will analyze its convergence with the related DERS, RSES, and GSE questionnaires, using Pearson linear correlation coefficients. For this analysis, we have pooled together both of our samples to increase statistical power (N=76).

Second, we will present the prediction performance results of our BAT v2 scoring algorithm on the francophone learning/validation sample, and on the anglophone test sample **(IV)**.

For both performance analyses, the algorithm was trained on the francophone learning/validation sample, and both of our neural networks (second and third fusion steps) were trained 100 times with the outputted models averaged. The performance metrics calculated to compare the BAT v2 predicted scores vs. the AAQ ground-truth scores are the Pearson linear correlation coefficient, the Concordant Correlation Coefficient, the Mean Absolute Error, and the Root Mean Squared Error.

Third, we will analyze the performance loss resulting from removing each of the sensory modalities from the BAT v2 (e.g. removing facial expressions) **(III)**. For this analysis, we will use the Pearson linear correlation coefficient as our metric and present the prediction performance results on the anglophone test sample.

Fourth, we will analyze the performance loss resulting from replacing each of the techniques we suggested to improve small data learning with standard procedures **(II)**. In concrete, for exploring how removing the GPS algorithm from fusion step 1 would alter performance, we used the BFGS optimization algorithm instead [69]. To explore how avoiding amplifying variety in the inputs affected performance, we only used the responses to the first BAT v2 stimuli set for all participants in the learning set (n=57). To explore how discarding our three-step fusion procedure would affect performance, we replaced it with a one-step fusion scheme consisting of a shallow neural network with Bayesian regularization. To explore how removing the Pearson correlation coefficient maximization learning objective from fusion step 1 would affect performance, we used instead an MSE minimization objective. Finally, to explore the effects of removing Bayesian regularization from our neural networks at fusion steps 2 and 3, we used instead BFGS optimized, regularized neural networks, iteratively looking for the best lambda parameter during 100 iterations using a validation set consisting of 20% of the learning set. For this analysis, we will use the Pearson linear correlation coefficient as our metric and present the prediction performance results on the anglophone test sample. All analyses were conducted using Matlab (version R2018a, The MathWorks Inc., commercial license).

3 Results

3.1 AAQ external convergent validity results

Emotion dysregulation as measured by DERS was correlated with both AAQ's attachment anxiety ($r(76) = .55$, $p < 0.001$) and avoidance ($r(76) = .43$, $p < 0.001$). Self-esteem as measured by RSES was negatively correlated with both AAQ's attachment anxiety ($r(76) = -.42$, $p < 0.001$) and avoidance ($r(76) = -.44$, $p < 0.001$). Self-efficacy as measured by GSE was negatively correlated with both AAQ's attachment anxiety ($r(76) = -.46$, $p < 0.001$) and avoidance ($r(76) = -.42$, $p < 0.001$). These results confirm our choice of the AAQ as ground-truth for attachment.

3.2 BAT v2's prediction performance (IV)

Predictions of attachment anxiety and avoidance made by the BAT v2's scoring algorithm on the francophone training / validation (N = 57) sample were correlated with AAQ's attachment anxiety ($r(57) = .71$, $p < 0.001$) and avoidance ($r(57) = .72$, $p < 0.001$). Predictions of attachment anxiety and avoidance made on the anglophone independent test sample (N = 19) were correlated with AAQ's attachment anxiety ($r(19) =.79$, $p < 0.001$) and avoidance ($r(19) = .74$, $p < 0.001$). The other performance metrics are exposed on Table II.

3.3 Prediction performance loss on the anglophone test sample resulting from removing sensory modalities from fusion (III)

The best performance on the anglophone independent test sample was obtained by fusing all sensory modalities (anxiety $r(19) =.79$, avoidance $r(19) = .74$). Table III shows BAT v2 scores' Pearson correlation with AAQ's scores as each sensory modality is removed from fusion while all others are retained (see Figure 4 for a simplified illustration of this loss).

3.4 Prediction performance on the anglophone test sample resulting from replacing each of the recommended methodologies with standard practices (II)

The best performance on the anglophone independent test sample was obtained by following all the recommendations described in section 2.3.3 (anxiety $r(19) =.79$, avoidance $r(19) = .74$). Table IV shows BAT v2 scores' Pearson correlation with AAQ's scores as, one at a time, each of the recommended methodologies is replaced with its most common methodological counterpart, while all other recommended methodologies are still applied.

4 Discussion

In this paper, we presented a second version of the BAT (v2), which notably uses an rPPG method to obtain heart rate data out of video. We also presented and evaluated a machine learning methodology capable of training an automatic scoring algorithm for the BAT v2, using a small learning sample. We then tested the scoring algorithm's convergent validity and generalizability using a cross-cultural independent test sample, finding promising results. Throughout the following sections, we will summarize our findings and interpret them from the vantage points of psychology, specifically attachment theory, and computer science, specifically machine learning.

4.1 The remote photoplethysmography method (I)

The first aim of the present work was to incorporate a remote photoplethysmography method to the BAT, enabling extraction of heart rate data out of video. Thanks to this method, the new BAT v2 can be deployed on consumer devices using only the webcam and microphone, without requiring cumbersome special devices such as smart bands.

The BAT test was perfectly suited for the application of this method, for participants remain relatively steady during the 9.3

Table II. PREDICTION PERFORMANCE

Sample	AAQ dimension	N	Pearson r	p-value	CCC	MAE	RMSE
francophone sample (training/validation)	Anxiety (0-100)	57	0.71	<0.001	0.66	9.69	12.37
	Avoidance (0-100)		0.72	<0.001	0.66	10.07	12.12
anglophone sample (test)	Anxiety (0-100)	19	0.79	<0.001	0.60	14.90	17.94
	Avoidance (0-100)		0.74	<0.001	0.37	22.35	26.25

CCC = Concordant Correlation Coefficient; MAE = Mean Absolute Error; RMSE = Root Mean Squared Error

Table III. PERFORMANCE LOSS WHEN REMOVING SENSORY MODALITIES

Sensory modalities included in fusion	AAQ dimension	Pearson r	Performance percentage
with all sensory modalities	Anxiety	0.79	100.0
	Avoidance	0.74	100.0
without facial expressions	Anxiety	0.54	67.9
	Avoidance	0.49	67.1
without head pose	Anxiety	0.74	94.1
	Avoidance	0.42	56.4
without rPPG	Anxiety	0.74	93.7
	Avoidance	0.54	72.8
without linguistic	Anxiety	0.71	89.5
	Avoidance	0.64	87.2
without paralinguistic	Anxiety	0.70	88.3
	Avoidance	0.22	30.1

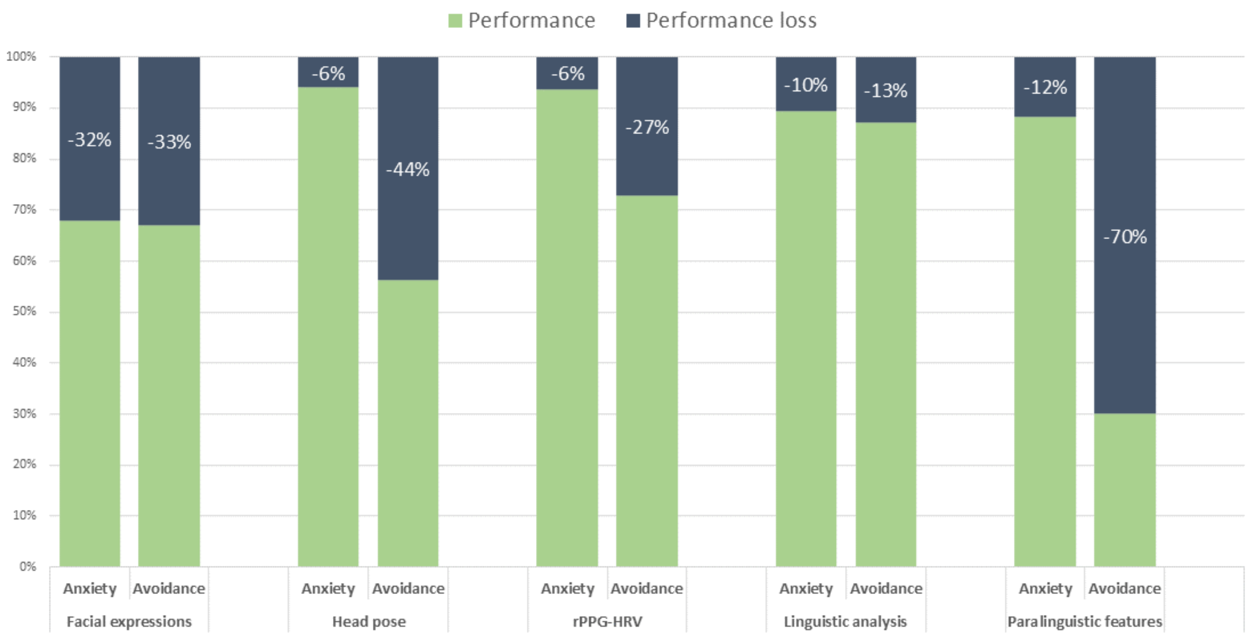


Figure 4. Percentage of prediction performance lost when removing each sensory modality from the attachment anxiety / avoidance fusion models. 100% is equivalent to the maximum performance achieved on each model by including all sensory modalities in the fusion (anxiety r(19) = .79, p < 0.001; avoidance r(19) = .74, p < 0.001).

minutes of stimuli exposure, always in front of the screen, and therefore in front of the camera.

The HRV features derived from the rPPG extracted heart pulse signal were important to the overall performance of our models. As seen in result 3.3, removing rPPG derived HRV features from our fusion algorithm produced a 6% decrease in the performance of our anxiety model and a 27% decrease in our avoidance model (we will discuss probable reasons for the difference between them in section 4.3).

This result attests to the quality of the extracted signal, which seems at least sufficient for our purposes of analyzing patterns of autonomic system reactions as a person is exposed to different attachment-related themes.

Table IV. PERFORMANCE LOSS WHEN REPLACING RECOMMENDED METHODS FOR STANDARD ONES

Bypassed recommended methodology	AAQ dimension	Pearson r	Performance percentage
Using all recommended methodologies	Anxiety	0.79	100.0
	Avoidance	0.74	100.0
No GPS optimization algorithm	Anxiety	0.63	79.1
	Avoidance	0.37	50.2
No amplifying variety of inputs	Anxiety	0.36	46.1
	Avoidance	0.40	53.7
No three-step fusion scheme	Anxiety	0.42	53.8
	Avoidance	0.06	7.6
No Pearson learning objective	Anxiety	0.12	15.3
	Avoidance	0.21	28.0
No Bayesian regularization	Anxiety	-0.30	0
	Avoidance	0.14	19.5

4.2 Our proposed methodology to overcome small data challenges in machine learning (II)

Our second aim was to describe and evaluate a set of methods we employed to overcome small data challenges in machine learning. From a machine learning perspective, the efficiency of our presented small data learning methodology was demonstrated. Our results (3.4) show that replacing even one of the proposed methods with more standard ones can induce a performance loss that spans from 21% up to 100%. More precisely, we found that, of the 5 proposed methods, replacing Bayesian regularization with standard regularization was the most damming - causing a loss in performance of 100% and 80.5% for the attachment anxiety and attachment avoidance models, respectively. Replacing our three-step fusion scheme with a one-step fusion approach also heavily penalized performance, losing 46.2% and 92.4% of the performance for the anxiety/avoidance models. Beyond anything else, though, these results underline the *combined power* of these methods.

We believe their implementation in the field could help further tending an interdisciplinary bridge between psychology and computer science. Small psychology studies abound, and being able to learn generalizable and robust predictive models from those small samples could be a gateway into exciting new potential collaborations. Additionally, and to the best of our knowledge, our results constitute the first preliminary validation of an automatic psychometric test powered by multimodal fusion and machine learning.

4.3 Development of an automatic scoring algorithm for the BAT v2 (III)

The third aim of this work was to use the aforementioned methodology to automatically learn from a francophone sample a scoring algorithm for the BAT v2, capable of fusing together hundreds of features into simple attachment anxiety/attachment avoidance scores inspired by our chosen ground-truth, the AAQ.

Since BAT v2 was to learn how to score based on AAQ dimensions, we decided to perform our own evaluation of the AAQ's construct validity. Our results (3.1) add support to the validity evidence for the AAQ measure. As predicted by attachment theory, both attachment anxiety and attachment avoidance as measured by the AAQ were associated with higher emotion dysregulation, lower self-esteem, and lower self-efficacy.

In terms of our new automatic scoring algorithm, we found (section 3.3) that removing any of the sensory modalities from it produced a marked decrease in performance, and that using them all together produced the highest convergence with AAQ ground-truth scores, proof of each feature's importance to the overall scoring model.

Specifically, facial expressions were important to detect both attachment anxiety and avoidance: when removed from our fusion algorithm, it lost more than 30% of its performance. For detecting avoidance, head movements were crucial, and removing them decreased BAT v2's performance by 44%. Paralinguistic features - "the tone" in which things are said - were likewise very important for avoidance detection since, when removed, our avoidance model's performance was undercut by a massive 70%. According to the attachment literature [5], people high in attachment avoidance tend to consciously conceal their emotions from others, whereas people high in anxiety tend to express their anxiety openly. This is probably why unconscious, automatic behavioral features such as head movements, sympathetic activation (detected by the Baevsky Stress Index derived from HRV), and, particularly, vocal stress detected by paralinguistic analysis of speech were more decisive for the avoidance detection model than they were for the anxiety model. Linguistic analysis contribution to the overall models was unsurprisingly more modest: removing it undercut performance by slightly above 10%, supporting the notion that attachment-related reactions are mostly non-verbal.

4.4 Evaluation and cross-cultural validation of BAT v2's scoring algorithm (IV)

The fourth aim of this work was to evaluate the robustness and generalizability of our new scoring algorithm using an anglophone, cross-cultural test sample.

From an attachment theory perspective and focusing on psychometrics, our main finding (section 3.2) is that the BAT v2 is indeed measuring very similar constructs compared to the validated AAQ.

The high correlation for both attachment anxiety (.79) and avoidance (.74) dimensions between the BAT v2 and the AAQ in the anglophone sample is remarkable when considering that the BAT v2 scoring algorithm was trained on a francophone sample.

These results constitute the first psychometric evidence in support of the use of the BAT v2 and its automatic scoring algorithm as a valid measure for adult attachment.

All features that were learned from the francophone sample and that were important for predicting attachment anxiety and avoidance in that sample were also important to predict attachment in the anglophone sample, supporting the notion, described in the introduction (1.4) and well described in the attachment literature, that attachment is a universal phenomenon that transcends culture.

One unexpected result contradicted a hypothesis we posited during that same section, that linguistic features are language-dependent and culture-dependent and thus patterns in a language related to attachment in one culture should not translate well to another culture. In opposition to that, our linguistic features, such as the use of the present tense, the entropy level and length of phrases, and the balance of positive vs. negative words used, were learned from a francophone sample yet were still powerful predictors for attachment dimensions when used to score an anglophone sample. This is remarkable, from an attachment theory standpoint, as it suggests there could be attachment-based dynamics underlying language that transcend the specificity of the language at hand. A good vantage point for the understanding of this phenomenon might be *speech act theory*, which posits a performative function of language and communication [70], seen as partially universal [71].

4.5 Advantages of the use of BAT v2 in adult attachment assessment

Our results also signal potential advantages of using BAT v2 in adult attachment assessment, in both research and clinical settings.

We would like to illustrate its potential contribution to research by commenting on a recent attachment theory controversy. In 2014, Groh and colleagues' provocative results rocked the attachment theory community [72]. Their work with a cohort of 819 participants studied longitudinally since birth till late adolescence showed that the participants' adult attachment measures (at 18 years) were almost not correlated at all to the participants' childhood attachment ones (at 15 months), challenging the notion of attachment continuity and stability across the life span - one of the theory's main tenets. From a multimodal perspective, however, we might read their results as meaning something else entirely.

Multimodal fusion imitates human cognition by cross-verifying hypotheses about what's happening using different sensory modalities at once [73]. For instance, an isolated phrase might be suggestive of anger, but uttered in a certain tone and accompanied by a certain facial expression and body posture it might suggest lightness and humor. The adult attachment measure implemented in Groh et al. study was the gold-standard Adult Attachment Interview (AAI, [74]), in which scores are obtained through a complex coding process applied to a structured interview transcript, containing what was *said* by the participant plus a minimalistic set of their annotated behaviors (e.g. crying). It is arguably a *unimodal* test since the analyzed information comes mostly from language. Except for rare cases in which the AAI interviewer doubles as the AAI coder, coders don't get to see or hear the participant at all. Compare this to the Strange Situation Procedure (SSP, [25]), the gold-standard test for children used in the same Groh et al. study, in which kids are *directly observed* through a one-way mirror as they experience separation from and then reunion with their mothers: movements, sounds, facial expressions, gaze... the SSP coder has access to all of these sensory modalities, being able to cross-verify hypotheses across them, and is able to observe different behaviors happening *simultaneously*, revealing their *combined* meaning. Attachment-related reactions, psycho-physiological traces, and behavioral patterns develop most critically between 6 months and 2 years of age, so they should be expected to be mostly non-verbal [5].

So what if the lack of correlation between SSP results at 15 months and AAI results at 18 years were revealing of an AAI failure to pick up the expected longitudinal traces due to its almost exclusive reliance on transcribed language? We can hypothesize that a multi-modal assessment such as the BAT v2, which is more similar in nature to the way the SSP works, could be better suited to find the "lost attachment traces", though of course, rigorous research needs to be performed to test such a hypothesis.

In a clinical setting, attachment is often assessed through the therapeutic relationship. The therapist tends to play the role of a temporary attachment figure for the patient and the latter's attachment anxiety and avoidance tends to arise and play out in the relationship [5]. Like the expert evaluators in the SSP, the clinician directly examines complex patterns of behavior: a phrase being repeated (linguistic analysis) in a specific tone of voice (paralinguistic features), the physical distance the patient chooses for their chair with respect to the therapist's (distance from stimulus), the fact that they keep looking away the whole time (head pitch/yaw)... the *combination* of all these behavioral features might hit the therapist as their patient relationally "pushing them away" (i.e. attachment avoidance). Clinician assessments offer an undeniable multimodal advantage over unimodal tests like the aforementioned AAI, but clinician assessments suffer from subjectivity, and different clinicians with different sensibilities will judge the attachment characteristics of a client differently.

The BAT v2 mimics this clinical multimodal advantage, analyzing many of the same sensory modalities a clinician assesses, but in an objective (i.e. replicable) manner.

Finally, both existing attachment tests as well as clinician assessments, good as they might be, miss the very earliest of attachment-related reactions: autonomic nervous system responses. Several studies suggest that physiological responses are sometimes the only way of differentiating among attachment characteristics, for example for separating a secure response from a dismissing (i.e. high in avoidance) one. Both a secure and a dismissing person might show a relatively calm facial expression while being reminded of attachment-related content (e.g. last separation), and yet the dismissing person will tend to additionally show an elevated physiological level of distress [75]. The importance of physiology for uncovering avoidance was seen in our results: removing rPPG features from our fusion algorithm represented a 27% performance loss in the attachment avoidance scoring model.

The BAT v2 uses this physiological information in conjunction with all the other extracted behavioral features to produce a multifaceted assessment that might rival the nuance of a psychotherapist's analysis while offering the objectivity of a blood test.

4.6 Limitations

Arguably and despite all the procedures implemented in this work, our test sample is rather small [76]. It could be the case for example that our algorithm would not work as well on another anglophone group.

Additionally, we want to acknowledge here that albeit self-reports (questionnaires) such as the AAQ are common for measuring attachment in social psychology, they are somewhat controversial in developmental psychology, which favors third-party assessments such as the AAI. In addition, the AAQ measure is designed to assess attachment to romantic partners, whereas other measures of adult attachment assess attachment to parents, constructs thought to be related but not interchangeable [36]. Therefore, developmental psychologists could contest our construct validity claims on those grounds. To such critique, we respond with our AAQ / DERS / RSES / GSE convergence results, which support the use of the AAQ as ground-truth for adult attachment.

Attentive readers might find intriguing that performance was worse, in terms of the Pearson linear correlation coefficient, on the francophone learning/validation sample compared to the anglophone test sample. We invite the reader to observe the other metrics that take error into account: CCC, MAE, and RMSE. They show an overall performance advantage on the francophone learning/validation sample, as expected. We explain the Pearson advantage for the anglophone test sample by the audio and video quality of that sample, which was somewhat better than in the francophone learning/validation sample.

Finally, we would like to point out that although Pearson linear correlations were high between BAT v2's scores and AAQ's scores for both attachment anxiety and avoidance, MAE and RMSE metrics - which focus on error rather than relationship - as well as CCC - which measures correlation but penalizes error - revealed a steep difference between the performance of our anxiety model vs. that of our avoidance model, with the latter showing a MAE of 22.35 and an RMSE of 26.25, pretty high given the 0-100 scale of the predicted dimension. Therefore, improvements should be made to the avoidance model in the future in order to reduce the error rate.

4.7 Future directions

At the time of writing, we are finishing the development of a Microsoft Xamarin [77] multi-platform application for iOS, Android, MacOs and Windows 10. Our application is capable of presenting different BAT v2 stimuli sets, analyzing video and audio feeds in real time, automatically extracting all necessary features, and predicting attachment scores.

We have three other projects in the queue at the moment: first, a large validation study using an attachment questionnaire as ground-truth with a far larger sample recruited online. For this, we will be using the new multi-platform application running on the participants' tablets or computers. This shall allow us to further assess convergence validity as well as test-retest reliability.

Second, we intend to use the aforementioned Xamarin application with a cohort that has taken the AAI in order for the BAT v2 to learn to classify attachment based on the patterns that better predict AAI classifications.

Lastly, using our small data learning methodology, in the future we expect to train different scoring algorithms for the BAT v2 based on different existing attachment validated tests. Specifically, we wish to have the algorithm trained over an adult sample previously evaluated with the SSP during childhood. We hypothesize that would result in a new BAT v2 scoring model capable of picking up longitudinal traces of attachment, that could, in turn, contribute to resolving the longitudinal traces controversy mentioned in section 4.5. Evidently, empirical research needs to be conducted to test such a hypothesis.

REFERENCES

[1] J. Cassidy and P. R. Shaver, *Handbook of Attachment, Third Edition: Theory, Research, and Clinical Applications*. The Guilford Press, 2016.

[2] D. W. Griffin and K. Bartholomew, “Models of the self and other: Fundamental dimensions underlying measures of adult attachment,” *Journal of Personality and Social Psychology*, vol. 67, no. 3, pp. 430–445, 1994. [Online]. Available: https://doi.org/10.1037%2F0022-3514.67.3.430

[3] M. Mikulincer and P. R. Shaver, “An attachment perspective on psychopathology,” *World Psychiatry*, vol. 11, no. 1, pp. 11–15, feb 2012. [Online]. Available: https://doi.org/10.1016%2Fj.wpsyc.2012.01.003

[4] L. A. Sroufe, B. Coffino, and E. A. Carlson, “Conceptualizing the role of early experience: Lessons from the Minnesota longitudinal study,” *Developmental Review*, vol. 30, no. 1, pp. 36–51, mar 2010. [Online]. Available: https://doi.org/10.1016%2Fj.dr.2009.12.002

[5] D. P. Brown and D. S. Elliot, *Attachment disturbances in adults: treatment for comprehensive repair*. W.W. Norton & Company, 2016.

[6] F. Parra, C. George, K. Kalalou, and D. Januel, “Ideal Parent Figure method in the treatment of complex posttraumatic stress disorder related to childhood trauma: a pilot study,” *European Journal of Psychotraumatology*, vol. 8, no. 1, p. 1400879, jan 2017. [Online]. Available: https://doi.org/10.1080%2F20008198.2017.1400879

[7] M. Mikulincer and P. R. Shaver, “Attachment orientations and emotion regulation,” *Current Opinion in Psychology*, vol. 25, pp. 6–10, feb 2019. [Online]. Available: https://doi.org/10.1016%2Fj.copsyc.2018.02.006

[8] S. L. Wright, K. M. Perrone-McGovern, J. N. Boo, and A. V. White, “Influential Factors in Academic and Career Self-Efficacy: Attachment Supports, and Career Barriers,” *Journal of Counseling & Development*, vol. 92, no. 1, pp. 36–46, jan 2014. [Online]. Available: https://doi.org/10.1002%2Fj.1556-6676.2014.00128.x

[9] L. A. Sroufe, B. Coffino, and E. A. Carlson, “Conceptualizing the role of early experience: Lessons from the Minnesota longitudinal study,” *Developmental Review*, vol. 30, no. 1, pp. 36–51, mar 2010. [Online]. Available: https://doi.org/10.1016%2Fj.dr.2009.12.002

[10] K. Bergman, P. Sarkar, V. Glover, and T. G. O'Connor, “Maternal Prenatal Cortisol and Infant Cognitive Development: Moderation by Infant–Mother Attachment,” *Biological Psychiatry*, vol. 67, no. 11, pp. 1026–1032, jun 2010. [Online]. Available: https://doi.org/10.1016%2Fj.biopsych.2010.01.002

[11] J. A. Feeney, “Adult romantic attachment: developments in the study of couple relationships,” in *Handbook of Attachment, Third Edition: Theory, Research, and Clinical Applications*. Guilford New York, NY, 2016, pp. 435–463.

[12] P. Dias, I. Soares, J. Klein, J. P. Cunha, and G. I. Roisman, “Autonomic correlates of attachment insecurity in a sample of women with eating disorders,” *Attachment & Human Development*, vol. 13, no. 2, pp. 155–167, mar 2011. [Online]. Available: https://doi.org/10.1080%2F14616734.2011.554005

[13] I. Bretherton and K. A. Munholland, “The Internal Working Model Construct in Light of Contemporary Neuroimaging Research,” in *Handbook of Attachment, Third Edition: Theory, Research, and Clinical Applications*. Guilford New York, NY, 2016, pp. 63–90.

[14] E. Lemche, V. P. Giampietro, S. A. Surguladze, E. J. Amaro, C. M. Andrew, S. C. Williams, M. J. Brammer, N. Lawrence, M. A. Maier, T. A. Russell, A. Simmons, C. Ecker, P. Joraschky, and M. L. Phillips, “Human attachment security is mediated by the amygdala: Evidence from combined fMRI and psychophysiological measures,” *Human Brain Mapping*, vol. 27, no. 8, pp. 623–635, 2006. [Online]. Available: https://doi.org/10.1002%2Fhbm.20206

[15] F. Parra, R. Miljkovitch, G. Persiaux, M. Morales, and S. Scherer, “The Multimodal Assessment of Adult Attachment Security: Developing the Biometric Attachment Test,” *Journal of Medical Internet Research*, vol. 19, no. 4, p. e100, apr 2017. [Online]. Available: https://doi.org/10.2196%2Fjmir.6898

[16] J. A. Simpson, W. S. Rholes, and D. Phillips, “Conflict in close relationships: An attachment perspective.” *Journal of Personality and Social Psychology*, vol. 71, no. 5, pp. 899–914, 1996. [Online]. Available: https://doi.org/10.1037%2F0022-3514.71.5.899

[17] C. George and M. West, “The development and preliminary validation of a new measure of adult attachment: the Adult Attachment Projective,” *Attachment & Human Development*, vol. 3, no. 1, pp. 30–61, apr 2001. [Online]. Available: https://doi.org/10.1080%2F14616730010024771

[18] R. Miljkovitch, E. Moss, A. Bernier, K. Pascuzzo, and E. Sander, “Refining the assessment of internal working models: the Attachment Multiple Model Interview,” *Attachment & Human Development*, vol. 17, no. 5, pp. 492–521, sep 2015. [Online]. Available: https://doi.org/10.1080%2F14616734.2015.1075561

[19] M. Bertamini and M. R. Munafò, “Bite-Size Science and Its Undesired Side Effects,” *Perspectives on Psychological Science*, vol. 7, no. 1, pp. 67–71, jan 2012. [Online]. Available: https://doi.org/10.1177%2F1745691611429353

[20] M. Buhrmester, T. Kwang, and S. D. Gosling, “Amazon's Mechanical Turk,” *Perspectives on Psychological Science*, vol. 6, no. 1, pp. 3–5, jan 2011. [Online]. Available: https://doi.org/10.1177%2F1745691610393980

[21] M. Bertamini and M. R. Munafò, “Bite-Size Science and Its Undesired Side Effects,” *Perspectives on Psychological Science*, vol. 7, no. 1, pp. 67–71, jan 2012. [Online]. Available: https://doi.org/10.1177%2F1745691611429353

[22] M. Amrhein, B. Srinivasan, D. Bonvin, and M. Schumacher, “On the rank deficiency and rank augmentation of the spectral measurement matrix,” *Chemometrics and Intelligent Laboratory Systems*, vol. 33, no. 1, pp. 17–33, may 1996. [Online]. Available: https://doi.org/10.1016%2F0169-7439%2895%2900086-0

[23] D. M. Hawkins, “The Problem of Overfitting,” *Journal of Chemical Information and Computer Sciences*, vol. 44, no. 1, pp. 1–12, jan 2004. [Online]. Available: https://doi.org/10.1021%2Fci0342472

[24] E. Keogh and A. Mueen, “Curse of Dimensionality,” in *Encyclopedia of Machine Learning and Data Mining*. Springer US, 2017, pp. 314–315. [Online]. Available: https://doi.org/10.1007%2F978-1-4899-7687-1_192

[25] M. D. S. Ainsworth, M. C. Blehar, E. Waters, and S. N. Wall, *Patterns of attachment: A psychological study of the strange situation*. Psychology Press, 1978.

[26] J. Mesman, M. H. van IJzendoorn, and A. Sagi-Schwartz, “Cross-cultural patterns of attachment,” in *Handbook of Attachment, Third Edition: Theory, Research, and Clinical Applications*. Guilford New York, NY, 2016, pp. 852–877.

[27] A. L. Hill-Soderlund, W. R. Mills-Koonce, C. Propper, S. D. Calkins, D. A. Granger, G. A. Moore, J.-L. Gariepy, and M. J. Cox, “Parasympathetic and sympathetic responses to the strange situation in infants and mothers from avoidant and securely attached dyads,” *Developmental Psychobiology*, vol. 50, no. 4, pp. 361–376, 2008. [Online]. Available: https://doi.org/10.1002%2Fdev.20302

[28] H. A. DeVon, M. E. Block, P. Moyle-Wright, D. M. Ernst, S. J. Hayden, D. J. Lazzara, S. M. Savoy, and E. Kostas-Polston, “A Psychometric Toolbox for Testing Validity and Reliability,” *Journal of Nursing Scholarship*, vol. 39, no. 2, pp. 155–164, jun 2007. [Online]. Available: https://doi.org/10.1111%2Fj.1547-5069.2007.00161.x

[29] D. Bouthillier, N. Tremblay, F. Hamelin, D. Julien, and P. Scherzer, “Traduction et validation canadienne-française d'un questionnaire évaluant l'attachement chez l'adulte.” *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, vol. 28, no. 1, pp. 74–77, 1996. [Online]. Available: https://doi.org/10.1037%2F0008-400x.28.1.74

[30] C. Hazan and P. Shaver, “Romantic love conceptualized as an attachment process.” *Journal of Personality and Social Psychology*, vol. 52, no. 3, pp. 511–524, 1987. [Online]. Available: https://doi.org/10.1037%2F0022-3514.52.3.511

[31] D. W. Griffin and K. Bartholomew, “Models of the self and other: Fundamental dimensions underlying measures of adult attachment,” *Journal of Personality and Social Psychology*, vol. 67, no. 3, pp. 430–445, 1994. [Online]. Available: https://doi.org/10.1037%2F0022-3514.67.3.430

[32] H. Stein, A. D. Koontz, P. Fonagy, J. G. Allen, J. Fultz, J. R. Brethour, D. Allen, and R. B. Evans, “Adult attachment: What are the underlying dimensions?” *Psychology and Psychotherapy: Theory Research and Practice*, vol. 75, no. 1, pp. 77–91, mar 2002. [Online]. Available: https://doi.org/10.1348%2F147608302169562

[33] B. L. Hankin, J. D. Kassel, and J. R. Z. Abela, “Adult Attachment Dimensions and Specificity of Emotional Distress Symptoms: Prospective Investigations of Cognitive Risk and Interpersonal Stress Generation as Mediating Mechanisms,” *Personality and Social Psychology Bulletin*, vol. 31, no. 1, pp. 136–151, jan 2005. [Online]. Available: https://doi.org/10.1177%2F0146167204271324

[34] W. S. Rholes, J. A. Simpson, and M. Friedman, “Avoidant Attachment and the Experience of Parenting,” *Personality and Social Psychology Bulletin*, vol. 32, no. 3, pp. 275–285, mar 2006. [Online]. Available: https://doi.org/10.1177%2F0146167205280910

[35] H. Stein, A. D. Koontz, P. Fonagy, J. G. Allen, J. Fultz, J. R. Brethour, D. Allen, and R. B. Evans, “Adult attachment: What are the underlying dimensions?” *Psychology and Psychotherapy: Theory Research and Practice*, vol. 75, no. 1, pp. 77–91, mar 2002. [Online]. Available: https://doi.org/10.1348%2F147608302169562

[36] P. Ravitz, R. Maunder, J. Hunter, B. Sthankiya, and W. Lancee, “Adult attachment measures: A 25-year review,” *Journal of Psychosomatic Research*, vol. 69, no. 4, pp. 419–432, oct 2010. [Online]. Available: https://doi.org/10.1016%2Fj.jpsychores.2009.08.006

[37] D. Bouthillier, N. Tremblay, F. Hamelin, D. Julien, and P. Scherzer, “Traduction et validation canadienne-française d'un questionnaire évaluant l'attachement chez l'adulte.” *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, vol. 28, no. 1, pp. 74–77, 1996. [Online]. Available: https://doi.org/10.1037%2F0008-400x.28.1.74

[38] K. L. Gratz and L. Roemer, “Multidimensional Assessment of Emotion Regulation and Dysregulation: Development Factor Structure, and Initial Validation of the Difficulties in Emotion Regulation Scale,” *Journal of Psychopathology and Behavioral Assessment*, vol. 26, no. 1, pp. 41–54, mar 2004. [Online]. Available: https://doi.org/10.1023%2Fb%3Ajoba.0000007455.08539.94

[39] M. Rosenberg, *Society and the adolescent self-image*. Princeton university press, 2015.

[40] M. Johnston, J. Weinman, and S. C. Wright, *Measures in health psychology : a user's portfolio*. Windsor : NFER-NELSON, 1995.

[41] M. Dascalu, “ReaderBench (2) - Individual Assessment through Reading Strategies and Textual Complexity,” in *Analyzing Discourse and Text Complexity for Learning and Collaborating*. Springer International Publishing, 2014, pp. 161–188. [Online]. Available: https://doi.org/10.1007%2F978-3-319-03419-5_8

[42] V. A. Mashin, “Nonstationarity and duration of the cardiac interval time series in assessing the functional state of operator personnel,” *Biophysics*, vol. 52, no. 2, pp. 241–247, apr 2007. [Online]. Available: https://doi.org/10.1134%2Fs0006350907020170

[43] R. M. Baevsky, V. V. Bogomolov, I. I. Funtova, I. N. Slepchenkova, and A. G. Chernikova, “Prospects of medical monitoring of long-duration space flights by means of non-contact recording of physiological functions during sleep time,” *Human Physiology*, vol. 37, no. 7, pp. 816–820, dec 2011. [Online]. Available: https://doi.org/10.1134%2Fs036211971107005x

[44] T. Baltrusaitis, A. Zadeh, Y. C. Lim, and L.-P. Morency, “OpenFace 2.0: Facial Behavior Analysis Toolkit,” in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*. IEEE, may 2018. [Online]. Available: https://doi.org/10.1109%2Ffg.2018.00019

[45] A. Piolat, R. Booth, C. Chung, M. Davids, and J. Pennebaker, “La version française du dictionnaire pour le LIWC : modalités de construction et exemples d'utilisation,” *Psychologie Française*, vol. 56, no. 3, pp. 145–159, sep 2011. [Online]. Available: https://doi.org/10.1016%2Fj.psfr.2011.07.002

[46] Y. R. Tausczik and J. W. Pennebaker, “The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods,” *Journal of Language and Social Psychology*, vol. 29, no. 1, pp. 24–54, dec 2009. [Online]. Available: https://doi.org/10.1177%2F0261927x09351676

[47] G. Degottex, J. Kane, T. Drugman, T. Raitio, and S. Scherer, “COVAREP A collaborative voice analysis repository for speech technologies,” in *2014 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*. IEEE, may 2014. [Online]. Available: https://doi.org/10.1109%2Ficassp.2014.6853739

[48] Y. Sun and N. Thakor, “Photoplethysmography Revisited: From Contact to Noncontact From Point to Imaging,” *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 3, pp. 463–477, mar 2016. [Online]. Available: https://doi.org/10.1109%2Ftbme.2015.2476337

[49] D. J. McDuff, J. R. Estepp, A. M. Piasecki, and E. B. Blackford, “A survey of remote optical photoplethysmographic imaging methods,” in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, aug 2015. [Online]. Available: https://doi.org/10.1109%2Fembc.2015.7319857

[50] Y. Benezeth, P. Li, R. Macwan, K. Nakamura, R. Gomez, and F. Yang, “Remote heart rate variability for emotional state monitoring,” in *2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*. IEEE, mar 2018. [Online]. Available: https://doi.org/10.1109%2Fbhi.2018.8333392

[51] R. Macwan, Y. Benezeth, and A. Mansouri, “Remote photoplethysmography with constrained ICA using periodicity and chrominance constraints,” *BioMedical Engineering OnLine*, vol. 17, no. 1, feb 2018. [Online]. Available: https://doi.org/10.1186%2Fs12938-018-0450-3

[52] P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features,” in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*. IEEE Comput. Soc, 2001. [Online]. Available: https://doi.org/10.1109%2Fcvpr.2001.990517

[53] C. O. Conaire, N. E. O'Connor, and A. F. Smeaton, “Detector adaptation by maximising agreement between independent data sources,” in *2007 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, jun 2007. [Online]. Available: https://doi.org/10.1109%2Fcvpr.2007.383448

[54] W. Wang, A. C. den Brinker, S. Stuijk, and G. de Haan, “Algorithmic Principles of Remote PPG,” *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 7, pp. 1479–1491, jul 2017. [Online]. Available: https://doi.org/10.1109%2Ftbme.2016.2609282

[55] G. de Haan and V. Jeanne, “Robust Pulse Rate From Chrominance-Based rPPG,” *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 10, pp. 2878–2886, oct 2013. [Online]. Available: https://doi.org/10.1109%2Ftbme.2013.2266196

[56] S. Poria, E. Cambria, R. Bajpai, and A. Hussain, “A review of affective computing: From unimodal analysis to multimodal fusion,” *Information Fusion*, vol. 37, pp. 98–125, sep 2017. [Online]. Available: https://doi.org/10.1016%2Fj.inffus.2017.02.003

[57] A. Holzinger, “Interactive machine learning for health informatics: when do we need the human-in-the-loop?” *Brain Informatics*, vol. 3, no. 2, pp. 119–131, mar 2016. [Online]. Available: https://doi.org/10.1007%2Fs40708-016-0042-6

[58] E. Keogh and A. Mueen, “Curse of Dimensionality,” in *Encyclopedia of Machine Learning and Data Mining*. Springer US, 2017, pp. 314–315. [Online]. Available: https://doi.org/10.1007%2F978-1-4899-7687-1_192

[59] G. I. Webb, C. Sammut, C. Perlich, T. Horváth, S. Wrobel, K. B. Korb, W. S. Noble, C. Leslie, M. G. Lagoudakis, N. Quadrianto, W. L. Buntine, N. Quadrianto, W. L. Buntine, L. Getoor, G. Namata, L. Getoor, X. J. J. Han, J.-A. Ting, S. Vijayakumar, S. Schaal, and L. D. Raedt, “Linear Regression,” in *Encyclopedia of Machine*

Learning. Springer US, 2011, pp. 603–606. [Online]. Available: https://doi.org/10.1007%2F978-0-387-30164-8_476

[60] A. C. Kakas, D. Cohn, S. Dasgupta, A. G. Barto, G. A. Carpenter, S. Grossberg, G. I. Webb, M. Dorigo, M. Birattari, H. Toivonen, J. Timmis, J. Branke, H. Toivonen, A. L. Strehl, C. Drummond, A. Coates, P. Abbeel, A. Y. Ng, F. Zheng, G. I. Webb, and P. Tadepalli, "Artificial Neural Networks," in *Encyclopedia of Machine Learning*. Springer US, 2011, pp. 44–44. [Online]. Available: https://doi.org/10.1007%2F978-0-387-30164-8_35

[61] V. Torczon, "On the Convergence of Pattern Search Algorithms," *SIAM Journal on Optimization*, vol. 7, no. 1, pp. 1–25, feb 1997. [Online]. Available: https://doi.org/10.1137%2Fs1052623493250780

[62] A. R. Conn, N. Gould, and P. L. Toint, "A globally convergent Lagrangian barrier algorithm for optimization with general inequality constraints and simple bounds," *Mathematics of Computation*, vol. 66, no. 217, pp. 261–289, jan 1997. [Online]. Available: https://doi.org/10.1090%2Fs0025-5718-97-00777-1

[63] X. Zhang, "Regularization," in *Encyclopedia of Machine Learning and Data Mining*. Springer US, 2017, pp. 1083–1088. [Online]. Available: https://doi.org/10.1007%2F978-1-4899-7687-1_718

[64] D. J. C. MacKay, "Bayesian Interpolation," *Neural Computation*, vol. 4, no. 3, pp. 415–447, may 1992. [Online]. Available: https://doi.org/10.1162%2Fneco.1992.4.3.415

[65] F. D. Foresee and M. Hagan, "Gauss-Newton approximation to Bayesian learning," in *Proceedings of International Conference on Neural Networks (ICNN'97)*. IEEE, 1997. [Online]. Available: https://doi.org/10.1109%2Ficnn.1997.614194

[66] K. Levenberg, "A method for the solution of certain non-linear problems in least squares," *Quarterly of Applied Mathematics*, vol. 2, no. 2, pp. 164–168, jul 1944. [Online]. Available: https://doi.org/10.1090%2Fqam%2F10666

[67] E. Waters, C. E. Hamilton, and N. S. Weinfield, "The Stability of Attachment Security from Infancy to Adolescence and Early Adulthood: General Introduction," *Child Development*, vol. 71, no. 3, pp. 678–683, may 2000. [Online]. Available: https://doi.org/10.1111%2F1467-8624.00175

[68] R. Leiper and P. Casares, "An investigation of the attachment organization of clinical psychologists and its relationship to clinical practice," *British Journal of Medical Psychology*, vol. 73, no. 4, pp. 449–464, dec 2000. [Online]. Available: https://doi.org/10.1348%2F000711200160651

[69] D. F. Shanno, "Conditioning of quasi-Newton methods for function minimization," *Mathematics of Computation*, vol. 24, no. 111, pp. 647–647, sep 1970. [Online]. Available: https://doi.org/10.1090%2Fs0025-5718-1970-0274029-x

[70] J. Austin, *How To Do Things With Words*. Oxford University Press, sep 1975. [Online]. Available: https://doi.org/10.1093%2Facprof%3Aoso%2F9780198245537.001.0001

[71] J. Dore, "Holophrases speech acts and language universals," *Journal of Child Language*, vol. 2, no. 01, apr 1975. [Online]. Available: https://doi.org/10.1017%2Fs0305000900000878

[72] A. M. Groh, G. I. Roisman, C. Booth-LaForce, R. C. Fraley, M. T. Owen, M. J. Cox, and M. R. Burchinal, "IV. STABILITY OF ATTACHMENT SECURITY FROM INFANCY TO LATE ADOLESCENCE," *Monographs of the Society for Research in Child Development*, vol. 79, no. 3, pp. 51–66, aug 2014. [Online]. Available: https://doi.org/10.1111%2Fmono.12113

[73] S. Poria, E. Cambria, R. Bajpai, and A. Hussain, "A review of affective computing: From unimodal analysis to multimodal fusion," *Information Fusion*, vol. 37, pp. 98–125, sep 2017. [Online]. Available: https://doi.org/10.1016%2Fj.inffus.2017.02.003

[74] C. George, N. Kaplan, and M. Main, "Adult Attachment Interview," Ph.D. dissertation, University of California at Berkeley, 1985.

[75] S. I. Powers, P. R. Pietromonaco, M. Gunlicks, and A. Sayer, "Dating couples' attachment styles and patterns of cortisol reactivity and recovery in response to a relationship conflict." *Journal of Personality and Social Psychology*, vol. 90, no. 4, pp. 613–628, 2006. [Online]. Available: https://doi.org/10.1037%2F0022-3514.90.4.613

[76] R. A. Charter, "Study Samples Are Too Small to Produce Sufficiently Precise Reliability Coefficients," *The Journal of General Psychology*, vol. 130, no. 2, pp. 117–129, apr 2003. [Online]. Available: https://doi.org/10.1080%2F00221300309601280

[77] D. Hermes, "The Art of Xamarin App Development," in *Xamarin Mobile Application Development*. Apress, 2015, pp. 387–388. [Online]. Available: https://doi.org/10.1007%2F978-1-4842-0214-2_10

5.3 Psychometric properties of the Biometric Attachment Test: French, American and Tunisian validation of the first adult attachment test powered by artificial intelligence

Our last article to date, it is currently in review at the European Journal of Psychological Assessment (impact factor 2.22). This paper represents a more classic psychometric validation study for the BAT, testing its convergent and cross-cultural validity as well as its test-retest reliability on samples from three different countries: Tunisia, France, and the United States.

Psychometric properties of the Biometric Attachment Test: French, American and Tunisian validation of the first adult attachment test powered by artificial intelligence

Federico Parra[1,2,3], Gwenaëlle Persiaux[4], Nahed Boukadida[5], Carole Marchand[1], Eva Hanras[1], Clara Falala-Séchet[1], Yannick Benezeth[2], Khalid Kalalou[6], and Susana Tereno[1]

[1]LPPS (EA 4057) Institut de Psychologie, Université Paris Descartes, Université de Paris, France
[2]Université Bourgogne Franche-Comté, LE2I EA 7508, Dijon, France
[3]Centre de Psychothérapie et de Psychotrauma de Ville-Evrard, Saint-Denis, France
[4]University Hospital Center Saint Etienne, France
[5]Université de Tunis El Manar, Institut Supérieur des Sciences Humaines de Tunis, Département de Psychologie
[6]Unité de Recherche Clinique de Ville-Evrard, Neuilly-Sur-Marne, France

September 2, 2019

Abstract

During the last few years we have developed the Biometric Attachment Test (BAT), the first psychometric tool to evaluate adult attachment powered by artificial intelligence. The BAT is a 9-minutes long test based on an automatic analysis of linguistic, behavioral and physiological responses to standardized musical and picture stimuli. In this paper, we examine the BAT's convergence level with the Relationship Scales Questionnaire (RSQ), its test-retest reliability, and its cross-cultural validity, using samples from three different countries/continents: France (N=35), Tunisia (N=33), and the United States (N=35). Results show good evidence of convergence validity with the RSQ for the three samples, taken separately or collectively, serving also as proof of cross-cultural validity. Good test-retest reliability was also found. These results are encouraging for the use of this innovative tool in the attachment field.

Keywords: attachment, psychometrics, artificial intelligence, relationships, RSQ, test

Introduction

Adult attachment is an essential construct in psychological development. Attachment security has an impact on individuals' capacity to build healthy intimate relationships (Feeney, 2016), their ability for effective emotion regulation (Mikulincer and Shaver, 2019), and their risk of developing psychopathology (Mikulincer and Shaver, 2012).

In 2017, we introduced the Biometric Attachment Test (BAT; (Parra et al., 2017)), a 9-minutes long psychometric test designed to assess adult attachment. It is based on an automated, artificial-intelligence-based analysis of the participants' linguistic, behavioral and physiological responses to standardized picture and music stimuli sets. In 2019, we described in length the algorithms powering a second version of the BAT

(BAT v.2; (Parra et al., 2019)). These algorithms allow for generating different "scoring models" for the BAT based on preexisting validated attachment measures. We also presented an algorithm capable of accurately detecting heart rate from video (RPPG). In that same study, we presented cross-cultural and convergent validity evidence for the BAT's first scoring model, which was based on the AAQ, using a French and an American sample and showing high levels of convergence.

In the present study, we intend to examine the convergent validity of a new scoring model for the BAT based on the Relationship Scales Questionnaire (RSQ; (Griffin and Bartholomew, 1994)). As we'll see in more detail later, the RSQ has the advantage of measuring current attachment in general, whereas the AAQ measures attachment to romantic partners only. Convergence will be evaluated on three samples from different countries/continents: France, Tunisia, and the United States, analyzed separately and in a composite sample, providing an additional assessment of the cross-cultural validity of the BAT. We'll finally examine the BAT's RSQ model test-retest reliability, crucial for evaluating whether it is apt for use in outcome studies where attachment might need to be measured at different points in time.

Method

Participants

French American scoring model training sample

This sample consisted of 171 participants (109 females, 62 males) recruited between January and July 2019. The French subsample (n=137, 89 females, 48 males) was formed from multiple sources in different regions of France: 10 clinical patients recruited at University Hospital Center Sainte-Étienne and 19 recruited at the Ville-Evrard Center of Psychotherapy and Psychotrauma in Saint-Denis; 42 volunteers enrolled in Paris, 10 in Lyon; 3 college students enrolled at Paris Descartes University and 7 at University Bourgogne Franche-Comté (Dijon); 31 clinical, private practice patients were enrolled in Paris and 5 in Lyon. Age for this subsample was (minimum 18, maximum 72, mean 39.5, SD 13.1). The American subsample (n=34, 20 females, 14 males) was recruited online using Amazon Mechanical Turk and Prolific services. Age for this subsample was (minimum 18, maximum 68, mean 33.3, SD 10.9). Clinical patients were included because on average they present higher levels of both anxiety and avoidance, a configuration that is rarer in non-clinical populations. This way the BAT algorithms can get trained to recognize all possible combinations of anxiety and avoidance.

French sample

This sample consisted of 35 French participants (25 females, 10 males) recruited between January and July 2019. The sample was formed from multiple sources in different regions of France: 3 clinical patients recruited at the Ville-Evrard Center of Psychotherapy and Psychotrauma in Saint-Denis; 7 volunteers enrolled in Paris, 9 in Lyon; 4 college students enrolled at University Bourgogne Franche-Comté (Dijon); 12 clinical, private practice patients were enrolled in Paris. Age was (minimum 19, maximum 69, mean 39, SD 14.1). Clinical patients were included to examine whether the BAT was capable of rightly assessing more extreme attachment insecurity cases.

American sample

This sample consisted of 35 Americans (19 females, 16 males) recruited over the summer of 2019 online using Amazon Mechanical Turk and Prolific services. The mean age was 36.8 (minimum 18, maximum 68, SD 14.1). We did not recruit any clinical participants for this sample.

Tunisian sample

This sample consisted of 33 Tunisian participants (21 females, 12 males) recruited in July 2019 in the city of Tunis. Age was (minimum 17, maximum 55, mean 37.6, SD 10.5). We did not recruit any clinical participants for this sample.

Measures

Biometric Attachment Test (BAT) v2 - RSQ model

The Biometric Attachment Test (BAT) is a psychometric test developed by Federico Parra designed to measure attachment in adults (Parra et al., 2017, 2019). During the test, 14 themes related to attachment are evoked, such as separation from a partner, loss of a loved one, or tender romantic love. To evoke these themes, several pictures and music clips have been selected using a standardized process (see (Parra et al., 2017)). Each time the test is taken, a different, randomized combination of the BAT's pictures and music clips is used by the program in order to avoid priming effects on participants taking the test several times. Participants' responses to different BAT stimuli sets do not differ significantly (Parra et al., 2017). Algorithms analyze the participants' facial expressions, head movements, gaze direction, language, voice characteristics and heart rate as they are taking the test, using only the computer or tablet's camera and microphone to do so, and without saving any identifiable data in the process. Once the BAT obtains the raw results from these analyses, it interprets this data according to a scoring model and outputs a value in the 0-100 range.

In order to create a scoring model for the BAT based on a given preexisting validated measure, a sample of participants (estimated minimum n=150) has to be evaluated with both the BAT and the target measure. Then, the BAT's learning algorithm can automatically create a new scoring model containing the best possible combination of BAT's raw results (e.g. facial expressions results, heart rate results) to produce scores that are highly correlated (i.e. convergent) with the target measure. Scoring models can be created based on validated questionnaires, interview-based tests, or any other measure, whether its scores are continuous or discrete/categorical.

In 2017 we provided the first content validity evidence for the BAT showing that its raw output, such as voice tone or language results, were correlated in a theoretically-meaningful way to attachment security as measured by three validated attachment measures: the Adult Attachment Projective (George and West, 2001), the Adult Multiple Model Interview (Miljkovitch et al., 2015), and the Adult Attachment Questionnaire (Simpson et al., 1996).

Since different BAT scoring models can be created, it is expected for each to have different psychometric properties. We preliminarily examined the convergence and cross-cultural validity of the BAT's AAQ model finding a convergence of .74 ($r(19) = .74$, $p < 0.001$) with AAQ's avoidance and .79 ($r(19) = .79$, $p < 0.001$) with AAQ's anxiety dimensions (Parra et al., 2019).

In the present study, we've used the BAT's RSQ model, the second we have created. It is based on Griffin and Bartholomew's RSQ questionnaire described below (Griffin and Bartholomew, 1994).

The methods implemented to automatically analyze video and audio, extracting facial expressions, gaze direction, heart rate, voice tone and language, as well as the techniques used in the scoring model creation process, are highly technical subjects of computer science that exceed the scope of the present paper. We invite the interested reader to consult our 2019 paper detailing and evaluating these procedures (Parra et al., 2019).

The BAT's RSQ model was "trained" on a mixed French and American sample of 171 participants described in the Participants section above.

Relationship Scales Questionnaire (RSQ)

Griffin and Bartholomew's 30-item Relationship Scales Questionnaire (RSQ) was originally designed to measure dimensions related to positive or negative models of self and others (Griffin and Bartholomew, 1994). Its items are derived from Hazan and Shaver's three-category attachment prototype descriptors, Bartholomew and Horowitz's four-category attachment prototype descriptors (Bartholomew and Horowitz, 1991), and Collins and Read's attachment measure. Items are scored on a 5-point Likert scale. The RSQ produces scores for four attachment patterns and two attachment dimensions. Roisman and his colleagues performed confirmatory factor analyses and found that using the dimensions of anxiety (α=.84) and avoidance (α=.86) was the best-fitting model (Roisman et al., 2007), so we'll be using these dimensions in this paper. Regarding the attachment patterns scores, Backstrom and Holmes found low reliability for both the secure (α=.32) and the preoccupied (α=.46) patterns (äckström2001**?**), whereas Griffin and Bartholomew's research found fearful (.79) and dismissing (.64) patterns scores to have higher Cronbach's α values. They also found that the reliability of "model of others" was acceptable (α=-.68) whereas that of "model of self" was low (α=.50). In a study by Fortuna and colleagues, high levels of attachment anxiety and avoidance on the RSQ were shown to predict psychopathology under conditions of high and low life stress. In comparison, the gold-standard Adult Attachment Interview (AAI, (Carol George, 1985)) only did so under conditions of high stress (Fortuna and Roisman, 2008). There is a high correlation between Bartholomew and Horowitz's Relationship Questionnaire (RQ; (Bartholomew and Horowitz, 1991)) and the RSQ. The RSQ is further correlated to NEO Personality Inventory personality factors without being identical to them (Griffin and Bartholomew, 1994).

In France and Tunisia, we used the French translation of the RSQ produced by Guédeney and colleagues (Guédeney et al., 2010). Although the translation was validated for a French population in France, it has already been used with francophone populations in Tunisia as well (e.g. (Menif et al., 2016)).

RSQ scores were scaled to the 0-100 range to simplify its comparison with the BAT.

Procedures

Each participant was assessed in a room all alone using a computer with the BAT application installed. Each theme (e.g. separation from a partner) was evoked by a stimulus (picture, music, or both) for 15 seconds. The program then invited the participant to say aloud what they felt, giving them 20 seconds to do so. A 5-seconds pause followed before the next theme was evoked by the next stimulus. The duration of the full test was of 9 minutes and 20 seconds. The RSQ was answered on the computer after the BAT administration.

It is worth noticing here that given the diversity of sources for our samples, testing conditions were somewhat diverse. The BAT software itself provides guidelines for the participant to establish a relatively uniform testing situation (e.g. asking the participant to take the test in a room alone; asking them to answer using their voice), and includes several tests to ensure those guidelines are followed (e.g. face detection to control if there is another person nearby; microphone check to verify the voice is being used; testing for good enough lighting conditions using the webcam). But uniformity is still limited: lighting conditions still may vary, computers differ greatly from one another, microphones and webcams are of very different quality, and background sound is uncontrollable. We make this remark here to emphasize the challenge we took on: our intention was not to conduct a validation study in controlled laboratory conditions, for the BAT is intended as a tool to be run in all sorts of circumstances, with all sorts of different hardware configurations. Instead, we attempted to validate the BAT in a real-life context, with its inherent variance and unpredictability.

For the recruitment of our samples, a declaration with the "Commission Nationale de l'Informatique et des Libertés" (CNIL) was filled (filling number 2159625 v 0) in France. Additionally, permission was obtained from the Paris Descartes Institutional Review Board (IRB) under the identifier 00012019-04, and from the University of Southern California Institutional Review Board (IRB) under the identifier UP-16-00447. All participants electronically signed informed consent forms.

About our sample size: the rule of thumb for the validation of psychological measures according to Boateng and colleagues is to recruit 10 participants per measure item (Boateng et al., 2018). Counting the BAT's 13 themes (theme 1 is a neutral warm-up) as items, our target sample size was set to 130 participants. We were only able to recruit 103 participants, 27 short of our target. This was due to budget and time limitations, as well as specific difficulties associated with recruiting in three different continents that use two different languages (e.g. lack of sufficient access to powerful-enough computer equipment in Tunisia). It seems relevant to mention that this study was entirely self-funded.

Analyses

The following analyses were performed using SPSS 23, except for the Concordance Correlation Coefficient, calculated using Matlab, version R2018A. All analyses include 95% confidence intervals calculated using 1000 bootstrap samples.

Convergent validity analysis

Convergent validity is the extent to which a measure produces results that are similar to other validated measures measuring the same construct (Boateng et al., 2018). A standard way of measuring it is by using Pearson's product-moment correlation (Swank and Mullen, 2017). We'll interpret Pearson's results based on Drummond, Sheperis, and Jones's review of the best practices for interpreting validity coefficients, for whom a correlation greater than .5 indicates a very high correlation, .40 to .49 a high correlation, .21 to .40 a moderate correlation and .2 an unacceptable one (Drummond, 2016). In addition to analyzing convergent validity for each country sample separately, we created a composite of the three countries samples, of size N=103, to analyze them collectively.

Test-retest reliability analysis

This reliability test examines whether a person taking an assessment twice within a short interval will obtain the same result at both moments. It represents the test scores' stability across time (Boateng et al., 2018). A standard way of measuring test-retest reliability is by using Pearson's product-moment correlation (Boateng et al., 2018), as we'll do in our analysis. Several authors, however, have criticized this approach as incomplete, arguing that Pearson r can prove reliability but not agreement (Berchtold, 2016). We'll therefore also include a measure of Concordance Correlation Coefficient into our analysis, which is capable of accurately measuring agreement (Berchtold, 2016).

This analysis required participants to take the BAT test twice at a time interval. Unfortunately, most participants were not able to do so. Our analysis will therefore be performed on a subset of the French (N=35) and American (N=35) samples that were able to take the BAT a second time. The subsample size was of N=30.

BAT "resilience" mechanism to deal with errors explain variance in sample sizes reporting

During the automatic BAT scoring process, errors in the data, such as impossible heart rate values or times at which the participant did not speak or looked away for a long period of time, are detected and dealt with using a "resilience" mechanism that falls back on other available data features to compensate for the lacking ones. For example, the BAT can compensate the lack of accurate heart rate for 20 seconds using the facial expressions, language used, gaze direction, head pose and voice tone used during the same period.

This mechanism is limited to errors in one feature at a time, and therefore on rare occasions were errors are present in several features simultaneously and persistently the model can't produce a score at all.

BAT scoring models are different for each target variable they are meant to score. The model for RSQ anxiety differs from the model for RSQ avoidance, they use different combination of BAT raw feature scores (e.g. facial expressions scores) to produce the desired outcome values. This explains why each BAT dimension reported in the following Results section can slighly vary in terms of sample size.

All comparisons with RSQ scores are made pairwise.

Results

Sociodemographic data

French American scoring model training sample (N=171): occupational status was n=15(9%) unemployed, n=117(68%) employed, n=35(20%) students, n=4(2%) retired; relationship status was n=24(14%) in a relationship, n=46(27%) married, n=19(11%) separated or divorced, n=75(44%) single, n=7(4%) widow; and n=54(32%) were clinical patients.

French sample (N=35)*:* occupational status was n=2 (6%) unemployed, n=27 (77%) employed, n=5 (14%) students, n=1 (3%) retired; relationship status was n=7 (20%) in a relationship, n=16 (46%) married, n=1 (3%) separated or divorced, n=9 (26%) single; and n=9 (26%) were clinical patients.

American sample (N=35)*:* occupational status was n=6 (17%) unemployed, n=19 (54%) employed, n=7 (20%) students, n=3 (9%) retired; relationship status was n=5 (14%) in a relationship, n=12 (34%) married, n=4 (11%) separated or divorced, n=12 (34%) single, n=2(6%) widow. We did not recruit any clinical participants for this sample.

Tunisian sample (N=33)*:* occupational status was n=29 (88%) employed, n=4 (12%) students; relationship status was n=11 (33%) married, n=6 (18%) separated or divorced, n=13 (39%) single, n=3(9%) widow. We did not recruit any clinical participants for this sample.

In Table 1 we can observe the distribution (mean and 95% confidence intervals) for each dimension of the RSQ versus those obtained using the BAT RSQ-model, in each of the samples and the composite sample.

Convergent validity results (Pearson r)

In order to assess the extent to which the BAT produces results that are similar to the RSQ, and that they measure the same construct, we have used Pearson's product-moment correlation.

Attachment anxiety dimension, as measured by BAT's RSQ model, was very highly correlated with RSQ anxiety in the French (r(35) = .52, p = .001, 95% CI [.32, .69]) and in the American (r(32) = .54, p = .002, 95% CI [.32, .72]) samples, and was highly correlated with RSQ anxiety in the Tunisian (r(31) = .47, p = .008, 95% CI [.08, .72]) and in the composite (r(98) = .45, p < .001, 95% CI [.32, .58]) samples.

Table 1: Means and 95% confidence intervals of RSQ and BAT results for the three samples and the composite sample.

		RSQ anxiety (0-100)			BAT anxiety (0-100)		
	N	Mean	Low 95% CI	High 95% CI	Mean	Low 95% CI	High 95% CI
French	35	49.57	42.14	57.14	49.33	45.86	52.65
American	32	41.88	32.04	51.87	53.34	49.15	57.92
Tunisian	31	40.65	33.71	48.22	54.89	50.45	59.91
Composite	98	44.23	39.19	48.88	52.40	49.96	54.74

		RSQ avoidance (0-100)			BAT avoidance (0-100)		
	N	Mean	Low 95% CI	High 95% CI	Mean	Low 95% CI	High 95% CI
French	35	41.79	35.45	48.66	44.66	41.61	47.97
American	34	43.93	36.58	51.65	47.11	44.40	49.74
Tunisian	32	43.26	37.79	49.22	43.96	40.47	47.53
Composite	101	42.98	39.33	46.72	45.26	43.49	47.08

Attachment avoidance dimension, as measured by BAT's RSQ model, was highly correlated with RSQ avoidance in the French ($r(35) = .48$, $p = .004$, 95% CI [0.18, 0.73]), American ($r(34) = .44$, $p = .01$, 95% CI [.18, .66]), Tunisian ($r(32) = .40$, $p = .025$, 95% CI [.01, .67]), and composite ($r(101) = .43$, $p < .001$, 95% CI [.26, .57]) samples.

Test-retest reliability results (Pearson r and Concordant Correlation Coefficient)

In order to examine the BAT scores' stability across time we have used Pearson's product-moment correlation and the Concordance Correlation Coefficient (CCC). The mean interval between the first (T1) and the second (T2) time participants took the BAT was 14.53 days (SD 17.71).

BAT attachment anxiety dimension scores at T1 and T2 were very highly correlated in our retest subsample ($r(28) = .74$, $p < .001$, 95% CI [.49, .88]), with a CCC of .73, 95% CI [.49, .86].

BAT attachment avoidance dimension scores at T1 and T2 were very highly correlated in our retest subsample ($r(29) = .79$, $p < .001$, 95% CI [.61, .90]), with a CCC of .78, 95% CI [.60, .89].

Discussion

The present study evaluated the BAT RSQ model's convergence validity, test-retest reliability, and crosscultural validity. We interpret our results as good evidence for convergence validity between the BAT's RSQ model and the RSQ questionnaire for both the anxiety and the avoidance dimensions in all of our samples, suggesting scores obtained using the BAT with its RSQ model are valid measures of adult attachment. Moreover, the fact that results were comparable for all three samples suggests the BAT RSQ model can be considered cross-culturally valid at least with regards to France, Tunisia, and the United States. This was further demonstrated by our analysis of the composite sample.

We also found evidence for test-retest reliability and agreement, suggesting that the BAT can be used in outcome studies with confidence.

To our knowledge, the BAT is the first test of its kind: a validated, exposure-based psychometric test that implements artificial intelligence in a standardized protocol in order to measure a psychological construct.

Its advantages over classical questionnaires and interview-based tests are manifold: the BAT takes only 9 minutes to complete, it can be taken at home with a computer or tablet and is resilient to unpredictable variability in the test conditions, it is scored automatically in minutes, it is objective and replicable in its observations, it is holistic, taking into account language, voluntary and involuntary behavior and physiology; it can be used in different cultures with only minimal translation efforts, and it can evolve over time learning new scoring models based on different validated psychometric measures. Our encouraging results suggest that the BAT RSQ model is ready for being implemented in research and clinical practice.

A limitation of our study is that, taken separately, the sample sizes of the three different countries as well as the subsample for test-retest reliability were arguably small. One could see a strength in this limitation, however, considering the low statistical power of small samples and yet how good the results were. In addition, we were able to replicate the individual samples' results in the larger composite sample (N=103). Another limitation is the fact that the RSQ was responded to immediately after taking the BAT, potentially affecting the participants' RSQ responses. This was a compromise made in order to simplify the procedure for our participants and thus maximize recruitment.

In 2020, a version of the BAT software will be made available for researchers interested in conducting their own validation studies. Collaborations to elaborate new scoring models for the BAT based on other validated attachment measures (e.g.; Adult Attachment Narratives, Adult Attachment Interview), to create new stimuli sets (i.e. pictures, music) better tailored to specific cultures, or to use the BAT in attachment research in general, will be welcomed.

Open Practices Statement

The experiment reported in this article was not formally preregistered. The data have been made available on the Hardvard Dataverse permanent archive: https://doi.org/10.7910/DVN/1CHHGB

The materials (the BAT software) will be made available in 2020 upon request that should be sent via email to the lead author federico.parra@hotmail.com.

References

Kim Bartholomew and Leonard M. Horowitz. Attachment styles among young adults: A test of a four-category model. *Journal of Personality and Social Psychology*, 61(2):226–244, 1991. doi: 10.1037/0022-3514.61.2.226. URL https://doi.org/10.1037%2F0022-3514.61.2.226.

André Berchtold. Test–retest: Agreement or reliability? *Methodological Innovations*, 9:205979911667287, apr 2016. doi: 10.1177/2059799116672875. URL https://doi.org/10.1177%2F2059799116672875.

Godfred O. Boateng, Torsten B. Neilands, Edward A. Frongillo, Hugo R. Melgar-Quiñonez, and Sera L. Young. Best Practices for Developing and Validating Scales for Health Social, and Behavioral Research: A Primer. *Frontiers in Public Health*, 6, jun 2018. doi: 10.3389/fpubh.2018.00149. URL https://doi.org/10.3389%2Ffpubh.2018.00149.

Mary Main Carol George, Nancy Kaplan. *Adult Attachment Interview*. PhD thesis, University of California at Berkeley, 1985.

Sheperis & Jones Drummond. *Assessment Procedures for Counselors and Helping Professionals, 8th Edition*. Pearson, 2016.

J. A. Feeney. Adult romantic attachment: developments in the study of couple relationships. In *Handbook of Attachment, Third Edition: Theory, Research, and Clinical Applications*. Guilford, 2016.

Keren Fortuna and Glenn I. Roisman. Insecurity stress, and symptoms of psychopathology: contrasting results from self-reports versus interviews of adult attachment. *Attachment & Human Development*, 10(1):11–28, mar 2008. doi: 10.1080/14616730701868571. URL https://doi.org/10.1080%2F14616730701868571.

Carol George and Malcolm West. The development and preliminary validation of a new measure of adult attachment: the Adult Attachment Projective. *Attachment & Human Development*, 3(1):30–61, apr 2001. doi: 10.1080/14616730010024771. URL https://doi.org/10.1080%2F14616730010024771.

Dale W. Griffin and Kim Bartholomew. Models of the self and other: Fundamental dimensions underlying measures of adult attachment. *Journal of Personality and Social Psychology*, 67(3):430–445, sep 1994. doi: 10.1037/0022-3514.67.3.430. URL https://doi.org/10.1037%2F0022-3514.67.3.430.

N. Guédeney, J. Fermanian, and A. Bifulco. La version française du Relationship Scales Questionnaire de Bartholomew (RSQ Questionnaire des échelles de relation) : étude de validation du construit. *L'Encéphale*, 36(1):69–76, feb 2010. doi: 10.1016/j.encep.2008.12.006. URL https://doi.org/10.1016%2Fj.encep.2008.12.006.

Leila Menif, Faten Medini, Sana Derouiche, and Wahid Melki. The Links between Attachment and Cyberaddiction. *Journal of Pharmacy and Pharmacology*, 4(8), aug 2016. doi: 10.17265/2328-2150/2016.08.006. URL https://doi.org/10.17265%2F2328-2150%2F2016.08.006.

Mario Mikulincer and Phillip R. Shaver. An attachment perspective on psychopathology. *World Psychiatry*, 11(1):11–15, feb 2012. doi: 10.1016/j.wpsyc.2012.01.003. URL https://doi.org/10.1016%2Fj.wpsyc.2012.01.003.

Mario Mikulincer and Phillip R Shaver. Attachment orientations and emotion regulation. *Current Opinion in Psychology*, 25:6–10, feb 2019. doi: 10.1016/j.copsyc.2018.02.006. URL https://doi.org/10.1016%2Fj.copsyc.2018.02.006.

Raphaële Miljkovitch, Ellen Moss, Annie Bernier, Katherine Pascuzzo, and Emmanuel Sander. Refining the assessment of internal working models: the Attachment Multiple Model Interview. *Attachment & Human Development*, 17(5):492–521, sep 2015. doi: 10.1080/14616734.2015.1075561. URL https://doi.org/10.1080%2F14616734.2015.1075561.

Federico Parra, R Miljkovitch, G Persiaux, M Morales, and S Scherer. The Multimodal Assessment of Adult Attachment Security: Developing the Biometric Attachment Test. *J Med Internet Res*, 19:e100, Apr 2017. doi: 10.2196/jmir.6898. URL https://www.jmir.org/2017/4/e100/.

Federico Parra, Stefan Scherer, Yannick Benezeth, Plamena Tsvetanova, and Susana Tereno. Development and cross-cultural evaluation of a scoring algorithm for the Biometric Attachment Test: Overcoming the challenges of multimodal fusion with small data. *IEEE Transactions on Affective Computing*, pages 1–1, 2019. doi: 10.1109/taffc.2019.2921311. URL https://doi.org/10.1109%2Ftaffc.2019.2921311.

Glenn I. Roisman, Ashley Holland, Keren Fortuna, R. Chris Fraley, Eric Clausell, and Alexis Clarke. The Adult Attachment Interview and self-reports of attachment style: An empirical rapprochement. *Journal of Personality and Social Psychology*, 92(4):678–697, apr 2007. doi: 10.1037/0022-3514.92.4.678. URL https://doi.org/10.1037%2F0022-3514.92.4.678.

Jeffry A. Simpson, W. Steven Rholes, and Dede Phillips. Conflict in close relationships: An attachment perspective. *Journal of Personality and Social Psychology*, 71(5):899–914, 1996. doi: 10.1037/0022-3514.71.5.899. URL https://doi.org/10.1037%2F0022-3514.71.5.899.

Jacqueline M. Swank and Patrick R. Mullen. Evaluating Evidence for Conceptually Related Constructs Using Bivariate Correlations. *Measurement and Evaluation in Counseling and Development*, 50(4):270–274, oct 2017. doi: 10.1080/07481756.2017.1339562. URL https://doi.org/10.1080%2F07481756.2017.1339562.

6 Discussion

This doctoral dissertation aimed to develop a new kind of adult attachment test, based on exposure to attachment-related image and music stimuli and powered in its analyses by artificial intelligence. Using this methodology, we attempted to overcome the limitations of existing self-report and interview-based measures of adult attachment.

Readers are aware by now of the discussions specific to the individual results of each of the three studies reproduced above. In the following discussion, we would like to focus rather on several insights that are inspired by the overall results. After that, we will spend some time reviewing some strengths of the present work, before discussing its limitations. We'll later share our intentions for this project moving forward, and finish up with a conclusion.

6.1 Insights from our studies

6.1.1 The flexibilities of the BAT

The studies we conducted helped reveal different ways in which the BAT is flexible compared to other psychometric tests.

The first such way is about the stimuli we use to trigger attachment-related reactions. We showed that, as long as stimuli are chosen following the standardized procedure and survey, stimuli can change without affecting the test's performance. In particular, in the first study, repeated measures ANOVAs were used to test differences in responses of the same participants to different stimuli sets. No significant differences were found for the majority of features. In the third study, BAT v3 was used, which automatically randomizes the stimuli set every time the test is taken (to see an example of stimuli representing each theme, see section 4.4.5) so that no BAT administration can ever be exactly the same. And yet, test-retest reliability was established for both anxiety and avoidance dimensions on an n=30 subsample, showing stability despite the different stimuli.

This flexibility implies that in the future, researchers or clinicians could decide to build new stimuli sets, perhaps to increase the empathic identification effect (see section 4.3.1), by selecting subjects more representative of a given population. This differentiates the BAT from all other exposure-based tests such as the AAP, which rely on a fixed set of stimuli.

Another flexibility was revealed in the second and third studies: the BAT three-step multimodal fusion procedure can help create scoring models that emulate other measures. For practicality sake, we used two questionnaires, the AAQ, and the RSQ, for these studies; but the same mechanism could be used to mimic the AAI or even the SSP (by training the BAT scoring model using a longitudinal sample, see section 6.2.4). As more scoring models are developed, the BAT becomes more useful: with a single 9-minutes-long test, a clinician or researcher could obtain a number of results (e.g., BAT SSP model result, BAT AAI model result, BAT RSQ model result). Depending on the problem being examined, users of the BAT could choose the score that better serves them, or combine some in a composite measure. Scores could be compared too. To obtain the same results without the BAT, one would require great time and budget investment to have a sample of participants take all the necessary tests.

A third flexibility we observed in the second and third article is the capacity of the BAT to accommodate to a new language easily. In principle, the BAT can accommodate any language for which a LIWC validated dictionary exists. As of today, this includes Spanish, English, French, Brazilian, and Portugal Portuguese, Dutch, German, Italian, Russian, simplified, and traditional Chinese. If a researcher wanted to make the BAT available in one of these languages, he or she would have to translate the BAT instructions, a set of simple phrases the likes of "What did you feel?" that, overall, amount to one page of text. Moreover, we are trying now a solution for languages for which a validated LIWC dictionary doesn't yet exist. The BAT will be used at IDC Herzliya center by Mario Mikulincer, where Israeli participants will be tested in their native tongue, Hebrew. Since there is no validated LIWC dictionary for Hebrew, we are inserting one more step in the linguistic analysis procedure: after Google Cloud Speech's automatic transcription, we implement Google Cloud Translation Services to translate the original Hebrew to English. Then, the translated text is analyzed as a regular English text. Empirical analyses of our results with the Hebrew sample will dictate if this is an appropriate line of work. If it is, this might mean that almost all world languages will be covered by the BAT, since Google Cloud Speech can detect a total of 120 languages and dialects, and Google Cloud Translation Services can translate 104 of them into English.

A fourth flexibility of the BAT is its capacity to interpret attachment both as a dimensional construct and as a categorical one. The three-step multimodal fusion procedure we developed can use in its last step either a regression neural-network or a classification neural network, allowing the test to emulate not only scores from other tests but also classifications from other tests. In the future, a person taking the BAT will get not only dimensional results, such as the level of attachment anxiety or avoidance, but also a classification, such as dismissing or preoccupied.

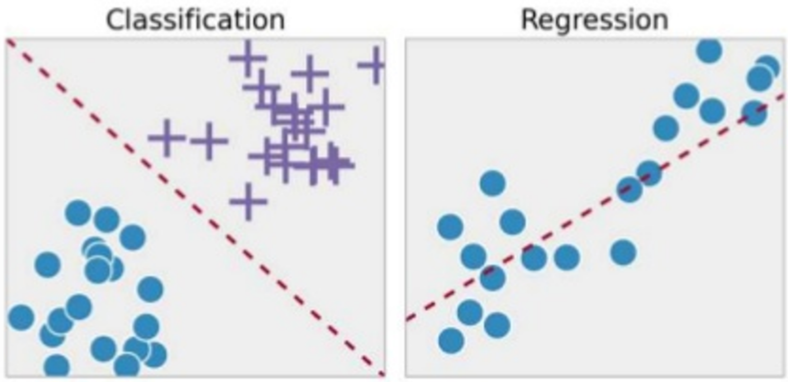


Figure 14: In the BAT's last fusion step, a classification or a regression can be used to emulate either attachment classifications or attachment continuous scores.

6.1.2 A cross-cultural perspective on our results: attachment as a universal adult phenomenon

As we began working on a new measure for adult attachment, our personal experience as immigrants that lived in three different countries, that speak three different languages, pushed us to reflect on the difficulties of translating and validating psychological measures, as well as potential ways to circumvent this hindrance for the BAT. As we've seen in our section on the flexibilities of the BAT (section 6.1.1), this goal was achieved. But what we had not foreseen were the ways in which our multi-cultural research could contribute to the understanding of adult attachment as a universal phenomenon.

Attachment as a construct has been widely studied in several different cultures, as shown by the fact that the theory was first investigated by Mary Ainsworth in Uganda (Ainsworth, Blehar, Waters, & Wall, 1978). Several studies show that attachment needs and their very early expressions are innate and thus arise cross-culturally (Mesman, van IJzendoorn, & Sagi-Schwartz, 2016). If we think of all the behavioral, physiological and cognitive reactions children have to an environment of insufficient care with regards to attachment, the earliest of said reactions are rather uninfluenced by culture. For example, children in Uganda and children in the United States both display very similar behaviors of avoidance during controlled experiments (Ainsworth, Blehar, Waters, & Wall, 1978), and the same seems to be true for psycho-physiological traits. These early developed patterns are carried into adulthood. As we age, however, other more complex attachment-related patterns develop, some of which are heavily influenced by the surrounding culture, with a salient case being language.

When we begun using the BAT with groups of people from different countries such as the United States, Portugal, France, and Tunisia, since the BAT measures both early attachment patterns (such as changes in sympathetic activation or spontaneous gaze behaviors) and more mature attachment patterns (such as linguistic characteristics of speech), we hypothesized two different sets of results. On the one hand, scores associated with sensory modalities that reveal innate or early developed behaviors, such as psychophysiological responses or facial expressions, should shield equivalent results across cultures, we thought, since these "low-level" characteristics are rather unaffected by culture. On the other hand, we thought that scores from linguistic analysis should not be equivalent across cultures, as we thought language was heavily influenced by culture. Since the BAT combines scores from all these sensory modalities in the assessment of adult attachment, we predicted that the BAT would work better (i.e., show higher convergent validity) in some cultures compared to others; in particular, when the BAT scoring method is trained using a sample from a particular culture, we thought it would work better in measuring attachment in said culture than in any other.

But we were wrong.

In our second study, we trained the BAT v2 scoring model using a sample of 59 French participants and yet, the BAT v2 was capable of scoring attachment in a sample of 19 American participants with even higher convergent validity metrics. And in our third study, we trained the BAT v3 using a French-American composite sample, yet the BAT v3 was perfectly capable of scoring a Tunisian sample, results which showed similar convergent validity metrics.

What does this say about the cross-cultural nature of adult attachment? About its universality?

Our linguistic features, such as the use of the present tense or the balance of positive vs. negative words used, were powerful predictors for attachment dimensions when used to score attachment in French (even Tunisian French) and English. This is remarkable, from an attachment theory standpoint, as it suggests there could be attachment-based dynamics underlying language that transcend the specificity of the language at hand. A good vantage point for the understanding of this phenomenon might be speech act theory, which posits a performative function of language and communication (Austin, 1975), seen as partially universal (Dore, 1975). In fact, a recently developed measure has used universal speech acts to measure attachment in adults (Talia, Miller-Bottome, & Daniel, 2015).

For us, our evidence further contributes to the view that Bowlby and Ainsworth were pointing to something much larger than childhood behaviors encoded in our genes: they were uncovering neuropsychophysiological, behavioral and linguistic patterns that transcend culture and affects all humans throughout our lifetimes.

6.1.3 Ethical dilemmas of using artificial intelligence in psychometric testing

"In November 2019, a Voight-Kampff test is being conducted in a room, on a man named Leon Kowalski. The test automatically measures bodily functions such as respiration, heart rate, blushing and eye movement in response to emotionally provocative questions. Halfway through the test, Mr. Kowalski gets violent and shoots Holden, the interviewer". Those lines, which probably caused confusion both for their temporal reference as well as for the resemblance to the BAT, are actually a description of a Blade Runner scene. The date, that we left unchanged, is striking.

Figure 15: Identification of "replicants" using the Voight-Kampff test (1981 science fiction film Blade Runner)

Aside from highlighting the brilliant way in which science fiction once and again predicts the future (Blade Runner was filmed in 1981), we bring up this scene because, for many, it conjures all the fears they might feel about a test such as the BAT.

In Blade Runner, the state has begun using such automatic tests to differentiate among kinds of people. Could the BAT, and similar technology, be used in a way that is not ethical?

From the moment we began thinking of using artificial intelligence to power an automated attachment test, we were confronted, both in our thoughts as well as in discussion with other fellow clinicians and researchers, with important and legitimate ethical questions. Now that the test is completed and validated, those questions deserve some answers.

One such question regards the feasibility of our methodology becoming generalized to psychological testing in general and, if so, the economic impact that such development could have on professionals that make a living out of psychometric testing. Psychiatrists, psychologists, and related professionals have for a century made part of their revenue out of psychometric testing: personality tests, intelligence tests, mental health screenings, and so on.

To answer the first part of that question, indeed we think the time is ripe for several psychological dimensions to be measured by equivalent methodologies to the one applied in the present work. Efforts are ongoing (see section 4.5.1) and might bear fruit in upcoming years, with tools made available to researchers and clinicians. With regards to the latter part of the question, our conclusion is that, if mental health is not a commodity but a human right, access to higher quality care - including higher-quality testing - should prevail, ethically speaking, over economic and political concerns such as revenue or joblessness. In addition, both the Hippocratic Oath (for psychiatrists) and the psychological deontological code (ethical principles of psychologists and code of conduct, in the US) imply the figure of a fiduciary relationship, by which the professional is bound to act in the interest of the client/patient. In cases were the financial interest of the professional (e.g., being able to make more income by performing tests manually) and the health interest of the patient (e.g., having access to a more objective test) collide, it is the client's interest that should be prioritized, according to the ethics of beneficence (Kinsinger, 2009). A different way to reflect about this dilemma is through a thought experiment: imagine the Theriak - an ancient mythical medicine that was supposed to cure all illnesses - recipe was rediscovered and validated as a real, effective drug. Shouldn't we produce it in mass? Should we care that, by doing so, we would be putting all doctors and related professions out of work?

Another ethical question we hear often is reminiscent of the Blade Runner scene: what if bad actors use the BAT, or similar technology, for negative things? Bad actors could include state or private actors, while negative things could include all sorts of manipulation, for example, inadvertently selecting social workers based on their attachment characteristics or having a dating site use technology akin to the BAT to unadvisedly filter out people with insecure attachment. If a test with similar technology measures a different construct, such as depression, possibilities for bad actors increase, for instance targetting depressed people with psychiatric drugs on facebook.

From an ethical standpoint, a serious risk-benefit assessment needs to be conducted to determine the scope of the threat as well as the predicted improvement in peoples' lives. While each of the steps we took while conducting this work has been scrutinized through risk-benefit assessments by expert ethical committees (University of Southern California Institutional Review Board under the identifier UP-16-00447, Paris Descartes Institutional Review Board under the identifier 00012019-04, "Commission Nationale de l'Informatique et des Libertés" under identifier 2159625 v 0), none of these committees was evaluating the potential harm the BAT could cause in the future, nor the potential harm that technology akin to the BAT could cause. It is clearly beyond our capabilities to attempt such an analysis, given the vast number of actors invested currently in developing similar technology with all sorts of different objectives, and given the fact that there is no overarching transnational control institution allowing for the regulation and supervision of them all. We can only offer our humble opinion, which is optimistic, by looking at the way in which mankind has got to regulate, in time, the worse applications of new technology (e.g, nuclear weapons), while simultaneously increasing the applications that promote the common good (e.g, nuclear medicine). For a review of the regulation of emergent technologies, see (Wiener, 2004)).

Another ethical question is about access. Should a software that can detect some form of mental health condition be accessible to the general public? In principle, no. That is the position of the American Psychological Association (APA; see https://www.apa.org/science/programs/testing/test-security) and the International Test Commission (ITC; see https://www.intestcom.org/page/17).

And yet, questionnaires for adult attachment, in particular, abound on the internet, often including automatic scoring and even therapeutic recommendations (e.g., Experiences in Close Relationships questionnaire, ECR https://openpsychometrics.org/tests/ECR.php, Relationships Attachment Style test https://www.psychologytoday.com/intl/tests/relationships/relationship-attachment-style-test). Note the last one of those is made available by a mainstream American psychology magazine.

Similar self-report measures with automated coding exist online for depression, PTSD, and most psychiatric conditions. In this reality, which is very distant from APA and ITC recommendations, would a better, more objective test do harm or good? For the time being, our conclusion is that it is better to stick to the guidelines of keeping psychometric tools strictly for the use of mental health professionals, but we are aware of the complexities of the issue and think that a revision of guidelines in light of new developments in the psychometric field might be of the essence.

6.1.4 Insecure attachment as a vulnerability factor in adulthood

During our introduction, we explored the way in which attachment insecurity, both longitudinally as well as cross-sectionally, is associated with deficits in psychological functioning. For example, we saw how adults with attachment insecurity of any kind tend to report more interpersonal problems (Bartholomew & Horowitz, 1991), loneliness, social isolation, lower relationship satisfaction, more frequent relationship breakups, greater physiological stress reaction to interpersonal conflict, and more frequent conflicts and violence in their relationships (Mikulincer & Shaver, 2012; Larose & Bernier, 2001; Powers, Pietromonaco, Gunlicks, & Sayer, 2006; Wallace & Vaux, 1993).

During our three studies, we included in our recruitments questionnaires to measure emotion regulation (Difficulties in Emotion Regulation Scale or DERS; (Gratz & Roemer, 2004)), self-esteem (Rosenberg Self-Esteem Scale or RSES; (Rosenberg, 1965)), self-efficacy (General Self-Efficacy Scale or GSE; (Johnston, Weinman, & Wright, 1995)), and childhood trauma (Adverse Childhood Experience or ACE; (Felitti et al., 2019)). See table 3 for a detail of the questionnaires used in each study.

Originally we meant to use this data after the dissertation, in further developments of the BAT technology (see section 6.4.4). But as our sample grew larger and diverse, it became self-evident that missing the opportunity to comment on exploratory findings regarding these questionnaires would be unforgivable.

By exploring correlations between the AAQ and the RSQ attachment questionnaires, on the one hand, and the DERS, RSES, GSE and ACE questionnaires, on the other, we found the following results:

Using our first and second study samples, which used the AAQ as attachment questionnaire (N=76, 57 French and 19 American participants), emotion dysregulation as measured by DERS was correlated with AAQ's attachment anxiety ($r(76) = .55$, $p < .001$) and avoidance ($r(76) = .43$, $p < .001$). Self-esteem as measured by RSES was negatively correlated with AAQ's attachment anxiety ($r(76) = -.42$, $p < .001$) and avoidance (r(76) = -.44, p < .001). Self-efficacy as measured by GSE was negatively correlated with both AAQ's attachment anxiety (r(76) = -.46, p < .001) and avoidance (r(76) = -.42, p < .001).

Using our third study samples, that included participants from France, the United States and Tunisia and that used the RSQ as the attachment measure, for the 252 participants for which RSQ and DERS measures were obtained, a correlation was found between DERS and RSQ's attachment anxiety (r(252) = .60, p < .001) and avoidance (r(252) = .45, p < .001). For the 253 participants for which RSQ and ACE measures were obtained, a correlation was found between ACE and RSQ's attachment anxiety (r(253) = .21, p < .001) and avoidance (r(253) = .28, p < .001).

If we use the BAT results instead of the results of the questionnaires, these exploratory findings are replicated. For example, using our third study sample, we found that emotion dysregulation as measured by DERS questionnaire is correlated with BAT attachment anxiety (r(274) = .42, p < .001) and avoidance (r(274) = .39, p < .001).

These results speak for themselves. Both attachment anxiety and attachment avoidance in adulthood, measured with either measure in all these different countries, are highly associated with deficits in emotion regulation, self-esteem, and self-efficacy. In addition, attachment anxiety and avoidance in adulthood is moderately correlated with retrospectively self-reported childhood trauma.

This should come at no surprise, as it only replicates results from the literature that we already commented in the introduction (section 4.1, for example (Bergman, Sarkar, Glover, & O'Connor, 2010; de Wolff & van IJzendoorn, 1997; Sroufe, Coffino, & Carlson, 2010)).

Deficits in these psychological dimensions augment psychological suffering and difficulty to cope with life experiences. For example, a deficit in emotion regulation can produce road rage, office rage, and even air rage, as well as spousal, child, and elder abuse (Gross, 2002). Deficits in self-esteem can increase the risk for poor mental and physical health and lower economic prospects, whereas self-efficacy deficits can produce feelings of powerlessness, incompetence, and victimization that directly affect mental health (Trzesniewski et al., 2006; Gecas, 1989).

What's interesting for us is not only that these psychological dimensions are associated with attachment insecurity in our data as well as in many other studies. What's salient for us is that increasing attachment security tends to have the effect of improving these dimensions. Specifically, using Brown and Elliott's Ideal Parent Figure method to treat attachment insecurity of any kind, tends to improve attachment security but also emotion regulation and self-esteem, in addition to the quality of life and an overall improvement in psychological symptoms (see (Brown & Elliott, 2016; Parra, George, Kalalou, & Januel, 2017)).

This sort of relationship, whereby experimentally increasing one variable has the effect of decreasing the other, and whereby the amount of improvement or deterioration seems also linked, is called a dose-effect relationship and has been used as proof of causation in the literature. For instance, Read et al. found that child abuse is a causal factor for adult hallucinations because a dose-effect relationship was found (Read, Os, Morrison, & Ross, 2005). Of course, the notion that attachment insecurity could cause these deficits remains only a hypothesis for now, that requires further research for verification.

What's sure is that these exploratory findings add to the importance of considering attachment insecurity in adult clinical settings, and that in that context, the BAT could contribute by allowing a simpler, cheaper, more objective measure of attachment in adults that can help identify attachment insecurity early on, pointing towards proven methods of treatment when required.

Finally, and as parents, these exploratory findings led us into questioning about the cost we pay as a society for not investing ourselves as parents more fully during the first years of a child's life. If essential developmental qualities such as emotion regulation, self-esteem, and self-efficacy are all interdependent with attachment; and if attachment depends in great deal of our early attachment interactions, how come we don't work hard on improving the latter more during the crucial first two years of our baby's life? We think that, even from a parental self-interest perspective, it probably makes sense to invest most of our parental energy on those first, attachment years. Those that can't be recalled yet are stocked in our "body" memory (procedural memory) and that seem to have such an influence on our development down the line. For, neglecting the attachment process, might favor what we've seen in all of our samples (e.g., emotion regulation deficits in adulthood), which might in turn demand much more of our attention and resources as parents in years to come.

6.2 Strengths of this work

In the following paragraphs, we would like to emphasize what we consider to be positive achievements of this dissertation work, taken as a whole.

6.2.1 Sampling efforts

We would like to underline here the recruitment efforts that were involved in the present work. If we add-up samples from all the cited experiments, the overall number of participants amounts to N=912 (see details in table 3 below). If we break up these numbers in terms of countries, another characteristic of this work shines forth: it's cross-cultural spread. 536 participants were from France, 172 from the United States, 33 from Tunisia, 131 from Argentina, 40 from Portugal (that we never got to exploit due to time constraints). Those 5 countries span 4 different continents.

N	Country	Article 1	Article 2	Article 3	SURVEY	BAT	AAQ	AAP	AMMI	RSQ	ECR-RS	DERS	RSES	GSE	ACE
302	France	X			X										
131	Argentina	X			X										
87	US	X			X										
59	France	X	X			X	X	X	27/59			X	X	X	
19	US		X			X	X					X	X	X	
175	France			X		X				X		X			X
66	US			X		X				X	47/66	37/66			36/66
33	Tunisia			X		X				X		X			X
40	Portugal					X					X	X			X
912	TOTAL	579	78	274	520	392	78	59	27	274	87	363	78	78	284

Table 3: Samples of this dissertation in the order they were recruited.

6.2.2 An interdisciplinary struggle

One of the difficulties with the present work was the need to learn about numerous subjects spanning a number of disciplines: from adult attachment, developmental psychology and developmental psychopathology, to psychometrics design and validation best practices, from heart physiology to facial expressions, from speech analysis to linguistics, from multimodal fusion and machine learning to programming in Matlab and Xamarin. Each of these areas required some form of training, sometimes in the form of internships or online courses. Additionally, they required constant advice from experts in order to verify our implementations of a given knowledge were correct.

The difficulties associated with interdisciplinary research are known (MacLeod, 2016). One such difficulty is finding reputed journals that publish interdisciplinary research (Rafols, Leydesdorff, O'Hare, Nightingale, & Stirling, 2012).

Interdisciplinary papers are difficult to write. A balance needs to be stroke between breath (i.e., amount of subjects involved) and depth (i.e., level of explanation and amount of actual paragraphs dedicated to each subject), in article formats that are still bound to classic length limitations.

Additionally, interdisciplinary journals typically want their papers to be accessible to readers from all of the areas covered by the journal, which means avoiding domain-specific jargon and using simple language when possible, and keeping all concepts explained within the text.

Despite the difficulties, this process was highly enriching for us as researchers. It pushed us to understand that what seems like an impossible goal to attain in one discipline might have actually been already resolved in another and that, by tending bridges between them, human problems are likely to be resolved much faster, with solutions that end being more comprehensive and integrative.

6.2.3 A project involving basic and applied research

According to the Lawrence Berkeley National Laboratory, basic research concerns "[...] research driven by a scientist's curiosity or interest in a scientific question [...] where the main motivation is to expand man's knowledge, not to create or invent something", and to add: "There is no obvious commercial value to the discoveries that result from basic research". In contrast and quoting from the same source, applied research is about "[...] solving practical problems of the modern world, rather than to acquire knowledge for knowledge's sake" adding "One might say that the goal of the applied scientist is to improve the human condition" (source http://www.sjsu.edu/people/fred.prochaska/courses/ScWk170/s0/Basic-vs.-Applied-Research.pdf).

This fundamental methodological and epistemological distinction offer yet another lens to evaluate the work put throughout this doctoral dissertation.

The work begun squarely as basic research, with open exploratory questions, such as: can music stimuli elicit attachment activation and deactivation in adults? Are automatically-detected facial expressions accurate enough to capture attachment-related reactions? Is it possible to measure heart rate variability changes between time windows of only 40 seconds? Is it possible to integrate data from different modalities, such as paralinguistic characteristics of speech and gaze direction, to produce an attachment score? The list of open questions went on, there were dozens of them.

Our first study was almost entirely situated on this side of the basic/applied science threshold: its approach was mostly exploratory. And at that time it was suggested to us by several colleagues, that we should safely plateau on that level of research, conducting more exploratory work, the rationale being that basic research was enough for a doctoral dissertation. We were warned also that embarking in applied research would necessarily require some form of funding and a team, and that attempting such within the constraints of a 3-year doctoral program was doomed to fail.

But as we progressed along the way, we became convinced that building an actual working tool was in our reach. Moving into applied research required a change in mindset and an increased standard of rigor. In applied research, we provide evidence of the actual functionality of a system or product, and in doing so, we launch it into the world, where actual consequences can occur (e.g., wrong assessment or interpretation of attachment insecurity). The stakes became much higher.

In addition, new concerns, such as application ergonomics (a discipline focused on making products and tasks comfortable and efficient for the user), or programming code efficiency, alien to us on the previous basic research phase, became paramount. While before we could accept the BAT software to process the analyses for 3 hours on our computer, and we could tolerate for it to have no graphical interface, software for other people to use demanded short analysis time and a simple, elegant, graphical interface. And whereas in our lab experiments for our first article it was enough for the software to be compatible with our own computer and camera, a "software for the masses" should be capable of functioning in diverse systems and configurations and to alert the user when something was not working and why.

It goes without saying that this added a large load of work and implied training in additional disciplines, such as application coding for mobile devices.

But the end product of that effort is an actual BAT application that works. Its beta version is available under invitation on the Microsoft Store (see https://www.microsoft.com/en-us/p/biometric-attachment-test/9n5nzdbh7wff), and it was successfully used by our sample of more than 300 people from Portugal, France, the United States, and Tunisia. The application works in several languages: Spanish, French, Portuguese, Hebrew, and English. It can perform all the analyses except for the multimodal fusion (see 6.4.1), without storing nor transmitting any identifiable data, thus upholding user privacy, another key principle of the BAT design. The program was coded in such a way that it can easily be ported to other platforms, such as iOS (to work on iPads and iPhones), Android (to work on mobile phones), and MacOs. Most importantly, the application, as well as its different algorithms, were the focus of two validation studies (articles 2 and 3; see 5.2 and 5.3).

6.2.4 The BAT as a potential solution to the SSP-AAI attachment stability controversy

In 2014, Groh and colleagues' provocative results shocked the attachment theory community (Groh et al., 2014). Their work with a cohort of 819 participants studied longitudinally from birth till late adolescence showed that the participants' adult attachment measures at 18 years, assessed with the AAI, were almost not correlated at all to the participants' childhood attachment ones, measured at 15 months through the SSP. This discovery challenged the notion of attachment continuity and stability across the lifespan - one of the theory's main tenets.

While an obvious line of explanation would be to posit that attachment might simply change more over time than we originally thought, this notion would go counter the evidence that shows strong correlations between childhood attachment, measured by the SSP, and adult psychological disturbances such as psychopathologies or deficits in emotion regulation (see 4.1). If the attachment developmental trace is strong enough to increase the risk for psychopathology or reduce emotion regulation capacities decades later, how come that the AAI, a test specifically designed to detect attachment in adults, cannot measure that trace at all?

Instead of choosing to explain Groh et al.'s results as a problem with the concept of stability in attachment theory, we hypothesize their discovery shows the limitations of the "gold-standard" AAI. As we've seen during our introduction (section 4.3), the AAI scores are obtained through a complex coding process applied to a structured interview transcript, containing what was said by the participant plus a minimalistic set of their annotated behaviors (e.g. crying). It is arguably a unimodal test since the analyzed information comes mostly from one form of information: language. Except for rare cases in which the AAI interviewer doubles as the AAI coder, coders don't get to see or hear the participant at all; they find themselves scoring a language-transcript with no cues as to how those words were uttered, in which tone, looking down or towards the eyes, accompanied by which facial expression, followed by what autonomic nervous system response. If we compare the AAI's scoring system to the SSP, the "gold-standard" test for children used in Groh et al. study, we discover that for scoring, children are directly observed through a one-way mirror (or in videotaped sessions) as they experience separation from and then reunion with their mothers: movements, sounds, facial expressions, gaze, the SSP coder has access to all of these sensory modalities, being able to cross-verify hypotheses across them, and is able to observe different behaviors happening simultaneously, revealing their combined meaning.

Attachment-related experiences, the psychophysiological traces they leave, and the behavioral reactions to them, develop most critically between 6 months and 2 years of age, whereas episodic memory develops between 3 and 4 years of age, which means that from a developmental perspective, these traces should be expected to be mostly nonverbal (Brown & Elliott, 2016).

If so, why do we expect them to appear on language transcripts? We understand the AAI coding system is not based on content but on linguistic features believed to be unconscious, but it seems fair to us to argue that spoken language should not be our main object of study when we are trying to detect attachment traces. Attachment-related language characteristics seem more like the remote echo of a phenomenon whose epicenter is located deep down within more primal (i.e., subcortical) layers of our brains. Preliminary proof for this was found in our second study, where removing language analysis from the BAT decreased its scoring performance by only 10-13% (for anxiety and avoidance respectively).

We speculate therefore that the lack of correlation between SSP results at 15 months and AAI results at 18 years were revealing of an AAI failure to pick up the expected longitudinal traces due to the AAI's almost exclusive reliance on transcribed language. There is some preliminary proof for holding this hypothesis: the AMMI, which is similar to the AAI but adds probes to gain insight about current behavior in addition to the recollection of the person's childhood relationships, has shown a correlation with the SSP for a longitudinal sample (Miljkovitch, Moss, Bernier, Pascuzzo, & Sander, 2015). Following our reasoning, we believe the AMMI was better than the AAI at this because it attempts to incorporate at least two modalities: language and self-reported behaviors.

We believe that the BAT could do even better in this task, perhaps helping to find the "lost longitudinal traces" of attachment, as it measures all the dimensions that are evaluated in childhood through the SSP, with the important addition of physiological measures - offering us privileged access to the phenomenon's epicenter.

We've been in contact with Glenn Roisman, coauthor of the Groh et al. paper and researcher in charge of the SECCYD longitudinal cohort that the paper is based upon. It seems plausible that in the foreseeable future he'll grant us access to that cohort to measure them using the BAT and have access to their historical SSP measures.

If the BAT would be trained to recognize SSP traces, by being fed a sample of people tested with the BAT as adults and that was tested with the SSP in childhood, the resultant scoring model would potentially be the most theoretically sound measure of adult attachment to date.

6.3 Limitations of this work

In addition to the limitations of each of the studies we conducted, that were addressed in the respective articles above (see sections 5.1, 5.2, and 5.3), there are more general limitations of this work that we would like to discuss.

The first one is related to studies 2 and 3 and is our reliance on self-report measures to test for validity. If self-report measures are biased (see section 4.3), and the BAT is basically learning to imitate them, wouldn't be the BAT biased as well? The answer is of course yes. In order for the BAT to become better than questionnaires in terms of construct validity, a BAT scoring model should be trained using something better than a questionnaire, such as the AAI, the Adult Attachment Narratives (Bretherthon, 2002) or, as we described in section 6.2.4, the childhood's SSP scores of an adult sample followed longitudinally. The reasons that forced us to choose questionnaires for our two validation studies were purely practical: we did not have the funds to test all those people with interview-based assessments, and even if we had, it would have been difficult to interview, transcribe, and score transcripts for 392 people within the time constraints of this doctoral program. While we have tested content validity of the BAT vis-a-vis two interview-based measures (the AAP and the AMMI) on our fist study (see 5.1), an argument could still be made that until and unless we do train a scoring model for the BAT based on AAI scores and successfully demonstrate convergent validity with it, in a strictly developmental sense our BAT test has not proven construct validity yet.

To this critique, we respond in two ways. The first is to mention that indeed developing BAT scoring models based on interview-based measures is a priority and definitely in our roadmap (see 6.4.1). The second is to mention that although we share construct validity concerns about questionnaires somewhat, our own analyses (see section 6.1.4) seem to show a high level of theoretical coherence for attachment as measured by RSQ and AAQ questionnaires, showing the expected associations with emotion regulation, self-esteem, self-efficacy and childhood trauma in all of our samples. If we add to this the ongoing AAI controversy regarding continuity of attachment (see section 6.2.4), which brings the construct validity of the AAI into question, we conclude that our choices of using questionnaires for preliminary convergent validity measures were not that bad.

Another critique of our work could be that we have not, even in the discussion of this dissertation, offered some systematic comparison of the BAT's psychometric properties versus those of existing measures. This point would be reasonable given the strong critiques we make of those other measures all throughout this work, and the constant implication that the BAT would be better than them. Is the BAT's test-retest better than human-coded tests? Does it possess better cross-cultural validity or higher convergence validity than questionnaires?

The truth is we have not enough results yet to attempt a systematic comparison. For example, we don't have a scoring model yet trained to produce classifications, to compare with AAI or similar categorical measures. Comparing convergence levels between the BAT and other measures would be unreasonable too since most measures do so against the "gold standard" AAI, something we have not yet been able to do. Perhaps the only comparison we can make with the limited data we have is evaluating test-retest in the BAT RSQ model versus other dimensional measures of attachment that evaluate anxiety and avoidance. We offer a non-exhaustive comparison on table 4, which suggests that the BAT scoring models still need to improve in terms of stability over time in order to live to the expectation of being more reliable than classic adult attachment measures.

Measure	Anxiety	Avoidance	Interval	N
BAT (RSQ model)	0.74	0.79	2 weeks	30
AAI *	0.66	0.65	8 months	144
ECR **	0.68	0.71	6 months	207
ECR-Short Form **	0.8	0.83	1 month	165
RSQ ***	0.85	0.8	2 days	106

Table 4: Non-exhaustive comparison of the test-retest reliability of the anxiety and avoidance dimensions in different measures of adult attachment. * (Scharfe & Bartholomew, 1994), ** (Lopez & Gormley, 2002), *** (Wei, Russell, Mallinckrodt, & Vogel, 2007),**** (Guédeney, Fermanian, & Bifulco, 2010)

Finally, another limitation of our work is the fact that we barely mentioned attachment disorganization, a crucial element of attachment theory that is measured by several of the existing measures. Disorganization is especially important in the context of childhood trauma, for internal working models of attachment formed within such a context tend to be dysregulated (Solomon & George, 2011). When these internal representations are invoked in adults, they induce a state in which behavior and thought become disorganized and disoriented, either through emotional flooding or through attempts to prohibit or block emotions from consciousness (Solomon & George, 2011; Bowlby, 1969).

The literature tells us that adults who have not been able to 'resolve' (i.e., re-organize) their internal working models are over-represented in psychiatric populations (Stovall-McClough & Dozier, 2016). Therefore, early traumatic interactions between children and their caregivers impede the development of regulation strategies that can buffer adults from severe psychological distress, and instead, can compound it (Bacon & Richardson, 2001). Given the importance of attachment disorganization in adults, and given the fact the BAT was designed to be able to measure it (see, for instance, section 4.3.3), why haven't we formally investigated the BAT's capacity to assess attachment disorganization in our studies? The response has to do with time and budget constraints. For our second French sample (N=59), that we explored in our first study, we got measures that included attachment disorganization, namely the AAP and the AMMI. But during that study, we decided to focus on the primary content validity experiment we could perform: to evaluate whether or not the different BAT modalities could measure attachment security. For our second and third studies, we used samples for which we only had questionnaires, AAQ and RSQ, respectively, that don't provide an accurate measure of disorganization. Therefore, it wasn't possible to create a scoring model for the BAT to measure disorganization using those samples. As we describe in section 6.2.4, it is clearly in our future roadmap to create scoring models for the BAT based on assessments that do measure attachment disorganization. Since the BAT, like the AAP, implements stimuli that represent relational trauma, it seems reasonable to expect that the BAT will be also able to measure attachment disorganization in the future.

6.4 Future research directions

6.4.1 Immediate roadmap: what needs to happen for the BAT to become useful

Reading this dissertation, and particularly the last article, one might get the impression that the BAT is a tool ready to be used by researchers and clinicians. This is not entirely true yet.

What are the steps that still need to be taken in order to make the BAT a useful tool?

The first step is to debug the software. The BAT v3 software implemented for testing the 274 recruits in our third study (plus the N=40 recruits from Portugal) worked well for testing them, but produced errors with other participants, that couldn't complete the procedure. The errors reported to us were varied, going from full crashes of the application at a certain moment during the test, to camera, microphone or system incompatibilities, to errors that appeared at last minute during the data transfer phase, when the anonymized results data is sent to us for analysis before it is erased. Although for our purposes this beta application was good enough, it is unacceptable to provide researchers or clinicians with software that can at times crash, and so the BAT requires a thorough process of debugging until the app becomes fully stable.

Second, in its current version, the BAT v3 application can present the images and music clips, and perform all the feature extraction analyses: it can obtain heart rate, facial expressions, paralinguistic and linguistic speech features, gaze direction and head pose. However, it cannot fuse the data yet, and so it cannot produce scores on its own.

All fusion algorithms still run on our computer using Matlab, processing the multimodal data generated remotely by the application. In order for the BAT app to be fully autonomous, the fusion algorithms, which produce attachment scores, should be translated from Matlab to Xamarin code and included in the actual application that people use. This task is not to be underestimated, both codes (Matlab source versus Xamarin translation) should be exhaustively compared in several situations to prove they do produce the exact same scores.

Finally and once the application is stable and can on its own produce final scores, we would need to facilitate other researchers with the software so that they can conduct independent validation studies. Indeed, we would feel more comfortable with the BAT being implemented in either clinical or research settings after undergoing at least some external validation studies, for the sake of scientific replicability and psychological deontology.

6.4.2 Potential implementations of the BAT in a clinical setting

As we conclude this dissertation, we would like to offer a concrete example of how the BAT can be helpful in a clinical setting, where we do not deal with research subjects, but with actual patients.

The BAT software was installed at the testing computer in the Center for Psychotherapy and Psychotrauma of Ville-Evrard, a public health clinic in Saint-Denis, France, devoted to the treatment of PTSD and CPTSD and where we work as clinical psychologists (http://www.eps-ville-evrard.fr/institution/projets/centre-de-psychotherapie/). That is the computer all patients use at intake and throughout treatment to respond to different kinds of questionnaires, such as quality of life and symptoms severity scales.

During the last year, the BAT has been used by patients that decided to collaborate with this research, who would take the test in that room. Many of those that did collaborate told us and a number of our colleagues that, of all the “tests” they had to undergo during intake (most of them self-report questionnaires), the BAT was the most entertaining one and thus the easiest to complete. Indeed, used to respond to several 20-minutes-long, repetitive questionnaires with sometimes very intrusive questions (which are necessary to gather important information to provide correct treatment), the BAT's picture and music stimuli seemed rather attractive and the task, to respond with their voice freely, mentioning what did they feel, apparently was also simple to complete.

Psychologists and psychiatrists in the center begun imagining a day when all intake patients were to take the BAT test, so that the center's director, who distributes the patients to the psychologists depending on their disturbances, could easily and timely direct those with an insecure attachment to therapists that practice an attachment-specific method. This is of particular importance to our center because PTSD and CPTSD often look similar in terms of salient symptoms but differ in terms of attachment classification. For a person suffering from CPTSD, a stabilization phase is recommended before beginning the actual treatment of the trauma, and such phase involves, in the presence of attachment insecurity, some form of therapy that deals with that first (Parra, George, Kalalou, & Januel, 2017; Brown & Elliott, 2016; Parnell & Siegel, 2015).

While the AAP is sometimes used in our center for the same purpose, it takes 20 minutes to administer and 30-40 minutes to score, and subjectivity remains a problem. The result is that most often the AAP is not administered, or not scored, and thus the patients' attachment characteristics are known only much later, in the context of an already started psychotherapy.

That is one example of how the BAT could be used in the public mental health system in France and elsewhere to improve treatment. We can only imagine how many other examples could exist, both in private and public mental health clinics like ours, but also private practices around the world.

6.4.3 Potential implementations of the BAT in research settings

Perhaps the single most important reason that moved us to create the BAT was our own experience as undergraduates attempting to perform research on adult attachment on a budget. We soon discovered that with a low budget, the options for accurately measuring adult attachment are reduced to self-report measures, which have limitations as we have seen (see section 4.3). We became aware of dozens of colleagues that, facing this limitation, decided to investigate something else.

And so our journey began trying to solve this problem. The BAT's future holds a version of the software that will be free for public University researchers to use.

We believe this could unleash attachment research, compounding the sheer number and the creativity of undergraduate researchers with the objectivity and construct validity of our measure.

This new wave of attachment research would be more highly replicable, for the studies, varied as they might be, would have a common measure of attachment.

Of course, low-budgets are not a matter exclusive to undergraduate researchers, but a common denominator in the sciences around the world (Jacob & Lefgren, 2011; Muñoz, Moreno, & Luján, 2010). We cannot promise all uses of the BAT, by all researchers in all scenarios, will be free of charge. But we do promise that we'll honor the principle of making the BAT available to all those who sincerely want to help the attachment field move forward. This might happen through some form of mutual collaboration agreement, single-use-case rental agreements, or other arrangements, as necessary.

6.4.4 A means to an end: towards the Multimodal Developmental Profile

We became interested in creating a holistic measure that could evaluate both observable behavior and invisible physiological traits when we first got in contact with the literature on developmental psychopathology (Sroufe & Rutter, 1984).

Having some expertise in the clinical treatment of PTSD and CPTSD ourselves, the idea of an adult experiencing severe symptoms as a result of having experienced one or several adverse events in the past was commonplace to us. But what developmental psychopathology proponents were bringing forward went much further than PTSD: they implied that many psychopathologies could have a developmental origin.

Electrified by this idea, we wondered what kind of a test could measure such developmental traces that could favor the development of psychopathologies. We knew attachment was a central component to that and therefore we made the Biometric Attachment Test to measure it.

But what about emotion regulation, self-esteem, or the sequelae of childhood trauma? These dimensions are currently assessed by different self-report measures (DERS, RSES, and ACE respectively) but, could one 9 minutes test measure them all?

In late 2019, we'll begin conducting research using the BAT with Israeli participants. The BAT will also be used in a sample of military subjects suffering from severe trauma in France. Participants will be measured for adult attachment, but also emotion regulation and adverse childhood experiences. We already have a database of more than 360 participants from different countries (Portugal, the United States, France, and Tunisia) for which we have these measures as well as the BAT.

The BAT has stimuli, both pictures, and music, that can elicit content related to adverse childhood experience, and during the BAT administration, there are several opportunities to measure emotion regulation.

We hypothesize that, with proper training, the BAT multimodal fusion algorithm should be able to pick up patterns of response that allow it to score these dimensions.

If we succeed, the resultant test will coin a new name that more clearly convey a capacity to measure different developmental traces in adults. We thought of rebaptizing the test, then, the Multimodal Developmental Profile.

6.4.5 The IPF trial

From 2014 to 2016, we led a pilot study at the Center of Psychotherapy and Psychotrauma of Ville-Evrard (Saint-Denis) to independently test the efficacy of Brown and Elliott's IPF method in the treatment of CPTSD patients (Brown & Elliott, 2016; Parra, George, Kalalou, & Januel, 2017). The method uses semi-structured visualizations in which the patient is instructed to imagine him or herself as a child interacting with a set of parents that embody the qualities that research shows favor secure attachment. Our results, as well as Brown and Elliott's own study, showed that patients improved in terms of attachment, quality of life, emotion regulation, relationship quality, as well as CPTSD and general symptoms, even after a short treatment. Moreover, results were stable 8 months later. After publishing that work, we imagined a larger trial that would include patients with other psychological disorders other than PTSD, so that we could evaluate the impact improving attachment quality could have on the course of psychopathology. In addition, we wanted to be able to evaluate the mechanism of change that is the heart of IPF, that is, the enactment of positive attachment-related experiences through the visualization, as separate from the therapeutic action of epiphenomena such as the positive impact of the therapeutic relationship or the positive effect of relaxation. We imagined a study design that would be automated: an application that would randomly administrate one of two prerecorded versions of the IPF visualization, one the true practice, the other a "sham" version. The sham version would be identical to the true one in all but one aspect: the imagined parents in the sham visualization would do positive but attachment-unrelated things towards the patient-as-child, such as buying them candy or taking them shopping. We thought that if a great difference in efficacy is found between the treatment conditions, this could support the notion that positive, attachment-related interactions between infants and their caregivers, whether real or imagined by adults, could represent a powerful psychological healing mechanism. If this would work on several psychopathologies, it would further demonstrate an important role of attachment within the maintaining of such psychopathologies.

But we couldn't execute that protocol because we were missing a key ingredient: a valid, objective, automatic attachment test that could be taken at home by participants on their phone or tablets. A test capable of avoiding priming effects, since we needed participants to take it regularly to measure progress. And a test short enough, engaging enough, that participants would actually take it.

Now that the BAT exists, this trial protocol becomes viable. We look forward to working with the IPF creators to conduct it.

6.5 Conclusion: the dawn of a new age of psychometrics?

We see our work as part of a wave occurring in many fields: humans, slowly learning how to tame the sheer power of computer intelligence, for the benefit of humans.

Psychometrics, a discipline born around the 19th century in both its Victorian and German streams, hasn't really changed much ever since. As we've seen throughout this work, an injection of artificial intelligence can update the psychometric tradition, and bring forth several qualities it so far missed.

In so doing, it can help the attachment field move forward, acquiring unprecedented objectivity and agility.

As we have seen as well, this is not without its ethical dilemmas, even though we are inclined to think the benefits are poised to outweigh the risks in the long run.

In our dreams, we imagine it could be a few years from now that, as we walk into a psychiatric hospital, we are presented with a computer or tablet that introduces us to some images, film scenes, or music, and that 10 minutes later can tell us what are the affected psychological dimensions that are most likely culprits of our symptoms.

But it might very well not turn out that way. A dear colleague from Harvard Divinity School once told us, that obtaining a doctoral degree in a given field was analogous to a peak of a narrow mountain rising away from the ground: the ground meant to represent common knowledge, the peak, the piercing effect of focalizing our attention on a single subject for a long period of time, and the narrowness of the mountain, representing the fragility of such acquired knowledge, in so far as it is incredibly limited in its scope.

We do subscribe to this prescription for humility. For example, we do not know if humans will accept or reject the power of artificial intelligence, tamed or otherwise, when it comes to assessing their mental health.

In all cases, our deepest hope, remains: that this work may eventually help to alleviate, if it is a little, the suffering associated with insecure attachment in our fellow human beings.

7 References

A. Schmidt, G. T. (1989). Autonomic Nervous System. In Human Physiology (2nd ed.) (pp. 333–370.). Springer-Verlag.

Ainsworth, M. D. (1985). Attachments across the life span. Bull N Y Acad Med, 61, 792–812.

Ainsworth, M. D. S., & Bell, S. M. (1970). Attachment Exploration, and Separation: Illustrated by the Behavior of One-Year-Olds in a Strange Situation. Child Development, 41(1), 49. https://doi.org/10.2307/1127388

Ainsworth, M. D. S., Blehar, M. C., Waters, E., & Wall, S. (1978). Patterns of attachment: A psychological study of the strange situation. Psychology Press, XVIII.

Austin, J. L. (1975). How To Do Things With Words. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198245537.001.0001

Bacon, H., & Richardson, S. (2001). Attachment theory and child abuse: an overview of the literature for practitioners. Child Abuse Review, 10(6), 377–397. https://doi.org/10.1002/car.718

Baevsky, R. M., Bogomolov, V. V., Funtova, I. I., Slepchenkova, I. N., & Chernikova, A. G. (2011). Prospects of medical monitoring of long-duration space flights by means of non-contact recording of physiological functions during sleep time. Human Physiology, 37(7), 816–820. https://doi.org/10.1134/s036211971107005x

Baltrusaitis, T., Robinson, P., & Morency, L.-P. (2016). OpenFace: An open source facial behavior analysis toolkit. In 2016 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE. https://doi.org/10.1109/wacv.2016.7477553

Barnas, M. V., & Cummings, E. M. (1994). Caregiver stability and toddlers' attachment-related behavior towards caregivers in day care. Infant Behavior and Development, 17(2), 141–147. https://doi.org/10.1016/0163-6383(94)90049-3

Bartholomew, K., & Horowitz, L. M. (1991). Attachment styles among young adults: A test of a four-category model.. Journal of Personality and Social Psychology, 61(2), 226–244. https://doi.org/10.1037/0022-3514.61.2.226

Baumgartner, T., Lutz, K., Schmidt, C. F., & Jäncke, L. (2006). The emotional power of music: How music enhances the feeling of affective pictures. Brain Research, 1075(1), 151–164. https://doi.org/10.1016/j.brainres.2005.12.065

Bavelier, D., & Neville, H. J. (2002). Cross-modal plasticity: where and how?. Nature Reviews Neuroscience, 3(6), 443–452. https://doi.org/10.1038/nrn848

Bergman, K., Sarkar, P., Glover, V., & O'Connor, T. G. (2010). Maternal Prenatal Cortisol and Infant Cognitive Development: Moderation by Infant–Mother Attachment. Biological Psychiatry, 67(11), 1026–1032. https://doi.org/10.1016/j.biopsych.2010.01.002

Bobbia, S., Benezeth, Y., & Dubois, J. (2016). Remote photoplethysmography based on implicit living skin tissue segmentation. In 2016 23rd International Conference on Pattern Recognition (ICPR). IEEE. https://doi.org/10.1109/icpr.2016.7899660

Boucsein, W. (2012). Electrodermal Activity. Springer Science & Business Media.

Bowlby, J. (1969). Attachment and loss. Basic Books.

Brooks, S. J., Nilsson, E. K., Jacobsson, J. A., Stein, D. J., Fredriksson, R., Lind, L., & Schiöth, H. B. (2014). BDNF Polymorphisms Are Linked to Poorer Working Memory Performance Reduced Cerebellar and Hippocampal Volumes and Differences in Prefrontal Cortex in a Swedish Elderly Population. PLoS ONE, 9(1), e82707. https://doi.org/10.1371/journal.pone.0082707

Brown, D. P., & Elliott, D. (2016). Attachment disturbances in adults: treatment for comprehensive repair. Norton.

Buchheim, A., & Mergenthaler, E. (2000). The Relationship Among Attachment Representation Emotion-Abstraction Patterns, and Narrative Style: A Computer-Based Text Analysis of the Adult Attachment Interview. Psychotherapy Research, 10(4), 390–407. https://doi.org/10.1093/ptr/10.4.390

Calvo, R. A., D'Mello, S. K., Gratch, J., & Kappas, A. (2015). Oxford Handbook of Affective Computing. Oxford University Press.

Cantazaro, A., & Wei, M. (2010). Adult Attachment Dependence, Self-Criticism, and Depressive Symptoms: A Test of a Mediational Model. Journal of Personality. https://doi.org/10.1111/j.1467-6494.2010.00645.x

Carlson, E. A. (1998). A Prospective Longitudinal Study of Attachment Disorganization/Disorientation. Child Development, 69(4), 1107. https://doi.org/10.2307/1132365

Carrick, O. K., Thompson, J. C., Epling, J. A., & Puce, A. (2007). It's all in the eyes: neural responses to socially significant gaze shifts.. Neuroreport, 18, 763–766.

Cassidy, J., & Shaver, P. R. (2016). Handbook of attachment: Theory, research, and clinical applications. 3rd edition. Guilford Press.

Cassidy, J., Sherman, L. J., & Jones, J. D. (2012). What's in a word? Linguistic characteristics of Adult Attachment Interviews. Attachment & Human Development, 14(1), 11–32. https://doi.org/10.1080/14616734.2012.636649

Cecchini, M., Iannoni, M. E., Pandolfo, A. L., Aceto, P., & Lai, C. (2015). Attachment style dimensions are associated with brain activity in response to gaze interaction. Soc Neurosci, 10, 282–293.

Christopher, G., & MacDonald, J. (2005). The impact of clinical depression on working memory. Cognitive Neuropsychiatry, 10(5), 379–399. https://doi.org/10.1080/13546800444000128

Collins, N. L., & Read, S. J. (1990). Adult attachment working models, and relationship quality in dating couples.. Journal of Personality and Social Psychology, 58(4), 644–663. https://doi.org/10.1037/0022-3514.58.4.644

Crawford, T. N., Livesley, W. J., Jang, K. L., Shaver, P. R., Cohen, P., & Ganiban, J. (2007). Insecure attachment and personality disorder: a twin study of adults. European Journal of Personality, 21(2), 191–208. https://doi.org/10.1002/per.602

Cummins, N., Scherer, S., Krajewski, J., Schnieder, S., Epps, J., & Quatieri, T. F. (2015). A review of depression and suicide risk assessment using speech analysis. Speech Communication, 71, 10–49. https://doi.org/10.1016/j.specom.2015.03.004

Dalsant, A., Truzzi, A., Setoh, P., & Esposito, G. (2015). Maternal bonding in childhood moderates autonomic responses to distress stimuli in adult males. Behavioural Brain Research, 292, 428–431. https://doi.org/10.1016/j.bbr.2015.06.026

Dan-Glauser, E. S., & Scherer, K. R. (2011). The Geneva affective picture database (GAPED): a new 730-picture database focusing on valence and normative significance. Behavior Research Methods, 43(2), 468–477. https://doi.org/10.3758/s13428-011-0064-1

Darwin, C. (1872). The Expression of the Emotions in Man and Animals. John Murray.

Degottex, G., Kane, J., Drugman, T., Raitio, T., & Scherer, S. (2014). COVAREP: A collaborative voice analysis repository for speech technologies. In 2014 IEEE International

Conference on Acoustics Speech and Signal Processing (ICASSP). IEEE. https://doi.org/10.1109/icassp.2014.6853739

Dhamija, S., & Boult, T. E. (2018). Automated Action Units Vs. Expert Raters: Face off. In 2018 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE. https://doi.org/10.1109/wacv.2018.00035

Diamond, L. M., & Fagundes, C. P. (2010). Psychobiological research on attachment. Journal of Social and Personal Relationships, 27(2), 218–225. https://doi.org/10.1177/0265407509360906

Ditzen, B., Schmidt, S., Strauss, B., Nater, U. M., Ehlert, U., & Heinrichs, M. (2008). Adult attachment and social support interact to reduce psychological but not cortisol responses to stress. Journal of Psychosomatic Research, 64(5), 479–486. https://doi.org/10.1016/j.jpsychores.2007.11.011

Dore, J. (1975). Holophrases speech acts and language universals. Journal of Child Language, 2(1), 21–40. https://doi.org/10.1017/s0305000900000878

Doron, G., Moulding, R., Nedeljkovic, M., Kyrios, M., Mikulincer, M., & Sar-El, D. (2011). Adult attachment insecurities are associated with obsessive compulsive disorder. Psychology and Psychotherapy: Theory Research and Practice, 85(2), 163–178. https://doi.org/10.1111/j.2044-8341.2011.02028.x

Eerola, T., & Vuoskoski, J. K. (2010). A comparison of the discrete and dimensional models of emotion in music. Psychology of Music, 39(1), 18–49. https://doi.org/10.1177/0305735610362821

Ein-Dor, T., Doron, G., Solomon, Z., Mikulincer, M., & Shaver, P. R. (2010). Together in pain: Attachment-related dyadic processes and posttraumatic stress disorder. Journal of Counseling Psychology, 57(3), 317–327. https://doi.org/10.1037/a0019500

Ekman, P., & Friesen, W. V. (1978). Manual for the Facial Action Code. Consulting Psychologist Press.

Emery, N. J. (2000). The eyes have it: the neuroethology function and evolution of social gaze. Neuroscience & Biobehavioral Reviews, 24(6), 581–604. https://doi.org/10.1016/s0149-7634(00)00025-7

Farina, B., Speranza, A. M., Imperatori, C., Quintiliani, M. I., & Marca, G. D. (2014). Change in Heart Rate Variability After the Adult Attachment Interview in Dissociative Patients. Journal of Trauma & Dissociation, 16(2), 170–180. https://doi.org/10.1080/15299732.2014.975309

Feeney, J. (2016). Adult romantic attachment: developments in the study of couple relationships. In Handbook of Attachment: Theory, Research, and Clinical Applications, Third edition. Guilford Press.

Felitti, V. J., Anda, R. F., Nordenberg, D., Williamson, D. F., Spitz, A. M., Edwards, V., ... Marks, J. S. (2019). Relationship of Childhood Abuse and Household Dysfunction to Many of the Leading Causes of Death in Adults: The Adverse Childhood Experiences (ACE) Study. American Journal of Preventive Medicine, 56(6), 774–786. https://doi.org/10.1016/j.amepre.2019.04.001

Fortuna, K., & Roisman, G. I. (2008). Insecurity stress, and symptoms of psychopathology: contrasting results from self-reports versus interviews of adult attachment. Attachment & Human Development, 10(1), 11–28. https://doi.org/10.1080/14616730701868571

Fraley, R. C., & Roisman, G. I. (2015). Do early caregiving experiences leave an enduring or transient mark on developmental adaptation?. Current Opinion in Psychology, 1, 101–106. https://doi.org/10.1016/j.copsyc.2014.11.007

Garbarino, M., Lai, M., Tognetti, S., Picard, R., & Bender, D. (2014). Empatica E3 - A wearable wireless multi-sensor device for real-time computerized biofeedback and data acquisition. In Proceedings of the 4th International Conference on Wireless Mobile Communication and Healthcare - Transforming healthcare through innovations in mobile and wireless technologies. ICST. https://doi.org/10.4108/icst.mobihealth.2014.257418

Gecas, V. (1989). The Social Psychology of Self-Efficacy. Annual Review of Sociology, 15(1), 291–316. https://doi.org/10.1146/annurev.so.15.080189.001451

George, C., & West, M. (2012). The adult attachment projective picture system: attachment theory and assessment in adults. Guilford Press.

George, C., & West, M. (2001). The development and preliminary validation of a new measure of adult attachment: the Adult Attachment Projective. Attachment & Human Development, 3(1), 30–61. https://doi.org/10.1080/14616730010024771

Giudice, M. D., & Colle, L. (2007). Differences between children and adults in the recognition of enjoyment smiles. Developmental Psychology, 43(3), 796–803. https://doi.org/10.1037/0012-1649.43.3.796

Gormley, B., & McNiel, D. E. (2009). Adult Attachment Orientations Depressive Symptoms, Anger, and Self-Directed Aggression by Psychiatric Patients. Cognitive Therapy and Research, 34(3), 272–281. https://doi.org/10.1007/s10608-009-9267-5

Gratz, K. L., & Roemer, L. (2004). Multidimensional Assessment of Emotion Regulation and Dysregulation: Development Factor Structure, and Initial Validation of the Difficulties in Emotion Regulation Scale. Journal of Psychopathology and Behavioral Assessment, 26(1), 41–54. https://doi.org/10.1023/b:joba.0000007455.08539.94

Griffin, D. W., & Bartholomew, K. (1994). Models of the self and other: Fundamental dimensions underlying measures of adult attachment. Journal of Personality and Social Psychology, 67(3), 430–445. https://doi.org/10.1037/0022-3514.67.3.430

Groh, A. M., Roisman, G. I., Booth-LaForce, C., Fraley, R. C., Owen, M. T., Cox, M. J., & Burchinal, M. R. (2014). IV. Stability of attachment security from infancy to late adolescence. Monographs of the Society for Research in Child Development, 79(3), 51–66. https://doi.org/10.1111/mono.12113

Gross, J. J. (2002). Emotion regulation: Affective cognitive, and social consequences. Psychophysiology, 39(3), 281–291. https://doi.org/10.1017/s0048577201393198

Grossmann, K., Grossmann, K., & Keppler, A. (2005). Universal and Culture-Specific Aspects of Human Behavior: The Case of Attachment. In Culture and Human Development: The Importance of Cross-Cultural Research to the Social Sciences. Hove: Psychology Press.

Guastella, A. J., Mitchell, P. B., & Dadds, M. R. (2008). Oxytocin Increases Gaze to the Eye Region of Human Faces. Biological Psychiatry, 63(1), 3–5. https://doi.org/10.1016/j.biopsych.2007.06.026

Guédeney, N., Fermanian, J., & Bifulco, A. (2010). La version française du Relationship Scales Questionnaire de Bartholomew (RSQ Questionnaire des échelles de relation) : étude de validation du construit. LEncéPhale, 36(1), 69–76. https://doi.org/10.1016/j.encep.2008.12.006

Halpern, Y., Hall, K., Schogol, V., Riley, M., Roark, B., Skobeltsyn, G., & Bäuml, M. (2016). Contextual Prediction Models for Speech Recognition. In Interspeech 2016. ISCA. https://doi.org/10.21437/interspeech.2016-1358

Hane, A., & Fox, N. (2016). Studying the biology of human attachment. In Handbook of Attachment: Theory, Research, and Cinical Applications, third edition. Guilford Press.

Hermes, D. (2015). The Art of Xamarin App Development. In Xamarin Mobile Application Development (pp. 387–388). Apress. https://doi.org/10.1007/978-1-4842-0214-2_10

Hunt, M., Auriemma, J., & Cashaw, A. C. A. (2003). Self-Report Bias and Underreporting of Depression on the BDI-II. Journal of Personality Assessment, 80(1), 26–30.

https://doi.org/10.1207/s15327752jpa8001_10

Illing, V., Tasca, G. A., Balfour, L., & Bissada, H. (2010). Attachment Insecurity Predicts Eating Disorder Symptoms and Treatment Outcomes in a Clinical Sample of Women. The Journal of Nervous and Mental Disease, 198(9), 653–659. https://doi.org/10.1097/nmd.0b013e3181ef34b2

Jacob, B. A., & Lefgren, L. (2011). The impact of research grant funding on scientific productivity. Journal of Public Economics, 95(9-10), 1168–1177. https://doi.org/10.1016/j.jpubeco.2011.05.005

Johnston, M., Weinman, J., & Wright, S. C. (1995). Measures in health psychology : a user's portfolio. NFER-NELSON.

Junghaenel, D. U., Smyth, J. M., & Santner, L. (2008). Linguistic Dimensions of Psychopathology: A Quantitative Analysis. Journal of Social and Clinical Psychology, 27(1), 36–55. https://doi.org/10.1521/jscp.2008.27.1.36

KE Grossmann, E. W., K Grossmann. (2005). Attachment from Infancy to Adulthood: The Major Longitudinal Studies. The Guilford Press.

Kinsinger, F. S. (2009). Beneficence and the professional's moral imperative. Journal of Chiropractic Humanities, 16(1), 44–46. https://doi.org/10.1016/j.echu.2010.02.006

Klagsbrun, M., & Bowlby, J. (1976). Responses to separation from parents: a clinical test for young children. British Journal of Projective Psychology & Personality Study, 21(2).

Kodish-Wachs, J., Agassi, E., Kenny, P. 3rd, & Overhage, J. M. (2018). A systematic comparison of contemporary automatic speech recognition engines for conversational clinical speech.. AMIA Annu Symp Proc, 2018, 683–689.

Koskimäki, H., Mönttinen, H., Siirtola, P., Huttunen, H.-L., Halonen, R., & Röning, J. (2017). Early detection of migraine attacks based on wearable sensors. In Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers on - UbiComp '17. ACM Press. https://doi.org/10.1145/3123024.3124434

Koulomzin, M., Beebe, B., Anderson, S., Jaffe, J., Feldstein, S., & Crown, C. (2002). Infant gaze head, face and self-touch at 4 months differentiate secure vs. avoidant attachment at 1 year: A microanalytic approach. Attachment & Human Development, 4(1), 3–24. https://doi.org/10.1080/14616730210123120

Kraepelin, E. (1921). Manic Depressive Insanity and Paranoia. The Journal of Nervous and Mental Disease, 53(4), 350. https://doi.org/10.1097/00005053-192104000-00057

Laksana, E., Baltrusaitis, T., Morency, L.-P., & Pestian, J. P. (2017). Investigating Facial Behavior Indicators of Suicidal Ideation. In 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017). IEEE. https://doi.org/10.1109/fg.2017.96

Lang, P., Bradley, M., & Cuthbert, B. International affective picture system (IAPS): Technical manual and affective ratings.

Larose, S., & Bernier, A. (2001). Social support processes: mediators of attachment state of mind and adjustment in late adolescence. Attachment & Human Development, 3(1), 96–120. https://doi.org/10.1080/14616730010024762

Lazzouni, L., Voss, P., & Lepore, F. (2012). Short-term crossmodal plasticity of the auditory steady-state response in blindfolded sighted individuals. European Journal of Neuroscience, 35(10), 1630–1636. https://doi.org/10.1111/j.1460-9568.2012.08088.x

Lemche, E., Giampietro, V. P., Surguladze, S. A., Amaro, E. J., Andrew, C. M., Williams, S. C. R., . . . Phillips, M. L. (2006). Human attachment security is mediated by the amygdala: Evidence from combined fMRI and psychophysiological measures. Human Brain Mapping, 27(8), 623–635. https://doi.org/10.1002/hbm.20206

Levy, K. N., Meehan, K. B., Weber, M., Reynoso, J., & Clarkin, J. F. (2005). Attachment and Borderline Personality Disorder: Implications for Psychotherapy. Psychopathology, 38(2), 64–74. https://doi.org/10.1159/000084813

Lopez, F. G., & Gormley, B. (2002). Stability and change in adult attachment style over the first-year college transition: Relations to self-confidence coping, and distress patterns.. Journal of Counseling Psychology, 49(3), 355–364. https://doi.org/10.1037/0022-0167.49.3.355

Lorber, M. F., & Egeland, B. (2009). Infancy parenting and externalizing psychopathology from childhood through adulthood: Developmental trends. Developmental Psychology, 45(4), 909–912. https://doi.org/10.1037/a0015675

Lucas, G. M., Gratch, J., Scherer, S., Boberg, J., & Stratou, G. (2015). Towards an affective interface for assessment of psychological distress. In 2015 International Conference on Affective Computing and Intelligent Interaction (ACII). IEEE. https://doi.org/10.1109/acii.2015.7344622

MacLeod, M. (2016). What makes interdisciplinarity difficult? Some consequences of domain specificity in interdisciplinary practice. Synthese, 195(2), 697–720. https://doi.org/10.1007/s11229-016-1236-4

Marchewka, A., Łukasz Żurawski, Jednoróg, K., & Grabowska, A. (2013). The Nencki Affective Picture System (NAPS): Introduction to a novel standardized, wide-range, high-quality, realistic picture database. Behavior Research Methods, 46(2), 596–610. https://doi.org/10.3758/s13428-013-0379-1

Mashin, V. A. (2007). Nonstationarity and duration of the cardiac interval time series in assessing the functional state of operator personnel. Biophysics, 52(2), 241–247. https://doi.org/10.1134/s0006350907020170

Mathôt, S., Schreij, D., & Theeuwes, J. (2011). OpenSesame: An open-source graphical experiment builder for the social sciences. Behavior Research Methods, 44(2), 314–324. https://doi.org/10.3758/s13428-011-0168-7

Maunder, R. G., Lancee, W. J., Nolan, R. P., Hunter, J. J., & Tannenbaum, D. W. (2006). The relationship of attachment insecurity to subjective stress and autonomic function during standardized acute stress in healthy adults. Journal of Psychosomatic Research, 60(3), 283–290. https://doi.org/10.1016/j.jpsychores.2005.08.013

Maxwell, S. E., Lau, M. Y., & Howard, G. S. (2015). Is psychology suffering from a replication crisis? What does "failure to replicate" really mean?. American Psychologist, 70(6), 487–498. https://doi.org/10.1037/a0039400

McCarthy, C., Pradhan, N., Redpath, C., & Adler, A. (2016). Validation of the Empatica E4 wristband. In 2016 IEEE EMBS International Student Conference (ISC). IEEE. https://doi.org/10.1109/embsisc.2016.7508621

McCorry, L. K. (2007). Physiology of the Autonomic Nervous System. American Journal of Pharmaceutical Education, 71(4), 78. https://doi.org/10.5688/aj710478

McDuff, D. J., Estepp, J. R., Piasecki, A. M., & Blackford, E. B. (2015). A survey of remote optical photoplethysmographic imaging methods. In 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE. https://doi.org/10.1109/embc.2015.7319857

Mesman, J., van IJzendoorn, M. H., & Sagi-Schwartz, A. (2016). Crosscultural patterns of attachment. In Handbook of Attachment, Third Edition: Theory, Research, and Clinical Applications. (pp. 852–877). Guilford.

Mikulincer, M., Dolev, T., & Shaver, P. R. (2004). Attachment-Related Strategies During Thought Suppression: Ironic Rebounds and Vulnerable Self-Representations. Journal of Personality and Social Psychology, 87(6), 940–956. https://doi.org/10.1037/0022-3514.87.6.940

Mikulincer, M., & Shaver, P. R. (2007). Attachment in adulthood: structure, dynamics, and change. Guilford.

Mikulincer, M., & Shaver, P. R. (2015). The psychological effects of the contextual activation of security-enhancing mental representations in adulthood. Current Opinion in Psychology, 1, 18–21. https://doi.org/10.1016/j.copsyc.2015.01.008

Mikulincer, M., & Shaver, P. R. (2012). An attachment perspective on psychopathology. World Psychiatry, 11(1), 11–15. https://doi.org/10.1016/j.wpsyc.2012.01.003

Miljkovitch, R., Moran, G., Roy, C., Jaunin, L., Forcada-Guex, M., Pierrehumbert, B., … Borghini, A. (2013). Maternal interactive behaviour as a predictor of preschoolers' attachment representations among full term and premature samples. Early Human Development, 89(5), 349–354. https://doi.org/10.1016/j.earlhumdev.2012.11.006

Miljkovitch, R., Moss, E., Bernier, A., Pascuzzo, K., & Sander, E. (2015). Refining the assessment of internal working models: the Attachment Multiple Model Interview. Attachment & Human Development, 17(5), 492–521. https://doi.org/10.1080/14616734.2015.1075561

Muñoz, A., Moreno, C., & Luján, J. L. (2010). Who is willing to pay for science? On the relationship between public perception of science and the attitude to public funding of science. Public Understanding of Science, 21(2), 242–253. https://doi.org/10.1177/0963662510373813

Nagasawa, M., Mitsui, S., En, S., Ohtani, N., Ohta, M., Sakuma, Y., … Kikusui, T. (2015). Oxytocin-Gaze Positive Loop and the Coevolution of Human-Dog Bonds. Obstetrical & Gynecological Survey, 70(7), 450–451. https://doi.org/10.1097/01.ogx.0000469196.99143.92

Niedenthal, P. M., Brauer, M., Robin, L., & Åse H. Innes-Ker. (2002). Adult attachment and the perception of facial expression of emotion. Journal of Personality and Social Psychology, 82(3), 419–433. https://doi.org/10.1037/0022-3514.82.3.419

Novak, D., Mihelj, M., & Munih, M. (2012). A survey of methods for data fusion and system adaptation using autonomic nervous system responses in physiological computing. Interacting with Computers, 24(3), 154–172. https://doi.org/10.1016/j.intcom.2012.04.003

Parnell, L., & Siegel, D. J. (2015). Attachment-focused EMDR: Healing relational trauma. W. W. Norton & Company.

Parra, F., George, C., Kalalou, K., & Januel, D. (2017). Ideal Parent Figure method in the treatment of complex posttraumatic stress disorder related to childhood trauma: a pilot study. European Journal of Psychotraumatology, 8(1), 1400879. https://doi.org/10.1080/20008198.2017.1400879

Peltola, M. J., Forssman, L., Puura, K., van IJzendoorn, M. H., & Leppänen, J. M. (2015). Attention to Faces Expressing Negative Emotion at 7 Months Predicts Attachment Security at 14 Months. Child Development, 86(5), 1321–1332. https://doi.org/10.1111/cdev.12380

Pestian, J. P., Sorter, M., Connolly, B., Cohen, K. B., McCullumsmith, C., Gee, J. T., . . . and, L. R. (2016). A Machine Learning Approach to Identifying the Thought Markers of Suicidal Subjects: A Prospective Multicenter Trial. Suicide and Life-Threatening Behavior, 47(1), 112–121. https://doi.org/10.1111/sltb.12312

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies.. Journal of Applied Psychology, 88(5), 879–903. https://doi.org/10.1037/0021-9010.88.5.879

Poria, S., Cambria, E., Bajpai, R., & Hussain, A. (2017). A review of affective computing: From unimodal analysis to multimodal fusion. Information Fusion, 37, 98–125. https://doi.org/10.1016/j.inffus.2017.02.003

Powers, S. I., Pietromonaco, P. R., Gunlicks, M., & Sayer, A. (2006). Dating couples' attachment styles and patterns of cortisol reactivity and recovery in response to a relationship conflict. Journal of Personality and Social Psychology, 90(4), 613–628. https://doi.org/10.1037/0022-3514.90.4.613

Rafols, I., Leydesdorff, L., O'Hare, A., Nightingale, P., & Stirling, A. (2012). How journal rankings can suppress interdisciplinary research. A comparison between Innovation Studies and Business & Management. Research Policy, 41(7), 1262–1282. Retrieved from https://arxiv.org/abs/1105.1227

Ravitz, P., Maunder, R., Hunter, J., Sthankiya, B., & Lancee, W. (2010). Adult attachment measures: A 25-year review. Journal of Psychosomatic Research, 69(4), 419–432. https://doi.org/10.1016/j.jpsychores.2009.08.006

Read, J., Os, J., Morrison, A. P., & Ross, C. A. (2005). Childhood trauma psychosis and schizophrenia: a literature review with theoretical and clinical implications. Acta

Psychiatrica Scandinavica, 112(5), 330–350. https://doi.org/10.1111/j.1600-0447.2005.00634.x

Robson, K. S. (1967). The Role Of Eye To Eye Contact In Maternal Infant Attachment. Journal of Child Psychology and Psychiatry, 8(1), 13–25. https://doi.org/10.1111/j.1469-7610.1967.tb02176.x

Rodin, G., Walsh, A., Zimmermann, C., Gagliese, L., Jones, J., Shepherd, F. A., ... Mikulincer, M. (2007). The contribution of attachment security and social support to depressive symptoms in patients with metastatic cancer. Psycho-Oncology, 16(12), 1080–1091. https://doi.org/10.1002/pon.1186

Roisman, G. I. (2007). The psychophysiology of adult attachment relationships: Autonomic reactivity in marital and premarital interactions. Developmental Psychology, 43(1), 39–53. https://doi.org/10.1037/0012-1649.43.1.39

Roisman, G. I., Holland, A., Fortuna, K., Fraley, R. C., Clausell, E., & Clarke, A. (2007). The Adult Attachment Interview and self-reports of attachment style: An empirical rapprochement.. Journal of Personality and Social Psychology, 92(4), 678–697. https://doi.org/10.1037/0022-3514.92.4.678

Roisman, G. I., Tsai, J. L., & Chiang, K.-H. S. (2004). The Emotional Integration of Childhood Experience: Physiological Facial Expressive, and Self-Reported Emotional Response During the Adult Attachment Interview.. Developmental Psychology, 40(5), 776–789. https://doi.org/10.1037/0012-1649.40.5.776

Rosenberg, M. (1965). Society and the Adolescent Self-Image. Princeton University Press. https://doi.org/10.1515/9781400876136

Sakakibara, M., Takeuchi, S., & Hayano, J. (1994). Effect of relaxation training on cardiac parasympathetic tone. Psychophysiology, 31(3), 223–228. https://doi.org/10.1111/j.1469-8986.1994.tb02210.x

Sbarra, D. A., & Borelli, J. L. (2013). Heart rate variability moderates the association between attachment avoidance and self-concept reorganization following marital separation. International Journal of Psychophysiology, 88(3), 253–260. https://doi.org/10.1016/j.ijpsycho.2012.04.004

Scharfe, E., & Bartholomew, K. (1994). Reliability and stability of adult attachment patterns. Personal Relationships, 1(1), 23–43. https://doi.org/10.1111/j.1475-6811.1994.tb00053.x

Scherer, S., Gale, L., Gratch, J., Rizzo, A., & Morency, L. P. (2016). Self-Reported Symptoms of Depression and PTSD Are Associated with Reduced Vowel Space in Screening Interviews. IEEE Transactions on Affective Computing, 7(1), 59–73. https://doi.org/10.1109/taffc.2015.2440264

Scherer, S., Morency, L.-P., Gratch, J., & Pestian, J. (2015). Reduced vowel space is a robust indicator of psychological distress: A cross-corpus analysis. In 2015 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP). IEEE. https://doi.org/10.1109/icassp.2015.7178880

Scherer, S., Stratou, G., Mahmoud, M., Boberg, J., Gratch, J., Rizzo, A., & Morency, L.-P. (2013). Automatic behavior descriptors for psychological disorder analysis. In 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG). IEEE. https://doi.org/10.1109/fg.2013.6553789

Schultz, D. S., & Loving, J. L. (2012). Challenges Since Wikipedia: The Availability of Rorschach Information Online and Internet Users' Reactions to Online Media Coverage of the Rorschach–Wikipedia Debate. Journal of Personality Assessment, 94(1), 73–81. https://doi.org/10.1080/00223891.2011.627963

Schäfer, A., & Vagedes, J. (2013). How accurate is pulse rate variability as an estimate of heart rate variability?. International Journal of Cardiology, 166(1), 15–29. https://doi.org/10.1016/j.ijcard.2012.03.119

Shaffer, F., McCraty, R., & Zerr, C. L. (2014). A healthy heart is not a metronome: an integrative review of the heart's anatomy and heart rate variability. Frontiers in Psychology, 5. https://doi.org/10.3389/fpsyg.2014.01040

Sharma, N., & Gedeon, T. (2012). Objective measures sensors and computational techniques for stress recognition and classification: A survey. Computer Methods and Programs in Biomedicine, 108(3), 1287–1301. https://doi.org/10.1016/j.cmpb.2012.07.003

Simpson, J. A., Rholes, W. S., & Phillips, D. (1996). Conflict in close relationships: An attachment perspective. Journal of Personality and Social Psychology, 71(5), 899–914. https://doi.org/10.1037/0022-3514.71.5.899

Solomon, J., & George, C. (2011). The disorganized attachment-caregiving system: Dysregulation of adaptive processes at multiple levels. In Attachment disorganization and caregiving (pp. 3–24). Guilford Press.

Sroufe, L. A., Coffino, B., & Carlson, E. A. (2010). Conceptualizing the Role of Early Experience: Lessons from the Minnesota Longitudinal Study. Dev Rev, 30, 36–51.

Sroufe, L. A., & Rutter, M. (1984). The Domain of Developmental Psychopathology.

Child Development, 55(1), 17. https://doi.org/10.2307/1129832

Sroufe, L. A., & Waters, E. (2017). Attachment as an Organizational Construct. In Interpersonal Development (pp. 109–124). Routledge. https://doi.org/10.4324/9781351153683-5

Stovall-McClough, C., & Dozier, M. (2016). Attachment states of mind and psychopathology in adulthood. In Handbook of attachment. Theory, research, and clinical applications (3rd ed.) (pp. 715–738). Guilford Press.

Sun, Y., & Thakor, N. (2016). Photoplethysmography Revisited: From Contact to Noncontact From Point to Imaging. IEEE Transactions on Biomedical Engineering, 63(3), 463–477. https://doi.org/10.1109/tbme.2015.2476337

Supriya, V., Baltrušaitis, T., Pennant, L., Ongür, D., Baker, J. T., & Morency, L.-P. (2016). Computational study of psychosis symptoms and facial expressions. In Computing and Mental Health Workshop at Computer and Human Interaction.

Talia, A., Miller-Bottome, M., & Daniel, S. I. F. (2015). Assessing Attachment in Psychotherapy: Validation of the Patient Attachment Coding System (PACS). Clinical Psychology & Psychotherapy, 24(1), 149–161. https://doi.org/10.1002/cpp.1990

Talia, A., Miller-Bottome, M., Wyner, R., Lilliengren, P., & Bate, J. (2019). Patients' Adult Attachment Interview classification and their experience of the therapeutic relationship: are they associated?. Research in Psychotherapy: Psychopathology Process and Outcome, 22(2). https://doi.org/10.4081/ripppo.2019.361

Tausczik, Y. R., & Pennebaker, J. W. (2009). The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. Journal of Language and Social Psychology, 29(1), 24–54. https://doi.org/10.1177/0261927x09351676

Terathongkum, S., & Pickler, R. H. (2004). Relationships among heart rate variability hypertension, and relaxation techniques. Journal of Vascular Nursing, 22(3), 78–82. https://doi.org/10.1016/j.jvn.2004.06.003

Toppelberg, C. O., & Shapiro, T. (2000). Language Disorders: A 10-Year Research Update Review. Journal of the American Academy of Child & Adolescent Psychiatry, 39(2), 143–152. https://doi.org/10.1097/00004583-200002000-00011

Trzesniewski, K. H., Donnellan, M. B., Moffitt, T. E., Robins, R. W., Poulton, R., & Caspi, A. (2006). Low self-esteem during adolescence predicts poor health criminal behavior, and limited economic prospects during adulthood.. Developmental Psychology, 42(2), 381–390. https://doi.org/10.1037/0012-1649.42.2.381

Tuber, S. (2014). Understanding personality through projective testing. Rowman & Littlefield Publishers.

Tucker, J. S., & Anders, S. L. (1998). Adult Attachment Style and Nonverbal Closeness in Dating Couples. Journal of Nonverbal Behavior, 22(2), 109–124. https://doi.org/10.1023/a:1022980231204

Wallace, J. L., & Vaux, A. (1993). Social Support Network Orientation: The Role of Adult Attachment Style. Journal of Social and Clinical Psychology, 12(3), 354–365. https://doi.org/10.1521/jscp.1993.12.3.354

Waters, E., Hamilton, C. E., & Weinfield, N. S. (2000). The Stability of Attachment Security from Infancy to Adolescence and Early Adulthood: General Introduction. Child Development, 71(3), 678–683. https://doi.org/10.1111/1467-8624.00175

Wei, M., Russell, D. W., Mallinckrodt, B., & Vogel, D. L. (2007). The Experiences in Close Relationship Scale (ECR)-Short Form: Reliability Validity, and Factor Structure. Journal of Personality Assessment, 88(2), 187–204. https://doi.org/10.1080/00223890701268041

Wiener, J. B. (2004). The regulation of technology and the technology of regulation. Technology in Society, 26(2-3), 483–500. https://doi.org/10.1016/j.techsoc.2004.01.033

Wörtwein, T., Baltrušaitis, T., Laksana, E., Pennant, L., Liebson, E. S., Öngür, D., . . . Morency, L.-P. (2017). Computational Analysis of Acoustic Descriptors in Psychotic Patients. In Interspeech 2017. ISCA. https://doi.org/10.21437/interspeech.2017-466

Zeng, Z., Hu, Y., Fu, Y., Huang, T. S., Roisman, G. I., & Wen, Z. (2006). Audio-visual emotion recognition in adult attachment interview. In Proceedings of the 8th international conference on Multimodal interfaces - ICMI '06. ACM Press. https://doi.org/10.1145/1180995.1181028

Zillmann, D. (1995). Mechanisms of emotional involvement with drama. Poetics, 23(1-2), 33–51. https://doi.org/10.1016/0304-422x(94)00020-7

Ziman, K., Heusser, A. C., Fitzpatrick, P. C., Field, C. E., & Manning, J. R. (2018). Is automatic speech-to-text transcription ready for use in psychological experiments?. Behavior Research Methods, 50(6), 2597–2605. https://doi.org/10.3758/s13428-018-1037-4

de Wolff, M. S., & van IJzendoorn, M. H. (1997). Sensitivity and Attachment: A Meta-Analysis on Parental Antecedents of Infant Attachment. Child Development, 68(4), 571. https://doi.org/10.2307/1132107

del Paso, G. A. R., Langewitz, W., Mulder, L. J. M., van Roon, A., & Duschek, S. (2013). The utility of low frequency heart rate variability as an index of sympathetic cardiac tone:

A review with emphasis on a reanalysis of previous studies. Psychophysiology, 50(5), 477–487. https://doi.org/10.1111/psyp.12027

van de Mortel, T. (2008). Faking It: Social Desirability Response Bias in Self-report Research. Australian Journal of Advanced Nursing, Volume 25 Issue 4.

Task Force of the European Society of Electrophysiology. (1996). Heart Rate Variability. Circulation, 93(5), 1043–1065. https://doi.org/10.1161/01.cir.93.5.1043