HEC MONTRÉAL

Essais sur l'évaluation de l'expérience psychophysiologique des utilisateurs dans un contexte bancaire

par

Marianne Veilleux

Sciences de la gestion
(Option Marketing)

Sous la codirection de

Sylvain Sénécal, Ph. D. et Pierre-Majorique Léger, Ph. D.

Mémoire présenté en vue de l'obtention
du grade de maîtrise ès sciences en gestion
(M. Sc.)

Mars 2020

Certification d'approbation éthique et Attestation d'approbation éthique complétée

HEC MONTRĒAL

Comité d'éthique de la recherche

CERTIFICAT D'APPROBATION ÉTHIQUE

La présente atteste que le projet de recherche décrit ci-dessous a fait l'objet d'une évaluation en matière d'éthique de la recherche avec des êtres humains et qu'il satisfait aux exigences de notre politique en cette matière.

Projet # : 2019-3419

Titre du projet de recherche : Testing the convergent validity of stimulated recall : An Exploratory Study

Chercheur principal :
Pierre-Majorique Léger
Professeur titulaire, Technologies de l'information - HEC Montréal

Cochercheurs :
David Brieugne; Marianne Veilleux; Marc Fredette; Sylvain Sénécal; Shang-Lin Chen; Bertrand Demolin; Frédérique Bouvier

Date d'approbation du projet : 21 mars 2019

Date d'entrée en vigueur du certificat : 21 mars 2019

Date d'échéance du certificat : 01 mars 2020

Maurice Lemelin
Président
CER de HEC Montréal

NAGANO Approbation du projet par le comité d'éthique suite à l'approbation conditionnelle
Comité d'éthique de la recherche - HEC Montréal 2/2

HEC MONTRĒAL

Comité d'éthique de la recherche

ATTESTATION D'APPROBATION ÉTHIQUE COMPLÉTÉE

La présente atteste que le projet de recherche décrit ci-dessous a fait l'objet des approbations en matière d'éthique de la recherche avec des êtres humains nécessaires selon les exigences de HEC Montréal.

La période de validité du certificat d'approbation éthique émis pour ce projet est maintenant terminée. Si vous devez reprendre contact avec les participants ou reprendre une collecte de données pour ce projet, la certification éthique doit être réactivée préalablement. Vous devez alors prendre contact avec le secrétariat du CER de HEC Montréal.

Nom de l'étudiant(e) : Marianne Veilleux

Titre du projet supervisé/mémoire/thèse :
L'évaluation d'une experience psychophysiologique avec et sans l'aide de l'utilisateur dans un context bancaire

Titre du projet sur le certificat :
Testing the convergent validity of stimulated recall : An Exploratory Study

Projet # : 2019-3419

Chercheur principal / directeur de recherche :
Pierre-Majorique Léger

Cochercheurs :
David Brieugne; Marianne Veilleux; Marc Fredette; Sylvain Sénécal; Shang-Lin Chen; Bertrand Demolin; Frédérique Bouvier

Date d'approbation initiale du projet : 21 mars 2019

Date de fermeture de l'approbation éthique pour l'étudiant(e) : 25 octobre 2019

Maurice Lemelin
Président
CER de HEC Montréal

NAGANO Document officiel du CER
Comité d'éthique de la recherche - HEC Montréal 1 / 1

Résumé

Dans ce mémoire par articles, le processus d'évaluation de l'expérience cognitive et émotionnelle de l'utilisateur est étudié dans un contexte de test utilisateur. Plus précisément, ce mémoire présente deux essais sur la mobilisation des données physiologiques ainsi que leur interprétabilité à travers le processus d'évaluation pour la prise de décision en matière de design d'interface.

Dans le premier essai, l'importance de la prise en compte de la perspective de l'utilisateur à travers le processus d'évaluation est étudiée. Il porte sur la qualité de l'interprétation de l'état cognitif et émotionnel de l'utilisateur, mesurée par des outils physiologiques. La capacité de l'utilisateur à évaluer rétrospectivement ses états affectifs est comparée à celle d'un groupe d'experts en UX et vise à répondre à la question suivante : Lorsque les utilisateurs interprètent eux-même rétrospectivement leur expérience en ligne, sont-ils plus précis que les experts en UX? Les résultats du premier essai suggèrent que les experts en UX, caractérisés par des individus n'étant pas engagés émotionnellement à l'expérience, évaluent plus précisément l'émotion vécue par l'utilisateur que l'utilisateur lui-même.

Ensuite, la deuxième question porte sur l'interprétabilité des données physiologiques par des experts en UX. Dans cet essai, la visualisation des données physiologiques est étudiée afin de permettre aux experts de comprendre l'expérience vécue par l'utilisateur. Par conséquent, cet essai vise à répondre à la question suivante : Comment présenter visuellement l'évolution de l'expérience vécue par les utilisateurs pour faciliter la compréhension de leur navigation ? Pour y répondre, l'essai présente et teste une nouvelle méthode de visualisation combinant la charge cognitive, la valence émotionnelle et le parcours client afin de schématiser l'expérience vécue par les utilisateurs au fil de leur navigation.

Collectivement, ces deux essais contribuent à la littérature en UX en analysant l'importance d'intégrer la perspective de l'utilisateur lors du processus d'évaluation. Les résultats présentent aussi des implications pratiques en proposant un schéma de visualisation facilitant l'identification de zones problématique sur une interface donnée.

Mots clés : Émotions, activation, charge cognitive, utilisateur, expert, évaluation

Table des matières

Liste des tableaux et des figures

Liste des figures

Premier essai

Deuxième essai

Liste des tableaux

Premier essai

Deuxième essai

Liste des abréviations

UX : « User Experience »

UX : Expérience utilisateur

HCI : « human-computer interaction »

CRD : « Cued-Recall Debrief »

CJM : « Customer Journey Mapping »

SEBE : « Subjective Evidence-Based Ethnography »

SCM : « Self-Confrontation Method »

SR : « Stimulated Recall »

Avant-propos

Ce mémoire a été rédigé sous la forme de deux essais complémentaires à la suite de l'approbation de la direction du programme de la Maîtrise ès sciences en gestion de l'option Marketing. Le premier essai compare la qualité d'interprétation entre un utilisateur et un expert UX dans la compréhension de l'expérience vécue par l'utilisateur. Cet article sera soumis à la revue Behaviour and Information Technology. Le second essai propose une nouvelle méthode afin de visualiser les données physiologiques et permettre aux experts en UX de comprendre l'expérience vécue par l'ensemble des utilisateurs. La méthode proposée permet de visualiser les hauts et les bas de l'expérience vécue. Ce deuxième essai est complémentaire au premier puisqu'il offre une alternative à la compréhension de l'expérience sans la participation de l'utilisateur dans le processus d'analyse de l'expérience. Cet article a été accepté à la conférence HCI International 2020, qui se déroulera à Copenhague (Danemark) en juillet 2020. L'accord des co-auteurs de tous les articles a été obtenu afin que ceux-ci puissent être présentés dans ce mémoire. Le comité d'éthique de la recherche de HEC Montréal a donné son approbation pour cette recherche le 21 mars 2019.

Remerciements

Ce projet n’aurait jamais été aussi enrichissant sans quelques personnes que je tiens à remercier infiniment.

J’aimerais tout d’abord remercier Pierre-Majorique Léger et Sylvain Sénécal qui m’ont offert l’opportunité de travailler au sein de la famille du Tech3lab. Merci pour votre patience et votre accompagnement tout au long de ce projet de longue haleine. Vous avez eu confiance en moi et je crois profondément que l’accomplissement de ce mémoire n’aurait pas été aussi enrichissant sans le support constant que vous m’avez offert.

Un énorme merci à l’équipe du laboratoire, dont Shang Lin, David, Bertrand et les assistantes de recherche pour le support lors de la collecte. Un merci particulier à Frédérique sans qui cette collecte aurait été longue et périlleuse.

Merci à tous mes amis, camarades et collègues, Audrey, Charles, Elisabeth, Noémie, Selma, Tanya, Sarah et Ariane de m’avoir supportée tout au long du processus. Merci également à Caroline et Ariane de m’avoir guidée dans ce début de rédaction qui me paraissait insurmontable.

Je tiens également à remercier ma famille d'avoir toujours été là à m’écouter et à m’encourager même lorsque je n’étais plus sûre de moi. Merci à mon père et à ma mère pour les nombreuses relectures qui m’ont été très précieuses et à Flo pour les conseils de rédaction.

Un dernier remerciement, et non des moindres, à Patrik pour avoir écouté toutes mes péripéties et mes insécurités sans perdre une confiance aveugle en mes capacités.

Pour finir, je voudrais remercier le Conseil de Recherches en Sciences Naturelles et Génie (CRSNG) et Prompt pour leur contribution financière me permettant de réaliser cette recherche dans les meilleures conditions.

Introduction

Problématique générale de l'étude

Les technologies gagnent en importance dans notre vie de tous les jours. Avec le web à portée de main, les consommateurs sont davantage susceptibles d'avoir une forte présence en ligne lors de leur magasinage. En effet, les ventes au détail en ligne ont augmenté de 21,1% au Canada en 2019 (eMarketer 2019). Cette tendance confirme l'impact des technologies sur nos comportements d'achats. Conséquemment, le virage des entreprises au numérique devient un phénomène courant. Avec un ratio de 89% des entreprises ayant l'intention ou ayant déjà adopté une stratégie d'affaire numérique (IDG, 2018), l'expérience client en ligne devient une priorité. L'expérience client consiste à placer les besoins du consommateur au centre des prises de décision de l'organisation et à s'assurer que tous les points de contact créés par l'entreprise permettent au consommateur de vivre une expérience sans faille (McKinsey, 2016). Tandis que l'expérience client réfère à l'étude de l'interaction avec chacun de ces points de contact, l'expérience utilisateur (UX) est une pratique de l'expérience client qui se caractérise par l'étude de l'interaction avec une interface. L'expérience utilisateur gagne en importance, car les points de contact humains avec la marque sont moins fréquents lors d'un achat en ligne. La perception de l'entreprise se reflète donc principalement via l'expérience vécue en ligne par l'utilisateur. Selon un rapport de Gomez (2010), 88% des utilisateurs en ligne ne retournent pas sur un site internet à la suite d'une mauvaise expérience. Par conséquent, une plus grande attention est accordée à l'optimisation de l'expérience en ligne. Sachant qu'un consommateur loyal peut être jusqu'à cinq fois plus profitable qu'un consommateur nouvellement acquit (Gupta & Kim, 2007), une expérience client positive est donc souhaitable.

Par ailleurs, selon une étude par Adobe et EConsultancy (2019) auprès de plus de 12 000 professionnels des domaines de la publicité, du commerce électronique, des technologies de l'information et du marketing, près de 50% des répondants affirment que leur organisation planifie augmenter les dépenses en expérience client pour 2019. Bien que l'expérience client soit reconnue comme étant un facteur important de succès, les organisations semblent avoir du mal à traduire ce concept de manière opérationnelle (Adobe & EConsultancy, 2019). En effet, seulement une organisation sur dix se catégorise comme étant à un stade très avancé en terme d'expérience client tandis que la majorité (52%) se positionne présentement comme étant peu ou pas avancé (Adobe & EConsultancy, 2019). Par ailleurs, les organisations ayant su comment opérationnaliser l'expérience client et se trouvant à un stade avancé sont trois fois plus susceptibles d'avoir dépassé leurs objectifs d'affaires pour 2018 (Adobe & EConsultancy, 2019).

Par conséquent, dû à l'importance grandissante d'une bonne mise en application de l'expérience client, il est intéressant d'explorer les différentes approches permettant de mesurer l'expérience utilisateur auprès des consommateurs. Plus précisément, cette étude s'intéresse à la mobilisation des données physiologiques ainsi que leur interprétabilité à travers le processus d'évaluation de l'expérience client dans le contexte de test utilisateur.

Problématiques spécifiques de l'étude

Une multitude d'approches et de méthodes se développent au sein de la littérature dans le but de trouver une méthode optimale permettant de mesurer l'expérience utilisateur (Bagdare & Jain, 2013; Lemon & Verhoef, 2016; Klaus & Maklan, 2011). L'expérience utilisateur vise à étudier l'interaction entre le consommateur et une interface (Vermeeren et al., 2015). Ces techniques visent autant à mesurer la perception de l'utilisateur envers son expérience globale que de comprendre une sélection de moments précis vécut par l'utilisateur. Certains auteurs prônent l'approche qualitative et favorisent l'échange avec l'utilisateur comme source principale d'information (Miles & Huberman, 1994; Ambert et al., 1995; Kaplan & Maxwell, 2005; Kelly, 2009). Les objectifs de la recherche qualitative visent la compréhension d'un événement du point de vue de l'utilisateur. Selon Kaplan & Maxwell (2005), l'utilisateur est en mesure de fournir des informations ne pouvant être égalées par des questionnaires et des mesures standardisés. Par ailleurs, une approche rétrospective qualitative a été développée au sein de la littérature dans le but de favoriser l'échange avec l'utilisateur sans toutefois interrompre l'utilisateur lors de son expérience. Cette méthode est le rappel stimulé, qui réfère à un ensemble de méthodes de recherche introspective étudiant les processus cognitifs. La principale caractéristique de ces différentes méthodes de recherche est que les participants sont invités à se remémorer, à l'aide de séquences vidéos montrant leurs comportements, leurs pensées simultanées durant ces événements (Lyle, 2003). La vidéo est du point de vue de la première personne et vise à favoriser la rétroaction de l'utilisateur ainsi que le souvenir d'une plus grande quantité d'informations. La méthode du rappel stimulé a été largement utilisée dans le milieu de la psychologie (Lyle, 2003; Bentley, 2005; Lahlou, 2011) et est défendue comme favorisant un meilleur rappel de l'expérience vécue par l'utilisateur. Selon une étude par De Grave et al. (1996), le fait que la procédure de rappel suit immédiatement l'expérience empêche les participants de "quitter" les processus cognitifs impliqués dans l'analyse de l'expérience vécue. Le but de cette méthode est de maximiser l'utilisation de la mémoire à court terme de l'individu (Lyle, 2003). Bref, la méthode du rappel stimulé devient concurrente aux autres méthodes qualitatives en simultanée – telle que la méthode "think-aloud" – qui sont critiquées pour être intrusives et pour potentiellement influencer les processus cognitifs de l'utilisateur (Bentley et al., 2005; Schooler, 2011).

Cependant, d'autres auteurs défendent que l'utilisateur est sujet à des biais cognitifs et n'est donc pas en mesure de se rappeler correctement son expérience (Fredrickson & Kahneman, 1993; Schooler, 2011; Guinea et al., 2014; Cockburn et al., 2017). Par conséquent, se fier principalement à l'utilisateur comme source d'information mène possiblement à des conclusions erronées. Plusieurs biais ont été identifiés à travers la littérature comme pouvant influencer la capacité de l'utilisateur à se souvenir rétrospectivement de son expérience. Notamment, le concept de pic émotionnel réfère au moment le plus intense d'une expérience et représente le moment dont l'utilisateur se souvient le plus fréquemment (Kahneman et al., 1993; Redelmeier & Kahneman, 1996). Aussi, les principes de premières et dernières impressions expliquent que l'utilisateur est plus susceptible de se souvenir davantage de la séquence du début et de la fin d'une expérience plutôt que les séquences du milieu (Cockburn et al., 2017; Lourties et al., 2018). Afin d'éviter ces biais, les mesures psychophysiologiques sont utilisées afin d'avoir un compte-rendu en temps réel de l'état émotionnel et cognitif de l'utilisateur sans toutefois l'interrompre durant son expérience. De cette manière, les mesures psychophysiologiques permettent d'éviter les biais rétrospectifs reliés à la mémoire de l'utilisateur. Finalement, ces mesures permettent donc d'avoir un compte-rendu complet de l'état émotionnel et cognitif d'un utilisateur lors de son interaction avec une interface (Picard, 1995; Ward & Marsden, 2003; Bentley et al., 2005). Une étude par Giroux-Huppé et al. (2019) suggère que lorsque questionné sur leur expérience, les utilisateurs ne se souviennent pas de 75% des points de friction vécus. Les points de friction font ici référence à des moments caractérisés par une émotion négative vécue et une intensité émotionnelle élevée. Ces résultats démontrent donc l'utilité des données psychophysiologiques puisqu'elles permettent d'avoir des données quantifiables en temps réel sur l'expérience émotionnelle de l'utilisateur sans toutefois être sujet aux nombreux biais rétrospectifs de l'utilisateur.

Plusieurs approches se positionnent comme étant la manière optimale de mesurer ce concept. Au cours des dernières années, les méthodes se multiplient et ce phénomène devient une source de questionnements quant à la méthode optimale permettant d'évaluer l'expérience utilisateur. Ces différences au sein de la littérature nous amène à approfondir le sujet dans le cadre d'une étude méthodologique permettant d'étudier les données psychophysiologiques et la performance de l'utilisateur en comparaison à un groupe d'experts en UX lors de l'évaluation de l'expérience utilisateur. Plus précisément, nous cherchons à comparer les mesures psychophysiologiques aux informations provenant du rappel stimulé afin de mettre en lumière l'importance d'inclure la perspective de l'utilisateur dans la découverte d'insights.

Dans ce mémoire, deux approches principales sont étudiées et croisées afin de mettre en lumière leur pertinence et leur efficacité. Tout d'abord, la méthode du rappel stimulé, approche qualitative utilisée afin de faciliter la rétroaction de l'utilisateur suite à son expérience. Ensuite, les mesures psychophysiologiques, approche quantitative gagnant en popularité, car ces mesures permettent de suivre en temps réel l'état émotionnel et cognitif de l'utilisateur sans toutefois être interrompue lors de son expérience. Les informations captées grâce à ces deux méthodes sont ensuite croisées afin de comparer la qualité de l'interprétation de l'utilisateur. La performance des utilisateurs est ensuite comparée à la performance d'un groupe d'experts en UX afin de mettre en lumière la qualité de leur interprétation respective. L'utilisation d'observateurs externes comme les experts est fréquemment utilisée à titre de comparatif quant à la perception d'une expérience vécue (Harrigan et al., 2005). Bref, ces deux approches ainsi que l'utilisation d'experts en UX représentent les principaux concepts abordés dans ce mémoire.

Questions de recherche

Alors que plusieurs approches prônent différentes méthodes d'analyse de l'expérience utilisateur, il est souhaitable de comprendre l'importance de tenir compte de la perspective de l'utilisateur dans le processus d'évaluation de l'expérience vécue par ce dernier. L'expérience vécue fait ici référence aux états émotionnels et cognitifs qu'à expérimenté l'utilisateur lors de l'étude. En effet, certaines approches au sein de la littérature défendent l'inclusion de l'utilisateur alors que d'autres critiquent les biais pouvant influencer sa capacité à se souvenir rétrospectivement de son expérience, tels les pics émotionnels. La compréhension du rôle de l'utilisateur dans le processus d'évaluation de son expérience permet donc de mettre en lumière son utilité.

Ce mémoire par article explore la mobilisation des données physiologiques et leur interprétabilité dans le processus d'évaluation de l'expérience vécue par l'utilisateur pour la prise de décision en matière de design d'interface. Ce mémoire se présente sous forme de deux essais. Le premier compare la précision d'interprétation entre l'utilisateur et un groupe d'experts en UX afin de répondre à la question de recherche suivante : Lorsque les utilisateurs interprètent eux-même rétrospectivement leur expérience en ligne, sont-ils plus précis que les experts en UX? Cette question de rechercher permet d'étudier la précision à laquelle l'utilisateur est en mesure d'évaluer rétrospectivement l'expérience vécue dans un contexte de test utilisateur. Les évaluations a posteriori sont croisées avec les mesure psychophysiologique de l'utilisateur et sont par la suite comparées à un groupe d'experts en UX afin d'évaluer leurs performances respectives. La qualité de l'interprétation réfère donc à la validité et la précision de l'interprétation.

La mobilisation des données physiologiques réfère également au processus de visualisation permettant de faciliter l'évaluation de l'expérience vécue par l'utilisateur lors d'un test utilisateur. Le deuxième essai aborde donc le développement d'une méthode permettant aux experts en UX de comprendre l'évolution de l'expérience vécue par les utilisateurs. Il vise à répondre à la question de recherche suivante : Comment présenter visuellement l'évolution de l'expérience vécue par les utilisateurs pour faciliter la compréhension de leur navigation ? Cet essai propose une approche permettant à des experts en UX de comprendre l'expérience vécue de manière approfondie. Il vise au développement d'une méthode permettant de visualiser et de contextualiser l'évolution de l'expérience psychophysiologique des utilisateurs lors de leur navigation sur une interface donnée.

Informations sur les articles

La collecte de données s'est déroulée à l'hiver 2019 alors que l'étudiante boursière de la Chaire de recherche industrielle CRSNG-Prompt en expérience utilisateur travaillait sur la rédaction de ce mémoire. L'analyse de la collecte de données auprès de 38 participants a permis la rédaction des deux essais. Le premier essai sera soumis pour publication au journal Behaviour and Information Technology alors que le deuxième essai a été accepté pour publication à la conférence HCI International 2020. Ce mémoire par articles est donc présenté en deux essais. Tout d'abord, la première partie aborde la contribution de l'utilisateur dans le processus d'interprétation de son expérience. Elle met en lumière la performance de l'utilisateur à interpréter son expérience psychophysiologique en comparaison à des experts en UX. Ensuite, la deuxième partie de ce mémoire aborde une nouvelle méthode de visualisation de l'expérience des utilisateurs. En effet, elle permet de suivre l'évolution émotionnelle et cognitive des utilisateurs au fil de leur navigation.

Résumé du premier essai

L’objectif de cet essai vise à comparer la qualité d’interprétation a posteriori de l’expérience vécue par l’utilisateur avec celle d’experts en UX dans un contexte de test utilisateur. Cette comparaison permet d’étudier l’importance d’inclure la perspective de l’utilisateur dans le processus d’évaluation. Dans le cadre de cette étude, l’utilisateur est un participant ayant complété le test utilisateur et étant émotionnellement engagé dans l’expérience tandis qu’un expert UX est un individu n’ayant pas participé à l’étude, mais ayant une connaissance dans le domaine de l'expérience utilisateurs. La comparaison entre ces deux groupes est basée sur les évaluations post-expérience d’une série d’événements en utilisant la méthode du rappel stimulé. La comparaison porte sur les moments les plus intenses de l’expérience, caractérisés par un état d’intensité émotionnelle élevé et une valence émotionnelle positive ou négative. Chaque moment est présenté sous forme de séquence vidéo du point de vue de la première personne afin de favoriser la réimmersion de l’utilisateur dans la tâche qu’il tentait d’accomplir. Les données de fixations oculaires sont superposées au vidéo afin d’être en mesure de suivre le regard de l’utilisateur sur l’interface. L’évaluation implique l’utilisation du Affective Slider (Betella & Verschure, 2016) ainsi qu’une description qualitative de l’intention de l’utilisateur à ce moment. La comparaison des performances entre l’utilisateur et les experts en UX en terme de précision d’interprétation permet de mettre en lumière l’importance d’inclure la perspective de l’utilisateur dans le processus d’évaluation de l’expérience. Les résultats suggèrent que les experts en UX ont une plus grande capacité que les utilisateurs à évaluer de manière précise l’expérience vécue par les utilisateurs pour les événements où il y a un accord interjuge, c’est-à-dire que les experts en UX ont une même interprétation de l’intention de l’utilisateur lors de l’événement.

Résumé du deuxième essai

L’objectif de cet essai est de présenter une méthode permettant la visualisation des données psychophysiologiques lors de la navigation des utilisateurs. Nous proposons une approche de visualisation qui présente, en conjonction avec le parcours réel d’un utilisateur sur une interface web, ses états cognitifs et émotionnels implicites au cours

de ce parcours. Plus précisément, nous présentons une nouvelle méthode de visualisation qui contextualise les données physiologiques et comportementales de l'utilisateur tout en interagissant avec un système d'information en ligne du secteur des services financiers. L'approche présente simultanément le comportement de l'utilisateur avec ses états cognitifs et émotionnels pour produire un aperçu complet de son expérience. La combinaison de ces méthodes permet d'obtenir des informations clés sur l'expérience de l'utilisateur et facilite la compréhension de l'évolution de l'expérience puisqu'elle montre où se trouvait l'utilisateur sur l'interface lorsqu'il a vécu un état cognitif et émotionnel donné. Afin de valider les formats de visualisation développés, ils ont été discutés avec des experts en UX.

Contributions et implications des résultats

Ce mémoire apporte des contributions théoriques ainsi que des implications managériales. En effet, du point de vue théorique, celui-ci contribue à la littérature de la rétrospection de l'utilisateur en utilisant l'approche du rappel stimulé (Lyle, 2003; Bentley, 2005; Lahlou, 2011; Bruun et al., 2016). Il permet également de bâtir sur la littérature actuelle en expérience utilisateur en définissant la qualité d'interprétation des utilisateurs lors de l'évaluation de leurs expériences. Enfin, il présente une nouvelle méthode servant également d'approche complémentaire à d'autres méthodes disponibles telles que les questionnaires et les entrevues qui permettent de recueillir des données autodéclarées et d'ajouter à la convergence de ces trois perspectives (émotions, cognition et comportements), ce que peu de méthodes permettent de faire (Coursaris et Kim, 2011).

Du point de vue managérial, cette recherche propose une méthode de visualisation de l'évolution des réactions émotionnelles des utilisateurs et permet ainsi de cibler facilement les sections et les zones à améliorer sur une interface. Cette méthode représente donc un outil intéressant pour les designers UX. Aussi, cette nouvelle méthode est utile pour comparer les expériences des utilisateurs sur différentes interfaces, notamment pour comparer les expériences utilisateurs d'une tâche spécifique sur des interfaces concurrentes. Enfin, les résultats sur la contribution de l'utilisateur dans le processus d'interprétation de l'expérience orientent les concepteurs lors de la sélection d'outils ainsi que sur l'importance accordée à la contribution de l'utilisateur.

Tableau 1 – Contributions aux différentes étapes du projet de recherche

Étapes du processus	**Contribution**
Définition des besoins du partenaire et de la question de recherche	Définir les objectifs d'affaires et traduire ces besoins en question de recherche - **70%** * Aide de mes directeurs ainsi que de l'équipe du Tech3Lab à établir les besoins d'affaire du partenaire. *Support de mes directeurs lors de la réflexion sur la question de recherche.
Revue de littérature	Recherche et lecture d'articles scientifiques reliés aux différents concepts utilisés dans le cadre de l'étude - **100%** Détermination des concepts pertinents à aborder afin de mettre en contexte la question de recherche - **100%** Rédaction de la revue de littérature - **100%** * Relecture et correction par mes directeurs de recherche. Définition des outils de mesures utilisés pour mesurer les construits - **75%** * Validation par l'équipe du Tech3Lab pour les outils de mesures utilisés afin de s'assurer qu'ils soient cohérents en fonction de la question de recherche. *Aide du laboratoire fournie concernant l'évaluation des outils des mesures psychophysiologiques.
Design expérimental	Complétion de la demande au CER - **100%** Rédaction du protocole d'expérimentation - **75%** * Suivi par l'équipe du Tech3Lab afin de s'assurer que le protocole soit complet. * Utilisation des formulaires standardisés (consentement et compensation) du Tech3lab.

Recrutement des participants	Recrutement des participants - **50%** *le partenaire de la Chaire a effectué la moitié du recrutement via une firme externe. Gestion des compensations - **100%** *le cartable comprenant les formulaires de compensation a été assemblé par l'équipe du laboratoire.
Collecte de données	Collecte des données - **50%** *Support technique des assistantes de recherche pour l'ensemble des collectes.
Extraction et transformation des données	Extraction et mise en forme des données physiologiques, psychométriques, cognitives et émotionnelles pour l'analyse statistique - **100%**
Analyse de données	Retranscription des entrevues qualitatives - **90%** * Aide de l'équipe du laboratoire à des fins de retranscription. Analyse des données physiologiques et oculométriques - **100%** Analyse des statistiques du mémoire - **90%** * Aide du statisticien afin d'effectuer des tests statistiques sur le logiciel SAS 9.4
Rédaction	Écriture des articles scientifiques et managériaux - **100%** *Mes directeurs ont apporté d'importants commentaires constructifs à des fins d'amélioration de la qualité des articles.

Structure du mémoire

Les prochains chapitres abordent le concept d’expérience utilisateur et se déclinent en deux essais. Tout d’abord, les deux essais présentent des approches distinctes à la compréhension de l’expérience vécue des utilisateurs. Il est à noter que les résultats obtenus pour l’ensemble des articles proviennent de la même collecte de données ainsi que du même design expérimental. Ensuite, la conclusion présente un rappel de la question de recherche et met en lumière les principaux résultats ainsi que les contributions au niveau académique et managérial suivi des limites de recherche et des potentielles avenues de recherche.

Chapitre 1 : Premier Essai

A comparative investigation on the quality of interpretation of the user's emotional experience

Marianne Veilleux, Sylvain Sénécal and Pierre-Majorique Léger

HEC Montréal, Montréal, Canada

marianne.veilleux@hec.ca

Abstract. Research in User Experience (UX) focuses on users and their emotional states. To date, various methods, both qualitative and quantitative, have been used to measure the user experience. This study addresses the quality of interpretation of users' emotional experience measured by psychophysiological tools. The user's ability to accurately assess retrospectively his/her affective states is compared to the ability of a group control of UX experts. A study involving 25 participants was performed to evaluate if users could retrospectively better assess their emotional state during specific events than UX experts. We found that users do not seem able to reliably assess their own emotional experience. However, UX experts appear to have a better ability to evaluate the user's experience in contexts where they agree on a common interpretation of the user's intentions. This tendency is even stronger for negative experience rather than positive experience. Also, there seem to be no recency effect influencing user's interpretation. This research contributes to the existing UX literature by providing more precision into the importance and abilities of users and UX experts into the assessment of the user's emotional experience.

Keywords. User Journey; Memory Recall; Psychophysiological Pain Points; Psychophysiological Pleasure Points; UX research

1 Introduction

Research in human-computer interaction (HCI) literature aims at gaining a deeper understanding of the reasoning behind the user's decision-making by developing approaches and techniques in order to facilitate user-centred design (Smith, 1997). More recently, user experience (UX), which can be defined as a study of the interaction between human and interface (Vermeeren et al., 2015), has gained popularity within the HCI literature. More specifically, user experience refers to the total effects felt internally by a user as a result of an interaction with a system (Hartson & Pyla, 2012). In general, research in UX focuses on the elements that make up a positive experience (Vermeeren, Roto et al. 2015). Consequently, emotional responses are put forward as important aspects of user-centred design as the user's remembrance of past events is mostly influenced by the intensity of emotions (MacKay et al., 2004). These constructs are measured using perceptual scales, psychophysiological measures, and observable behaviours (Bradley & Lang, 1994; Bruun et al., 2016; Hockenbury & Hockenbury, 2016).

Self-reports are the dominant method used to evaluate the user's behaviours and subjective experiences within social and behavioural science (Schwarz, 2007). In fact, the UX literature suggests the use of retrospective tools, such as the Self-Assessment Manikin (SAM) (Bradley & Lang, 1994), to assess users' emotional states (Bruun & Ahm, 2015). However, research outside the UX context suggests that biases remain with such subjective methods and that the subjective emotional responses will not reflect concurrent experiences (Bargas-Avila & Hornbaek, 2011; Bruun & Ahm, 2015). Indeed, studies show that what users recall after an experiment is not fully the same as what they felt during that same experiment (Wilson & Sasse, 2004; Schooler, 2011; Cockburn et al., 2017; Hetland et al., 2018). Indeed, users are not fully aware or able to report on their own cognitive processes (Nisbett & Wilson, 1977). A study by Kahneman and Riis (2005) states that self-evaluations are often biased, and people are less accurate when evaluating their own performance than when evaluating the performance of others. Therefore, retrospective evaluations of past emotional experiences are prone to errors (Kahneman & Riis, 2005). The memory-experience gap explains such discrepancy between subjective and concurrent ratings of emotional states. According to a study by Scherer (1984), emotions are short terms and intensive peaks which makes them difficult to recall at a later point in time. Thus, using only retrospective self-report measures to assess emotional and cognitive processes can represent an incomplete approach to fully understand the user’s experience. For this reason, several studies use video-situated recall techniques in order to recollect more accurately the user’s experience during specific events (Lim, 2002; Bentley et al., 2005; Lahlou, 2011; Bruun et al., 2016). These methods allow the re-immersion of users into past experiments in order to facilitate the retrospective verbal reports without interrupting them during that same experiment.

Psychophysiological data represents an interesting avenue for measuring emotional responses. Studies suggest that psychophysiological measures help identify significant moments such as frustration (Forne, 2012; Giroux-Huppé et al., 2019). However, interpretation of data remains a challenge when using psychophysiological measures (Cacioppo & Tassinary, 1990; Kecklund & Akerstedt, 2004; Forne, 2012). Hence, many studies now use verbal in conjunction with nonverbal measures with the aim of finding the right balance (Mandryk et al., 2006; Ivonin et al., 2014; Hetland et al., 2018).

In this article, we address the discrepancy between the user’s concurrent affective state and their subjective evaluation of their affective state following the experiment. We study such discrepancy by conducting an experiment with users and UX experts. We explore the level of congruency between the user’s and the UX expert’s respective interpretations of specific events and the psychophysiological measures of these events. In our study, users refer to participants who are emotionally involved and completed the experiment whereas UX experts represent impartial individuals that did not participate in the experiment and thus, are more distant emotionally when asked to evaluate sequences from the experiment. Several studies analyze the congruency between psychophysiological measures and what a user recall from that experiment (Bentley et al., 2005; Bruun et al., 2016; Lourties et al., 2018) or the efficiency of a recall method as a complementary tool to psychophysiological measures (Lim, 2002). However, little research focuses on the user's quality of interpretations of their own experience in comparison to a group of UX experts. This reflection brings us to the following question: When retrospectively interpreting their peak experiences, are users more accurate than UX experts?

Such a questioning also has practical implications, as it makes it possible to define the usefulness of keeping the user in the loop for the interpretation of psychophysiological data in UX evaluation. Moreover, it identifies the ability of UX expert to understand accurately the user's emotional experience.

2 Literature Review

2.1 Continuous Psychophysiological Measures of User Experience

Psychophysiological measures have gained popularity through recent years as they allow researchers to assess users' automation and non-conscious reactions to a specific stimulus (Riedl & Léger, 2016). Psychophysiological measures represent physical signals generated as a response to psychological changes and measured in real time (Dirican & Göktürk, 2011). A variety of physiological and psychophysiological measures are used in order to better understand users' responses to stimulus such as electrocardiography (ECG), skin-based measures (EDA), ocular measures, brain measures (EEG), respiration rates, and blood pressure (Charles & Nixon, 2019). Psychophysiological measures represent useful indicators of mental effort and stress (Vicente et al., 1987; Picard, 1995; Duric et al., 2002; Maia & Furtado, 2016) and give access to a more natural reaction from participants since data from users' reaction is available in real time without being intrusive (Dirican & Göktürk, 2011; Guinea & Webster, 2013; Guinea et al., 2014). Having data moment by moment allows researchers to gain insights into automatic and unconscious emotional reactions without interrupting users (Ivonin et al., 2014). Studies show that interrupting users while completing a task often leads to biased results (Zijlstra et al., 1999; Bailey et al., 2006; Bailey & Konstan, 2006). Moreover, Kahneman and Riis (2005) suggest that measures such as heart rate, skin conductance level, or electroencephalography offer uninterrupted reports of emotions. Psychophysiological measures thus consider non-conscious reactions from users (Guinea & Markus, 2009;

Dimoka, 2010; Dimoka et al., 2011). Researchers are then able to avoid memory bias from users (Guinea & Webster, 2013). These measures give a more complete view of the human-computer interaction and represent a useful tool to assess elements or events of cognitive or emotional relevance to the users (Picard, 1995; Ward & Marsden, 2003; Bentley et al., 2005).

Emotional experiences are often defined by two dimensional constructs well established within the UX literature, which are valence and arousal (Bradley & Lang, 1994; Guinea & Markus, 2009; Léger et al., 2014). The circumplex model of affect is a popular dimensional approach to define all affective states according to these two fundamental psychophysiological dimensions (Russell, 1980). Valence varies from positive to negative emotions (Colombetti, 2005; Maia & Furtado, 2016) or as other researchers define it, from pleasure to displeasure (Boucsein, 2012; Forne, 2012; Guinea & Webster, 2013). Burton-Jones and Gallivan (2007) describe valence as "what a user feels" and facial expression analysis remains the most reliable way to measure this construct since individuals tend to express their emotions with micro-movements of the facial muscles (Uyl & Kuilenburg, 2005). The complementary construct is arousal. It represents the level of intensity, ranging from calm to excited (Boucsein, 2012; Maia & Furtado, 2016). The electrodermal activity is an accurate analysis tool used in the literature to better determine levels of arousal. It collects variation of electrical properties of the skin from the secretion of sweat in the hands (Hassenzahl & Tractinsky, 2006; Boucsein, 2012). Overall, psychophysiological measures allow translating emotions and affect into specific and measurable constructs. Nevertheless, the interpretation of such measures remains an important challenge and memory recall method helps to provide such insights about users' emotional states.

2.2 Concurrent Methods

A diversity of methods have been developed in order to gain greater insights into the cognitive processes of users during the experience. Indeed, understanding the reasoning behind users' decision-making from their perspectives is an important goal in the HCI literature (Rosenthal, 1966; Smith, 1997). Multiple techniques such as 1) concurrent think aloud, where users are asked to explicitly mention their thoughts while completing a task or 2) thought sampling, where participants report what they were thinking just prior to the interruption (Omodei & McLennan, 1994) are used in order to better understand the users' point of view during the experience. However, these concurrent verbal protocols are criticized for being intrusive and for potentially influencing cognitive processes (Bentley et al., 2005; Schooler, 2011). Moreover, a study by Van Gog et al. (2005) suggests that neither concurrent reporting – such as thought sampling – nor retrospective reporting - such as retrospective think aloud - are enough to provide a comprehensive review of users' processes. Consequently, situated recall methods, which are methods to elicit information about users' affect during system use (Bentley et al., 2005), were developed to enlighten users' affective responses without interfering with the user's behaviour during the experiment. In the next section, we discuss the memory recall literature focusing on the variety of methods using similar procedures to elicit cognitive processes. We conclude by defining the proposed approach used in our study.

2.3 Retrospective Methods

2.3.1 Cued-Recall Debrief (CRD)

Cued-Recall Debrief (CRD) is a recall method developed by Omodei and McLennan (1994) in order to evoke affective responses during specific episodes without interfering with the user's behaviour in a natural context. This method aims to enhance users' ability to recall affective experiences of past episodes by showing videos sequences of an entire event. Users are encouraged to express their thoughts, intentions, and emotions during the process. A key component of this approach is that the material shown to users must be in a first-person point of view in order to fully re-immerse participants into their experience (Bruun et al., 2016). This technique is applied in different contexts such as usability testing (Bruun et al., 2016) or assessing decision-making (Omodei & McLennan, 1994). CRD has shown to provide more detailed responses compared to other free recall methods where participants are more inclined to organize their responses logically during free recall and thus, not reporting exactly what happened (Omodei & McLennan, 1994; Bentley et al., 2005; Bruun et al., 2016). In fact, CRD creates up to four times the number of recalls compared to others free recall methods (Omodei & McLennan, 1994). Research by Bentley (2005) suggests that CRD provides a deeper understanding of the participants affect and can identify and differentiate affective experiences.

2.3.2 Subjective Evidence-Based Ethnography (SEBE)

Mostly used in social sciences, Subjective Evidence-Based Ethnography (SEBE) is used in different settings for a variety of research objectives such as collaborative design, individual, and group decision-making, and counselling (Lahlou, 2011). SEBE is described as an investigation technique which builds on advantages of information technology to record, comment, and analyze data (Goldman et al., 2007; Hollan & Hutchins, 2009; Lahlou, 2010; Lewis et al., 2010). The procedure is based on audiovisual recordings from a first-person point of view followed by a self-confrontation of participants with the video sequences to collect their subjective thoughts on the experience, and analysis of the recordings along with discussion of final interpretations with participants (Lahlou, 2011). These steps allow participants to better describe their goals and affective states at the moment of action without interfering with the cognitive process since the event already took place (Lahlou, 2011). The purpose of this method is to understand the experience felt by the user from her perspective and to analyze her experience without it being influenced by inadequate preconceptions (Lahlou, 2011).

2.3.3 Self-Confrontation Method (SCM)

The Self-Confrontation Method (SCM) is a technique traditionally used in psychology and counselling research (Hermans, 1990) and more recently used in the study of online behaviour (Lim, 2002). It triangulates interviews with the observation method in order to have a more complete analysis of participants' actions and cognition (Lim, 2002). SCM follows procedures similar to other methods presented previously. First, the participant's actions, while performing different tasks, are recorded on video. Second, the participant is asked to review the recordings of the event. Last, the recording is paused at strategic moments for the participant to comment and explain retrospectively any thoughts, intentions or emotions when performing these tasks (Lim, 2002). A recent study by Barbier (2018) uses the Self-Confrontation Method with drivers and played a video sequence of their driving activities for them to comment and share their subjective thoughts on the experience. Researchers select video sequences based on critical segments such as positive acceleration or a steering wheel, and then views the sequence with participants (Barbier et al., 2018). SCM facilitates the understanding of participants' intentions since they are asked to explain it and it is unlikely for participants to be systematically wrong about their intentions even when there is a certain delay between the moment they actually felt their intentions and the moment they are asked about it (Harré & Secord, 1972).

2.3.4 Stimulated Recall (SR)

Stimulated Recall (SR) enters the category of introspective research method in which participants are asked to recall their intentions during specific episodes which are supported by video-recordings (Lyle, 2003). This method is used in teaching (Zanting et al., 2001; Chittenden, 2002), counselling (Halford & Sanders, 1990) and nursing (Lindgren, 2002). The use of SR varies across studies, but the normal procedure is to replay the video sequence to participants while asking open-ended questions in a structured way. Stimulated recall is considered useful for exploring cognitive processes, but in order to ensure validity, time between the experiment and the recall session must be minimized to prevent participants from leaving the cognitive process involved with the task completion (Lyle, 2003). Researchers have shown that in the context of stimulated recall, video recordings did not affect the user's decision-making during the experiment and that the verbal report had some resemblance to the "think aloud" techniques (Lyle, 2003).

Table 1. Review of Situated Recall Methods

Method	**Description**	**Studies**
Cued-Recall Debrief	First-person point of view recordings. Debrief facilitators and participants watch together the replay of video sequences of an entire event. Participants are encouraged to narrate and express their thoughts during the process. Facilitator observes participant to see if there is any reaction from participants If there is a reaction, he waits to see if the participant will comment. If the participant doesn't comment, the facilitator may ask what they were thinking or feeling at that time. Each Debrief is carefully analyzed (comments can be quantified into categories; qualitative analysis to gain insight on psychological processes).	(Omodei & McLennan, 1994) (Bentley et al., 2005) (Bruun et al., 2016)
Subjective Evidence-Based Ethnography	First-person point of view audiovisual recordings (video-camera worn at eye-level). Confrontation of recordings with participants to collect their subjective experience. (Analysis of events with researchers when looking at the recordings). Researchers formulate findings. Discussion on final interpretations with participants to check the validity of the interpretations.	(Hollan & Hutchins, 2009) (Lahlou, 2010) (Lahlou, 2011)

Self-Confrontation Method	Video-recording of subjects' actions Review of the video recordings by participants of their own actions. The video is paused at certain moments and asking subjects to recount any thoughts and emotions they had during those actions. Interviews are transcribed and correlated with the actor's thoughts and emotions with their respective actions.	(Hermans, 1990) (Lim, 2002) (Guérin et al., 2004) (Lyddon et al., 2006) (Barbier et al., 2018)
Stimulated Recall	Video-recording of subjects' actions Participants are asked to recall their thoughts during those events. Series of structured, but relatively open-ended, questions (centred on description-thinking-noticing-alternative behaviours).	(Halford & Sanders, 1990) (Zanting et al., 2001) (Lindgren, 2002) (Lyle, 2003)

Retrospective methods are more interesting than concurrent methods since they allow users to have more natural responses. However, these retrospective methods present several limits such as retrospective biases from users. Thus, a comparative analysis between interpretations of users and UX experts, represents an interesting avenue in order to evaluate the reliability of users in their quality of interpretation.

2.4 Judgment Studies

The term 'judgment studies' refers most generally to studies in which behaviours, persons, objects, or concepts are evaluated by one or more judges, raters, decoders, or categorizers (Harrigan et al., 2005). Moreover, this approach was first developed due to the several biases revolving around users' perceptions. Indeed, the accuracy of self-perceptions have been challenged within the psychology literature (Robins & John, 1997) due to the ambiguity in the ability of people to be objective when evaluating their own behaviour (John & Robins, 1994; Gosling et al., 1998). Moreover, observers are also less likely to succumb to the same biases found in self-reports (Dmello, 2016). A study by Borkenau and Ostendorf (1987) found that observers had a high level of accuracy in differentiating categories of behaviours. Several studies have compared the accuracy of self-reports in comparison to evaluations by peers (John & Robins, 1994; Kolar & Funder, 1996).

The starting point of a procedure in judgment studies implies that encoders, represented as users in our study, experiences a variety of states (e.g., anxiety, joy). Then, these states are observed by decoders (i.e., UX experts) who make judgments about the users' behaviours. The judgments employ a variety of metrics and measures from physical units of measurement to psychological units of measurement. The relationship between the encoder's attitude and the decoder's accuracy of judgment characterizes judgment studies (Harrigan et al., 2005). The most common purpose of judgment studies is to ask decoders to assess encoder states in order to correlate their judgments with the determined definition of encoder states. These studies are often in the context of accuracy studies in order to delimit the differences among the decoders in degree to the accuracy shown (Harrigan et al., 2005). Several researches have suggested that nonverbal behaviours can be assessed accurately by decoders (Ekman, 1965, 1973). Thus, in this study, we seek to compare the accuracy of interpretations of online behaviours between users (i.e., encoders) and UX experts (i.e., decoders) in order to investigate their quality of interpretation.

3 Proposed Approach and Hypothesis Development

The multiple approaches presented above shows a common purpose, which is to contextualize the experience from the user's point of view. In our study, the situated recall method is used to re-immerse users and immerse UX experts into the experiment in order to facilitate the assessment of the user's affective states. Throughout our research, we seek to compare the ability of users and UX experts to accurately assess the user's emotional responses during specific events. The use of impartial individuals, such as UX experts, to interpret and evaluate social behaviour is being widely used in personality, social, and other fields of psychology (John & Robins, 1994). In fact, one of the most common approaches to portray participants' affective states is to use impartial individuals as evaluators (Dmello, 2016). This strategy usually involves a small group of skilled observers in order to portray participants' affective states (Bernardin & Buckley, 1981; Dmello, 2016). By comparing impartial individuals' evaluations with the user's evaluation, we explore their ability to assess emotional states of specific sequences. Thus, this study aims to explore the extent to which the user brings more reliability into the interpretation process.

3.1 Hypothesis Development

According to Kahneman and Riis (2005), introspective evaluations of past events by users depend on two main criteria which are an accurate recollection of emotions and a reasonable integration of experiences spread over a period of time. However, several studies challenge user objectivity in their self-evaluation of past emotional states (John & Robins, 1994; Kahneman & Riis, 2005; Miron-Shatz et al., 2009; Bruun & Ahm, 2015; Dmello, 2016). A study by John & Robins (1994) suggests that people are less accurate when evaluating their own performance than when evaluating the performance of others. Indeed, users are more likely to alter or misrepresent their affective states due to a variety of biases such as social desirability bias (Krosnick, 1999) or differences in levels of narcissism (John & Robins, 1994). A study by Bruun and Ahm (2015) found that there is a significant discrepancy between retrospective and concurrent assessments of emotions. Consequently, self-evaluations differ from evaluations by peers or impartial individuals (Tesser & Campbell, 1983; John & Robins, 1994).

Judgment studies were initially developed due to the many biases surrounding the user's perception of the user and the user's ability to be objective when evaluating his or her own behaviour (Robins & John, 1997). The recency effect represents one of those biases. Indeed, it is a well known and established concept in the literature (Murdock, 1962; Waugh & Norman, 1965; Glanzer & Cunitz, 1966; Watkins, 1974;

de Gelder & Vroomen, 1997; Bloom & Watkins, 1999; Lourties et al., 2018), which refers to the increased memory and greater influence of the last moment of the experience (Kahneman et al., 1993). As a part of our study, we explore the extent to which the accuracy of users' interpretations is affected by the recency effect. Given the nature of our study, the recency effect can only influence the user's evaluation since the expert is not subject to such effect. Therefore, the recency effect should influence the user's ability to accurately interpret its emotions negatively.

The evaluation of the user experience varies according to who oversees the interpretation of results. We have previously discussed the user's contribution as an evaluator of his own experience, but the role of evaluators can also be taken on by an external observer. Consequently, several studies compare the ability of the external observer and the participant to evaluate the experience (Kolar & Funder, 1996; Robins & John, 1997). A study by Borkenau and Ostendorf (1987) suggests that observers have a high degree of precision in differentiating different categories of behaviour. Moreover, several studies suggest that personality assessment made by experienced observers can also provide accurate information about the participant (Oh et al., 2011). Hence, the use of experts as evaluators has also been explored as an alternative to the user's self-reporting. A study by Jarodzka et al. (2010), which aims at identifying expertise effects in the perception and interpretation of visual stimuli, found that experts were able to perform more accurately than novices. Given all of these elements, we expect UX experts to be more accurate in their interpretation of the user's experience than users themselves since they are less prone to such biases. Which leads us to the following hypothesis:

H1. When exposed to the user's experience recordings, UX experts perform better at identifying accurately users' emotional state during the online experience than users themselves.

Research shows that the intensity of events represents a key element in the remembrance of past events. (Bradley et al., 1992; Ariely, 1998; Hamann, 2001). In fact, the most intense event of an experience is defined as a "peak moment" represents a key concept which partially explains why individuals evaluate experiences based on short episodes representing the most intense moments (Kahneman et al., 1993). Hence, users tend to remember the most intense events of an experience when asked to do so retrospectively (Kahneman et al., 1993). Moreover, negative memories are usually better recalled than positive memories (Thomas & Diener, 1990; Baumeister et al., 2001). A study by Finkenauer & Rime (1998) found that when participants were asked to recall past emotional events, including both positive and negative events, not only are negative moments better recalled, but they are also easier to report. Frustration is a typical example of a negative event and is defined as high arousal and negative emotional valence according to the Russell circumplex of affect. Thus, we expect users to better assess their emotions during a negative experience than a positive experience.

H2. Users identify more accurately their emotional states for negative experience than positive experience when asked to report on specific video recordings.

4 Method

For this study, we have collected data on users while they were completing tasks on a website. Then, we identified positive and negative events from the user's experience based on the psychophysiological data collected. The selection of events was based on a study by Giroux-Huppé et al. (2019), which proposes a method that uses psychophysiological data in order to identify implicit pain points in a user's online journey. An implicit pain point is defined as a moment in reaction to a situation, during which the user experiences an automatic physiological response characterized by a high level of emotional arousal and a negative emotional valence (Giroux-Huppé et al., 2019). Likewise, an implicit pleasure point is defined as a reaction to a situation, during which the user experiences high levels of emotional arousal and a positive emotional valence. Thus, we used the same approach in order to define our positive and negative experiences. Then, users and UX experts were asked to evaluate these sequences in order to investigate the added value of including the user and to identify potential biases. To test our hypotheses, we use psychophysiological measures, psychometric measures and qualitative interpretations. These measures allow us to define both groups' ability to evaluate accurately the user's psychophysiological experience. For this research, we studied the statistical relationship between psychophysiological data and the respective evaluations of users and UX experts. The correlations between variables indicate whether the users and UX experts can accurately assess the user's emotional experience. This project was approved by the Ethics Committee of our institution.

4.1 Sample

A total of 25 participants were recruited in this study, composed of sixteen men (64%) and nine women (36%). Their age ranged from 23 to 61 years old (W: mean = 35.4, SD = 10.3; M: mean = 38.8, SD = 12.2). Prior to the study, participants were asked to rate the organization whose website will be navigated on the Net Promoter Score (NPS) Scale (Markey & Reichheld, 2011). To reduce any potential influence of the brand on participants' experience, only participants who rated the organization between a score of 3 to 8 were invited to participate in our study. NPS is a one question metric used to assess the customer's overall perception of the brand using a likert scale from 0 to 10, 0 being not at all likely to recommend the brand, and 10 being extremely likely to. Each participant received a moderate financial compensation to participate. To participate, users needed a normal vision and were pre-screened for glasses, laser eye surgery, astigmatism, epilepsy, neurological and psychiatric diagnoses.

4.2 Procedure

The test sessions took place in a laboratory and lasted around 75 minutes including participant greeting and the final debrief. Upon arrival, participants were asked to complete a consent form, after which they were explained the purpose of the study and fitted with physiological sensors. Participants were told that their participation was required in order to evaluate a financial website and were only informed in debriefing that the purpose of the study was to investigate their ability to recall episodes from their experiment. All tasks were divided into three types of tasks (e.g., searching for specific information on the website, using a tool in order to find answers, and find ways to contact the financial institution). As shown in Figure 1, each participant was assigned three sets of three types of tasks on the website out of all groups of tasks available. Each set allows participants to perform tasks which are frequently performed on the website. This approach allows to minimize the effect of the task by having a varied number of different tasks of similar complexity.

Once all three sets of tasks were completed, a facilitator came into the room for a debrief with each user on their experience during which users were asked about their overall experience. During the debriefing process, psychophysiological data was analyzed in another room for the selection of sequences for a period of 10 minutes. A lab management system allowed us to triangulate UX measures in order to synchronize and visualize in a timely manner series of positive and negative events (Léger et al., 2019). The two most intense negative events (high arousal, low valence) and the two most intense positive events (high arousal, high valence) were selected out of all events, which respectively represent implicit psychophysiological pain points and pleasure points (Giroux-Huppé et al., 2019). An additional positive moment and an additional negative moment were selected in case of an inappropriate choice, for instance, a supposedly intense event which was not remembered by the user. Hence, the additional two sequences were selected if one or more sequences out of the set of 4 is not appropriate. The order in which the types of moments are shown to the participants was randomized.

When the debrief about the user's overall experience was over, the researcher went into the room with the chosen sequences. A total of four sequences were shown to participants for them to become familiar with the re-immersion without participants getting too comfortable in the process. The reporting procedure consists of asking participants to verbalize their thought processes during task completion, based on a cue of their performance. The cue consisted of the recordings from the user's own eye movements, with eye tracking, superimposed onto the video of the screen (van Gog et al., 2005). Users were re-immersed into the experience for a total period of ten seconds (i.e., five seconds to the peak event followed by an additional five seconds). A time interval of 10 seconds allows for latency between the recognition of the stimuli and the triggering of emotional responses (Bruun et al., 2016). Moreover, they were asked to elicit intentions at the time of action. More concretely, when shown the sequences they had to answer the main open-ended question, "What was your intention at that time?” in order to guide users in their verbal answers. They also had to evaluate their levels of valence and arousal using the Affective Slider (Betella & Verschure, 2016). The Affective Slider is a digital self-reporting tool composed of two slider controls for assessment of pleasure and arousal. Each participant was shown up to four short sequences. However, due to time limitations and inconclusive selection of events, the total of sequences shown to participants differs. Therefore, we reached a total of 89 sequences for the overall sample.

Once all study sessions were done, two UX experts were recruited to analyze users’ experience recordings. UX experts were impartial individuals selected on a voluntary basis and both understood usability and had experiences in the analysis of a user journey. UX experts were asked to interpret all the 89 sequences using the same approach as users. For each sequence, UX experts had a time interval of 30 seconds (15 seconds prior to the specific point followed by an additional 15 seconds). This time interval was defined following a pre-test with four UX experts who had to select the most appropriate time interval. Since UX experts were not as involved in the experiment as users, a greater time interval allowed them to have more contextual information. UX experts were asked to evaluate levels of arousal and valence of participants using the Affective Slider (Betella & Verschure, 2016). Also, they had to give a description of the user's intentions during that event. The visual support consists of sequences of the same events shown previously to users. Having both users and UX experts’ evaluations of the same events allows investigating potential biases from engaging the user into the process and to understand the difference in performance between both parties. The overall procedure of the complete experience is shown below in Figure 1.

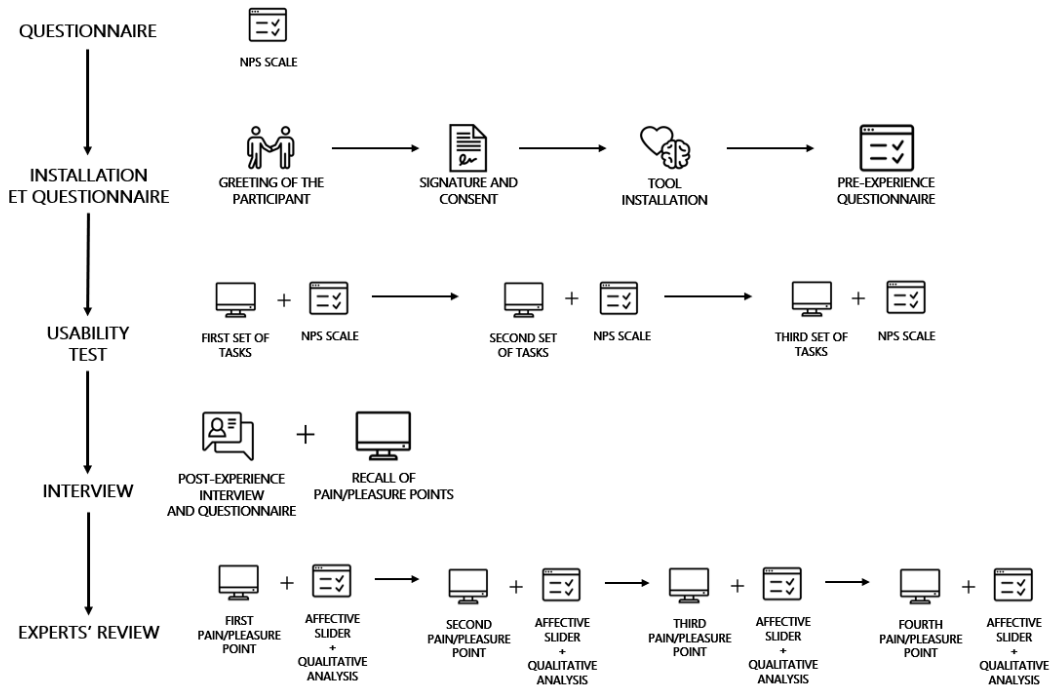

Fig. 1. Experimental Procedure

4.3 Measures

To evaluate users' and UX experts' ability to interpret accurately specific sequences of users' experiences, we first need to capture the right data in order to analyze and identify negative and positive moments. These points refer to the "peak effect" developed by Kahneman et al. (1993), which implies that in a retrospective evaluation, participants tend to give excessive weight to the most intense moment of an experiment. Thus, "peaks" are defined as extreme emotions during an experience. As mentioned previously, according to the circumplex model of affect, valence and arousal allow us to easily identity those emotional "peaks".

First, a Tobii X-60 eye-tracker (Stockholm, Sweden) sampled at 60 Hz was used to capture eye-tracking data (Laeng et al., 2012), and Tobii Studio was used to record the experience. The recordings of the eye-tracking device were used for contextualization during the subjective recall and shown to both users and UX experts in order to facilitate re-immersion into the experiment without interfering with the user's interaction with the website. Then, arousal is measured using electrodermal activity (EDA) collected with the Acknowledge software sampled at 500 Hz (BIOPAC, Goleta, USA). EDA has been proven to be a useful tool to measure arousal (Hassenzahl & Tractinsky, 2006; Boucsein, 2012; Braithwaite, Watson, Jones, & Rowe, 2013). In fact, EDA measures the variation of the electrical properties of the skin in response to the secretion of sweat from the palm of the hand in our case, which indicates cognitive and emotional stimulation (Stern, Ray & Quigley, 2001). Lastly, facial expression and micro-movements of the facial muscles represent a measure of emotional valence using the analysis tool FaceReaderTM (Noldus, Wageningen, Netherlands). Indeed, this tool is used to observe facial reaction and assess emotional valence with a score ranging from -1 (most negative emotions) to +1 (most positive emotions) with a temporal precision of 30 inferences per second (Ekman & Friesen, 2003; Uyl & Kuilenburg, 2005). The valence ratio is calculated as the intensity of the positive expression minus the intensity of the negative expression with the highest intensity. Then to synchronize all the signals between Acknowledge, FaceReader and Tobii we used Observer XT (Noldus, Wageningen, Netherlands) as recommended by Léger et al. (2014). All these tools were used in order to identify the negative and positive "peaks". Lastly, all physiological measures were triangulated using Cube Hx (Patent US10, 368, 741 B2, 2019) in order to identify at what point was experiencing such emotional "peaks" (Courtemanche et al., 2019). This lab management system detects in a timely manner series of positive and negative events (Léger et al., 2019). The tool used by both users' and UX experts' evaluation of their emotional state is the Affective Slider (Betella & Verschure, 2016). The Affective Slider is a scale from 0 to 100, used by both parties, in order to rate the perceived levels of valence and arousal of the user.

4.4 Analysis Process

Firstly, the sequences per participant were chosen based on their psychophysiological data. The emotional "peaks", as mentioned previously, represent the most intense (high arousal) events from the user's experience (Fredrickson & Kahneman, 1993). Moreover, to be characterized as a negative moment, the data associated needed to be in the ninetieth percentiles of EDA (i.e., high arousal) and in the tenth percentile of valence (i.e., high negative valence). Consequently, the data is associated with a positive moment when it is in the ninetieth percentiles of EDA and valence (i.e., high positive valence). “Peak” emotional experiences have been demonstrated to influence the affective assessment (Cockburn et al., 2017). Therefore, the preselected sequences shown to participants represent both positive and negative experiences.

Moreover, we define our cued retrospective approach as a video-assisted recall showing preselected sequences from a first-person point of view. Furthermore, time between the experience and the recall session is minimized for participants to recall easily on their experience. As mentioned, the preselected video sequences showed to participants are chosen based on a psychophysiological analysis of users’ experience, selecting only the most intense positive and negative events. Consequently, users are asked to contextualize their intentions and evaluate their emotions during those negative and positive events.

Once all the data is captured, the codification concerns mostly users’ and UX experts’ evaluations. The Affective Slider (Betella & Verschure, 2016) is a scale from 0 to 100, used by both parties, in order to rate perceived levels of valence and arousal. Since we want to compare the evaluation of negative events and positive events with the psychophysiological data, we determined a threshold based on all users’ evaluations on the Affective Slider (Betella & Verschure, 2016). From the total observations, we established a median in order to separate positive from negative events. Consequently, a moment is defined as positive when the valence and arousal have a score higher than the median of 0.1 multiplied by the standard deviation. Hence, a moment is defined as negative when the valence has a score lower than the median of 0.1 multiplied by the standard deviation and the arousal has a score higher than the median by the same amount. The table below shows this codification in order to compare both parties’ interpretations and detect any (non)congruence within evaluations of negative and positive events.

Table 2. Categorization of Psychophysiological and Qualitative Data

	Positive event	Neutral event	Negative event
Psychophysiological measures	Valence > median + 0.1*std Arousal > median + 0.1*std	All other possible combinations	Valence < median + 0.1*std Arousal > median + 0.1*std
Qualitative analysis	0 : Inaccurate interpretation of intentions 1: Accurate interpretation of intentions		

An analysis of users' and UX experts' verbatim was also done in order to identify any similarities between users and UX experts' responses, which consisted of a textual analysis of all verbatim. This type of analysis allowed to rapidly identify similar content throughout interpretations and determine if UX experts agree on a common understanding. Since it is unlikely for participants to be systematically wrong about their intention (Harré & Secord, 1972), we assume that users have the right interpretations of intent. Therefore, if the UX expert's content analysis was similar to the user's content analysis, the UX expert's interpretation was defined as correct. UX experts' verbatim evaluations were given a score of 0 by the researcher if it was incorrect, and a score of 1 if it was correct (see Table 3). In sum, each part of the interpretations – affective sliders and verbatim – allowed us to evaluate more specifically users and UX experts' ability to accurately assess users' emotions and intentions.

Table 3. Example of Qualitative Coding Scheme

User	UX expert 1	UX expert 2
P17 : "I was **a little bit lost**, but I was not frustrated nor angry".	"The user **is lost on the page** […] and is browsing the webpage for approx. Ten seconds". 1 : Accurate interpretation of intentions.	"The user **does not know where to go** on the website and seeks instructions to help him/her". Accurate interpretation of intentions.

Since the evaluations by UX experts rely on an analysis by two individuals, we first need to investigate the level of agreement between both UX experts in order to confirm the inter-rater reliability. According to all 89 sequences, the level of agreement between both UX experts is very low for valence (0.329) and arousal (0.223). Thus, these results suggest that UX experts do not seem to agree on common interpretations of valence and arousal. According to these results, a comparison of performance between users and UX experts can't lead to generalizable conclusions due to the unreliability of UX experts' performance. Consequently, we want to investigate whether a variable could influence these results. Therefore, we have included the qualitative variable, being the interpretation of the user's intention, in order to establish if this variable could impact the UX experts' reliability. The addition of this qualitative variable was done by dividing UX experts' interpretation into two clusters: ambiguous events and unambiguous events. This division was done by grouping sequences where there was a common understanding of the user's intentions. Thus, ambiguous events refer to sequences where UX experts disagree on a common understanding of the user's intentions, and unambiguous events refer to sequences where both UX experts agree on interpretations. Ambiguous events include 63 sequences whereas unambiguous events include 26 sequences. As expected, we found that the inter-rater reliability coefficient is higher for unambiguous events in terms of valence (0.606) and arousal (0.78) and remains low for ambiguous events both in terms of valence (0.099) and arousal (-0.011). Hence, UX experts seem to agree on interpretations of valence and arousal for unambiguous events. Therefore, the following results address unambiguous sequences since these are the results that can be generalized. The results of all the sequences are also presented in order to identify general trends.

5 Results

In this section, the results compare the performance of users and UX experts across different events (i.e., overall events and unambiguous events). The aim is to compare the accuracy of interpretation for each group and the factors influencing their interpretations. The performance of users and UX experts will first be compared to all the 89 events. However, based on the results found previously, we also need to investigate unambiguous events in order to test our first hypotheses. Lastly, the effect of emotion will be investigated in order to identify if emotions influence the quality of users and UX experts' interpretations. Indeed, negative affect has been known to have a stronger impact on the user's experience than positive affect (Baumeister et al., 2001; Miron-Shatz et al., 2009). Therefore, we want to explore whether the user's affective responses influence both groups' quality of interpretation.

Accuracy of Interpretations of User's Emotional Experience

In the first section, we compare users' and UX experts' ability to accurately assess users' emotional experience for all the 89 sequences. In order to evaluate both groups quality of interpretations, we compared their reported levels of arousal and valence on the Affective Slider (Betella & Verschure, 2016) with their psychophysiological data. In the table below, the lines refer to psychophysiological data categorized as either positive events (i.e., high emotional valence and high arousal) or negative events (i.e., low emotional valence and high arousal). Moreover, the columns represent the type of interpretations by users and UX experts (i.e., positive, negative, or neutral) according to the thresholds presented previously.

Table 4. Comparative Analysis of Users and UX experts' Performances on Assessment of User's Emotional Experience

	Negative Interpretation	Neutral Interpretation	Positive Interpretation	Total
Overall Experience				
Users' Performance				
Negative Event	3 6.67%	33 73.33%	9 20%	45
Positive Event	6 13.64%	28 63.64%	10 22.73%	44
Total	9 10.11%	61 68.54%	19 21.35%	89
UX experts Performance				
Negative Event	18 40%	27 60%	0	45
Positive Event	0	23 52.28%	21 47.73%	44
Total	18 20.22%	50 56.18%	21 23.59%	89
Unambiguous Events				
Users' Performance				
Negative Event	1 7.14%	10 71.43%	3 21.43%	14
Positive Event	0	8 66.67%	4 33.33%	12
Total	1 3.85%	18 69.23%	7 26.92%	26
UX experts Performance				
Negative Event	7 50%	7 50%	0	14
Positive Event	0	8 66.67%	4 33.33%	12
Total	7 26.92%	15 57.69%	4 15.38%	26

In order to compare both groups' quality of interpretation, a Fisher's test investigates whether there is a significant difference between users and UX experts' interpretations and the psychophysiological data. This difference indicates whether one group can significantly assess the user's experience. According to the "Overall Experience" section of the table above, UX experts seem to better perform in identifying accurately positive events (47.73%) and negative events (40%) compared to users who performed less well for both positive events (22.73%) and negative events (6.67%). Indeed, UX experts can accurately assess the user's emotional state ($p < 0.0001$) whereas users do not seem able to ($p = 0.5366$). However, given the low coefficient of inter-rater reliability of UX experts for both valence (0.329) and arousal (0.223) for the overall events, it is not possible to conclude a clear difference in the skills of these two groups. Therefore, we need to explore whether this tendency remains relevant for events with a high level of inter-rater reliability (i.e., unambiguous events).

In this second section, we compare the accuracy of interpretation of users and UX experts for unambiguous events. The statistical test still is Fisher's Test in order to identify the significant convergence between both groups' interpretations and psychophysiological data. According to the "Unambiguous Events" section of Table 4, there seems to be no significant convergence between user's interpretation of valence and arousal and psychophysiological data ($p = 0.8238$). Thus, users do not seem able to accurately report levels of emotional valence and arousal for positive events (33.33%) and negative events (7.14%). Furthermore, UX experts still seem able to accurately assess the user's emotional state of unambiguous events ($p = 0.0026$). Indeed, UX experts seem to have interpretations of sequences that are in line with the emotions experienced by users during the real events. Based on unambiguous events, UX experts have a success rate of interpretations of 50% for negative experience and 33.33% for positive experience. Thus, given these results, H1 is supported, but only for events with no ambiguity.

Moreover, in order to compare both group's ability to assess the user's emotional state, we need to determine which group is more likely than the other to accurately interpret events. Therefore, a logistic regression was performed to model the probability of accurately interpret psychophysiological data between users and UX experts. Hence, the independent variable refers to the individual evaluating the sequence (e.g., user and UX expert) and the dependent variable refers to the correctness of interpretation.

Table 5. Comparison Between Users and UX Experts' Quality of Interpretation of Reported Valence and Arousal

	Estimate	Std Err	t-value	p-value	Alpha	Reference
Overall Experience						
UX experts Affective Slider	1.6348	0.7485	2.18	**0.0343**	0.05	Users Affective Slider
Unambiguous Events						
UX experts Affective Slider	1,1249	0,6365	1.77	**0,0833**	0,1	Users Affective Slider

Based on these results, UX experts seem to be more accurate in their interpretations of the user's experience in comparison to users. Indeed, UX experts perform significantly better than users for the evaluation of the overall experience (p = 0.0343). Based on the 23 sequences of unambiguous events, UX experts are closer to accurately interpret psychophysiological events than users do (p = 0.0833), and the difference between interpretations is almost significant.

Lastly, recency effect was also evaluated as potential influence on the user's quality of interpretation. Hence, we compare both group's ability to assess the user's emotional state for sequences over time in order to determine if this construct can impact the results found previously. A logistic regression modelling the probability of accurately evaluating the sequence is performed using time and the individual evaluating as independent variables and the correctness as a dependent variable. It allows to see if the recency effect could influence the results shown previously.

Table 6. Recency Effect on Accuracy of Users' Interpretations

	Estimate	StdErr	t-Value	p-value	Alpha	Reference
UX experts Affective Slider	-0,00023	0,00116 5	-0,2	**0,8459**	0,05	Users Affective Slider

The results show that there is no recency effect influencing the conclusions shown previously. Indeed, there is no influence on the fact that UX experts are more likely to accurately interpret the psychophysiological points than users do. Moreover, the quality of users' interpretations of emotions ($p = 0.6956$) does not seem to change over time, both in terms of valence ($p = 0.4487$) and arousal ($p = 0.445$). Thus, there is no pattern of recency effect on the user's evaluation of emotions for all 89 sequences. The results are not affected by this retrospective bias.

All things considered, H1 is partially supported, because it is only supported in cases where there is no ambiguity (i.e., there is a common agreement on the interpretation of the user's intention). Indeed, for unambiguous events, UX experts can accurately evaluate the user's emotion while the user is not able to do so. Lastly, the concept recency does not influence the tendency for which UX experts are more precise in their interpretations than users. Thus, H1 is supported in a context with no ambiguity, given all this evidence.

Furthermore, in order to test H2 regarding users and UX experts' ability to accurately interpret emotions, we need to investigate if the type of moment – negative moment or positive moment – affects the quality of interpretations. Interpretations of emotions by users and UX experts are compared to psychophysiological data.

Table 7. Comparison Between Users and UX Experts' Quality of Interpretation for Positive and Negative Events

	Estimate	Std Err	t-value	p-value	Alpha	Reference
Overall Experience						
Negative Event						
UX experts Affective Slider	2.6391	1.1802	2.24	**0.0614**	0.05	Users Affective Slider
Positive Event						
UX experts Affective Slider	0.539	1.0494	0.51	**0.6102**	0.05	Users Affective Slider
Unambiguous Events						
Negative Event						
UX experts Affective Slider	2.5649	1.1673	2.2	**0.0658**	0.1	Users Affective Slider
Positive Event						
UX experts Affective Slider	2.31E-17	0.866	0	**1**	0.1	Users Affective Slider

Results show that UX experts are even more likely to accurately interpret events than users when the moment is negative rather than positive. Indeed, this tendency is observed for the overall experience when we compare the p-value of negative events ($p = 0.0614$) with positive events ($p = 0.6102$). Also, the conclusions are the same for unambiguous events. Thus, UX experts seem better able to report user's experience than users, and that tendency is stronger for negative events rather than positive events. Hence, H2 is rejected since users do not seem able to evaluate negative moments more accurately than positive moments whereas the ability of UX experts to accurately evaluate user experience seems to be more acute for negative than for positive moments.

6 Discussion

In this study, we compared the users and UX experts' accuracy of interpretations when watching specific sequences from the user's experience. We found that, for events with no ambiguity, UX experts seem to be accurate when evaluating the user experience whereas users do not seem to be. We explored several factors that might influence our results and found that UX experts tend to be more accurate in their interpretation of negative events.

In the HCI literature, one of the main goals is to gain a deeper understanding of the reasoning behind the user's decision-making. With the increased attention for UX design, HCI research has started to focus towards a deeper understanding of the user's experience and has put emphasis on the user journey (Barbier et al., 2018). Consequently, different techniques and methods were developed in order to gain insights into the inner user experience. The study presented in this article aims to investigate the contribution of users and UX experts into the analysis of the user's experience by comparing the accuracy of interpretation between both parties during positive and negative experiences. Our results show that, for unambiguous events, UX experts seem able to accurately assess the user's emotional experience whereas users don't seem able to accurately self-report levels of emotional valence and arousal. This finding suggests that UX experts can understand and interpret with accuracy the user's experience for events where they have a common understanding of the user's intentions. This finding is consistent with various studies addressing the ability of impartial individuals and users in the evaluation of the user's experience (White & Younger, 1988; Robins & John, 1997; Mehu & Scherer, 2012; Dmello, 2016). Indeed, impartial individuals can detect the obvious actions and behaviours of users when it is easily visually perceivable (e.g., facial expressions) (Mehu & Scherer, 2012). However, less observable behaviours such as intentions are accessible mainly to the self (i.e., users) and are perceived as potentially more insightful (White & Younger, 1988; Robins & John, 1997). Therefore, we expect UX experts to have accurate interpretations of emotions for events where they can contextualize and understand the user's intentions. Our research also suggests that users do not seem able to accurately evaluate their experience when asked to do so retrospectively. This result is congruent with a study by Giroux-Huppé et al. (2019) which states that less than 25% of pain points (negative moments) were identified qualitatively by the participant. Moreover, a comparison between both groups' ability to assess user experience allowed us to conclude that UX experts seem closer to accurately interpret events than users and that tendency is more pronounced for negative events than positive events. Hence, these results demonstrate the usefulness of psychophysiological data, as participants are not able to report these data on their own. Surprisingly, the recency effect had no influence on the accuracy of the user's interpretations. Thus, we can conclude that UX experts seem more reliable than users in their interpretations of the user's emotional experience during unambiguous events.

This study contributes to the existing user experience literature by further investigating the contribution of users and UX experts into the analysis of the user experience. Thus, providing more precision into the importance and abilities of UX experts and users into the process. In sum, users are not able to recall accurately on their own experience and UX experts seem better equipped to interpret the user's emotional experience accurately during unambiguous events. Therefore, relying heavily on users for the revelation of insights might encompass more biases into the interpretation of results.

Our research also contributes methodologically to the retrospective recall of user's affective states by using jointly the psychophysiological measures along with situated recall method in order to test their convergent validity. Several studies analyze the congruency between psychophysiological measures and what a user recall from that experiment (Bentley et al., 2005; Bruun et al., 2016; Lourties et al., 2018) or the efficiency of a recall method as a complementary tool to psychophysiological measures (Lim, 2002). However, little research focuses on the user's accuracy of interpretations on their own experience in comparison to a group control of UX experts.

Lastly, throughout this paper, the aim is to define the user's and UX expert's ability to accurately interpret the user's experience. Our results show that users don't seem able to assess retrospectively their emotional state during peaks when using a simulated recall method. This finding conflicts with past studies suggesting that users are able to successfully report their emotions during peaks (Fredrickson & Kahneman, 1993; Kahneman et al., 1993).

Our results also involve managerial implications for the UX industry. Indeed, while several approaches encourage user involvement to evaluate the user experience, our paper shows that UX experts, characterized as impartial individuals who are familiar with UX, such as UX practitioners, can evaluate the user's emotional experience without involving users. Therefore, this information can influence the approach used to analyze the user experience as UX experts require additional time to analyze the results.

Furthermore, negative peak events have shown to be more accurately interpreted by UX experts in comparison to positive peak events. Consequently, the purpose of a UX research is to optimize the user’s experience, notably through the identification of problematic elements in order to fulfill the user’s need and the improvement of the overall experience (Beauregard & Corriveau, 2007) (Beauregard & Corriveau, 2007; Hassenzahl, 2008). By being more accurate in the interpretation of negative events, UX experts are more than qualified to evaluate the user experience in order to meet the UX testing objectives.

Limits and Avenues of Research

The study presented in this paper has several limitations. Also, the nature of the tasks completed by the users allowed low fluctuations in terms of arousal which could have possibly influence user’s remembrance of past events. Since the tasks are utilitarian, they result in a smaller variation in the emotional intensity experienced in comparison to hedonic tasks such as video games. Indeed, an overall higher level of arousal, which translates into greater emotional intensity, in all events could affect positively users’ ability to accurately interpret events. Therefore, the conclusions of our study can only be generalized to utilitarian tasks with low activation and not all kinds of tasks. Moreover, a large majority of user interpretations are categorized as neutral (i.e., users are not able to categorize the sequence as positive or negative). In addition, users had to evaluate all sequences, which may have influenced the categorization of the sequences. Indeed, users who were unable to identify or recall the emotion experienced could have potentially categorized the moment as neutral. Therefore, adding the option of not commenting on the interpretation would be interesting in future research in order to have more accurate evaluations in the user's memory.

After addressing such limitations, a deeper investigation into the UX expert’s ability to evaluate users’ emotions would be interesting in future research. Indeed, throughout our research, UX experts seem able to accurately assess the user experience. However, interpretation of intentions remained a challenge with a minority of events being categorized as unambiguous. Therefore, future work should be done on tools and methods allowing a better evaluation of intentions from the UX expert’s point of view.

7 Conclusion

The evaluation of user experience remains a topic of investigation and the study presented in this paper explores the roles of users and UX experts into the analysis of such experience. We explored the process of evaluation of the user journey within the HCI literature by testing users and UX experts’ ability to accurately interpret and evaluate emotions for positive and negative experience. Our results show that users are not able to self-assess their affective state when shown the most intense (high arousal) moments from their own experience. Also, UX experts seem able to accurately assess the user’s emotional experience and that tendency is even stronger for negative experience. Overall, this suggests that UX experts are more accurate than users when evaluating and interpreting users’ emotions. We foresee that psychophysiological measures will provide researchers the necessary richness of data in order to understand the user journey.

References

Ariely, D. (1998). Combining Experiences over Time: The Effects of Duration, Intensity Changes and on-Line Measurements on Retrospective Pain Evaluations. *Journal of Behavioral Decision Making*, 11, 11-45.

Bailey, B. P., Adamczyk, P. D., Chang, T. Y., & Chilson, N. A. (2006). A Framework for Specifying and Monitoring User Tasks. *Computers in Human Behavior*, 22(4), 709-732. doi:10.1016/j.chb.2005.12.011

Bailey, B. P., & Konstan, J. A. (2006). On the Need for Attention-Aware Systems: Measuring Effects of Interruption on Task Performance, Error Rate, and Affective State. *Computers in Human Behavior*, 22(4), 685-708. doi:10.1016/j.chb.2005.12.009

Barbier, C., Guyonvarc'h, L., Guillaume, A., & Tattegrain, H. (2018). Is the Self-Confrontation Method Applicable to Naturalistic Driving Studies? *Safety Science*. doi:10.1016/j.ssci.2018.11.005

Bargas-Avila, J. A., & Hornbaek, K. (2011). Old Wine in New Bottles or Novel Challenges? A Critical Analysis of Empirical Studies of User Experience. Paper presented at the *CHI*, Vancouver BC, Canada.

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad Is Stronger Than Good. *Review of General Psychology*, 5(4), 323-370. doi:10.1037/1089-2680.5.4.323

Beauregard, R., & Corriveau, P. (2007). User Experience Quality: A Conceptual Framework for Goal Setting and Measurement. Paper presented at the *International Conference on Digital Human Modeling*.

Bentley, T., Johnson, L., & Baggo, K. v. (2005). Evaluation Using Cued-Recall Debrief to Elicit Information About a User's Affective Experiences. Paper presented at the *OZCHI*, Canberra, Australia.

Bernardin, H. J., & Buckley, M. R. (1981). Strategies in Rater Training. *Academy of Management Review*, 6(2), 205-212.

Betella, A., & Verschure, P. F. (2016). The Affective Slider: A Digital Self-Assessment Scale for the Measurement of Human Emotions. *PLoS ONE*, 11(2), e0148037. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/26849361. doi:10.1371/journal.pone.0148037

Bloom, L. C., & Watkins, M. J. (1999). Two-Component Theory of the Suffix Effect : Contrary Findings. *Journal of experimental psychology*, 25(6), 1452-2474.

Borkenau, P., & Ostendorf, F. (1987). Retrospective Estimates of Act Frequencies: How Accurately Do They Reflect Reality? *Journal of Personality and Social Psychology*, 52(3), 626-638. doi:10.1037/0022-3514.52.3.626

Boucsein, W. (2012). Electrodermal Activity: Springer.

Bradley, M. M., Greenwald, M. K., Petry, M. C., & Lang, P. J. (1992). Remembering Pictures: Pleasure and Arousal in Memory. *Journal of experimental psychology*, 18(2), 379-390.

Bradley, M. M., & Lang, P. J. (1994). Measuring Emotion: The Self-Assessment Manikin and the Semantic Differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25(1), 49-59. doi:10.1016/0005-7916(94)90063-9

Braithwaite, J. J., Watson, D. G., Jones, R., & Rowe, M. (2013). A guide for analysing electrodermal activity (EDA) & skin conductance responses (SCRs) for psychological experiments. *Psychophysiology*, 49(1), 1017-1034.

Bruun, A., & Ahm, S. (2015). Mind the Gap! Comparing Retrospective and Concurrent Ratings of Emotion in User Experience Evaluation. *Human-Computer Interaction – Interact* 2015 (pp. 237-254).

Bruun, A., Law, E. L.-C., Heintz, M., & Alkly, L. H. A. (2016). Understanding the Relationship between Frustration and the Severity of Usability Problems. Paper presented at the Proceedings of the 2016 *CHI Conference on Human Factors in Computing Systems* - CHI '16.

Burton-Jones, A., & Gallivan, M. J. (2007). Toward a Deeper Understanding of System Usage in Organizations : A Multilevel Perspective. *MIS Quarterly*, 31(4), 657-679.

Cacioppo, J. T., & Tassinary, L. G. (1990). Principles of Psychophysiology: Physical, Social, and Inferential Elements: Cambridge University Press.

Charles, R. L., & Nixon, J. (2019). Measuring Mental Workload Using Physiological Measures: A Systematic Review. *Applied Ergonomics*, 74, 221-232. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/30487103. doi:10.1016/j.apergo.2018.08.028

Chittenden, A. H. R. (2002). A Pastoral Care Teacher's Theory of Action, Interactive Thinking and Effective Teaching Practice. *Pastoral Care in Education*, 20(1), 3-10. doi:10.1111/1468-0122.00214

Cockburn, A., Quinn, P., & Gutwin, C. (2017). The Effects of Interaction Sequencing on User Experience and Preference. *International Journal of Human-Computer Studies*, 108, 89-104. doi:10.1016/j.ijhcs.2017.07.005

Colombetti, G. (2005). Appraising Valence. *Journal of Consciousness Studies*, 12(8-10), 103-126.

Courtemanche, F., Fredette, M., Senecal, S., Leger, P.-m., Dufresne, A., Georges, V., & Labonte-lemoyne, E. (2019). Method of and System for Processing Signals Sensed from a User. In: Google Patents.

De Gelder, B., & Vroomen, J. (1997). Modality Effects in Immediate Recall of Verbal and Non-Verbal Information. *European Journal of Cognitive Psychology*, 9(1), 97-110. doi:10.1080/713752541

Dimoka, A. (2010). What Does the Brain Tell Us About Trust and Distrust? Evidence from a Functional Neuroimaging Study. *MIS Quarterly*, 34(2), 373-396.

Dimoka, A., Pavlou, P. A., & Davis, F. D. (2011). Neurois: The Potential of Cognitive Neuroscience for Information Systems Research. *Information Systems Research*, 22(4), 687-702. doi:10.1287/isre.1100.0284

Dirican, A. C., & Göktürk, M. (2011). Psychophysiological Measures of Human Cognitive States Applied in Human Computer Interaction. *Procedia Computer Science*, 3, 1361-1367. doi:10.1016/j.procs.2011.01.016

Dmello, S. K. (2016). On the Influence of an Iterative Affect Annotation Approach on Inter-Observer and Self-Observer Reliability. *IEEE Transactions on Affective Computing*, 7(2), 136-149. doi:10.1109/taffc.2015.2457413

Duric, Z., Gray, W. D., Heishman, R., Li, F., Rosenfeld, A., Schoelles, M. J., Wechsler, H. (2002). Integrating Perceptual and Cognitive Modeling for Adaptive and Intelligent Human-Computer Interaction. *Proceedings of the IEEE*, 90(7).

Ekman, P. (1965). Differential Communication of Affect by Head and Body Cues. *Journal of Personality and Social Psychology*, 2(5), 726.

Ekman, P. (1973). Cross-Cultural Studies of Facial Expression. Darwin and facial expression: A century of research in review, 169222(1).

Ekman, P., & Friesen, W. V. (2003). In M. Books (Ed.), Unmasking the Face: A Guide to Recognizing Emotions from Facial Clues.

Finkenauer, C., & Rimé, B. (1998). Socially Shared Emotional Experiences Vs. Emotional Experiences Kept Secret: Differential Characteristics and Consequences. *Journal of Social and Clinical Psychology*, 17(3), 295-318.

Forne, M. (2012). Physiology as a Tool for UX and Usability Testing. (Media Technology Master), Royal Institute of technology,

Giroux-Huppé, C., Sénécal, S., Fredette, M., Chen, S. L., Demolin, B., & Léger, P.-M. (2019). Impact D'une Expérience Psychophysiologique Négative Sur La Satisfaction Des Consommateurs Dans Un Contexte D'épicerie En Ligne. (Master), HEC Montréal, Montréal.

Glanzer, M., & Cunitz, A. R. (1966). Two Storage Mechanisms in Free Recall. *Journal of verbal learning and verbal behavior*, 5(4), 351-360.

Goldman, R., Pea, R., Barron, B., & Derry, S. (2007). Video Research in the Learning Sciences. Mahwah, NJ: Lawrence Erlbaum Associates.

Gosling, S. D., John, O. P., Craik, K. H., & Robins, R. W. (1998). Do People Know How They Behave? Self-Reported Act Frequencies Compared with on-Line Coding by Observers. *Journal of Personality and Social Psychology*, 74(5), 1337-1349.

Guérin, J., Riff, J., & Testevuide, S. (2004). Étude De L’activité « Située » De Collégiens En Cours D’eps : Une Opportunité Pour Examiner Les Conditions De Validité

Des Entretiens D'autoconfrontation. *Revue française de pédagogie*, 147(1), 15-26. doi:10.3406/rfp.2004.3116

Guinea, A. O. d., & Markus, L. (2009). Why Break the Habit of a Lifetime? Rethinking the Roles of Intention, Habit, and Emotion in Continuing Information Technology Use. *MIS Quarterly*, 33(3), 433-444.

Guinea, A. O. d., Titah, R., & Léger, P.-M. (2014). Explicit and Implicit Antecedents of Users' Behavioral Beliefs in Information Systems: A Neuropsychological Investigation. *Journal of Management Information Systems*, 30(4), 179-210. doi:10.2753/mis0742- 1222300407

Guinea, A. O. d., & Webster, J. (2013). An Investigation of Information Systems Use Patterns: Technological Events as Triggers, the Effect of Time and Consequences for Performance. MIS Quarterly, 37(4), 1165-1188.

Halford, W. K., & Sanders, M. R. (1990). The Relationship of Cognition and Behavior During Marital Interaction. *Journal of Social and Clinical Psychology*, 9(4), 489-510.

Hamann, S. (2001). Cognitive and Neural Mechanisms of Emotional Memory. *TRENDS in Cognitive Sciences*, 5(9).

Harré, R., & Secord, P. F. (1972). The Explanation of Social Behaviour. Oxford, England.

Harrigan, J. A., Rosenthal, R., & R.Scherer, K. (2005). The New Handbook of Methods in Nonverbal Behavior Research.

Hartson, R., & Pyla, P. (2012). The UX Book: Process and Guidelines for Ensuring a Quality User Experience.

Hassenzahl, M. (2008). User Experience (UX): Towards an Experiential Perspective on Product Quality. Paper presented at the *IHM*.

Hassenzahl, M., & Tractinsky, N. (2006). User Experience - a Research Agenda. *Behaviour & Information Technology*, 25(2), 91-97. doi:10.1080/01449290500330331

Hermans, H. J. M. F., Riet;de Groot, Roel;Nauta, Jan F. (1990). Self-Confrontation as a Method for Assessment and Intervention. *Journal of Counseling and Development*, 69(2), p.156.

Hetland, A., Vitterso, J., Oscar Bo Wie, S., Kjelstrup, E., Mittner, M., & Dahl, T. I. (2018). Skiing and Thinking About It: Moment-to-Moment and Retrospective Analysis of Emotions in an Extreme Sport. *Frontiers Psychology*, 9, 971. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/29973894. doi:10.3389/fpsyg.2018.00971

Hockenbury, D. H., & Hockenbury, S. E. (2016). Discovering Psychology: Worth Publishers.

Hollan, J. D., & Hutchins, E. L. (2009). Opportunities and Challenges for Augmented Environments: A Distributed Cognition Perspective. *Designing User Friendly Augmented Work Environments* (pp. 237-259).

Ivonin, L., Chang, H.-M., Díaz, M., Català, A., Chen, W., & Rauterberg, M. (2014). Beyond Cognition and Affect: Sensing the Unconscious. *Behaviour & Information Technology*, 34(3), 220-238. doi:10.1080/0144929x.2014.912353

Jarodzka, H., Scheiter, K., Gerjets, P., & van Gog, T. (2010). In the Eyes of the Beholder: How Experts and Novices Interpret Dynamic Stimuli. *Learning and Instruction*, 20(2), 146-154. doi:10.1016/j.learninstruc.2009.02.019

John, O. P., & Robins, R. W. (1994). Accuracy and Bias in Self-Perception: Individual Differences in Self-Enhancement and the Role of Narcissim. *Journal of Personality and Social Psychology*, 66(1), 206-219.

Kahneman, D., Frederickson, B. L., Schreiber, C. A., & Redelmeier, D. A. (1993). When More Pain Is Preferred to Less : Adding a Better End.

Kahneman, D., & Riis, J. (2005). Living, and Thinking About It : Two Perspectives on Life. *The Science of Well-Being* (pp. 285-304): Oxford University Press.

Kecklund, G., & Akerstedt, T. (2004). Apprehension of the Subsequent Working Day Is Associated with a Low Amount of Slow Wave Sleep. *Biological Psychology*, 66(2),169-176.Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/15041138. doi:10.1016/j.biopsycho.2003.10.004

Kolar, D. W., Funder, D. C., & Colvin, C. R. (1996). Comparing the Accuracy of Personality Judgments by the Self and Knowledgeable Others. *Journal of Personality*, 64(2), 311-337. doi:10.1111/j.1467-6494.1996.tb00513.x

Krosnick, J. A. (1999). Survey Research. Annual Review of Psychology, 50, 537-567.

Laeng, B., Sirois, S., & Gredeback, G. (2012). Pupillometry: A Window to the Preconscious? *Perspectives on Psychological Science*, 7(1), 18-27. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/26168419. doi:10.1177/1745691611427305

Lahlou, S. (2010). Digitization and Transmission of Human Experience. *Social Science Information*, 49(3), 291-327. doi:10.1177/0539018410372020

Lahlou, S. (2011). How Can We Capture the Subject's Perspective? An Evidence-Based Approach For the Social Scientist. *Social Science Information*, 50(3-4).

Léger, P.-M., Courtemanche, F., Fredette, M., & Sénécal, S. (2019). A Cloud-Based Lab Management and Analytics Software for Triangulated Human-Centered Research, Cham.

Léger, P.-M., Davis, F. D., Cronan, T. P., & Perret, J. (2014). Neurophysiological Correlates of Cognitive Absorption in an Enactive Training Context. *Computers in Human Behavior*, 34, 273-283. doi:10.1016/j.chb.2014.02.011

Lewis, S., Pea, R., & Rosen, J. (2010). Beyond Participation to Co-Creation of Meaning: Mobile Social Media in Generative Learning Communities. *Social Science Information*, 49(3), 351-369. doi:10.1177/0539018410370726

Lim, S. (2002). The Self-Confrontation Interview: Towards an Enhanced Understanding of Human Factors in Web-Based Interaction for Improved Website Usability. *Journal of Electronic Commerce Research*, 3(3).

Lindgren, E. (2002). The Effect of Stimulated Recall on 14-Year-Olds' Swedish and Efl Writing and Revision. *Language Teaching Research*, 6(3), 267-268.

Lourties, S., Léger, P.-M., Sénécal, S., Fredette, M., & Chen, S. L. (2018a). Testing the Convergent Validity of Continuous Self-Perceived Measurement Systems: An Exploratory Study. Paper presented at the *HCI International*, Las Vegas, NV, USA.

Lourties, S., Léger, P.-M., Sénécal, S., Fredette, M., & Chen, S. L. (2018b). Testing the Convergent Validity of Continuous Self-Perceived Measurement Systems: An Exploratory Study. Paper presented at the *HCI International*, Las Vegas, NV, USA.

Lyddon, W. J., Yowell, D. R., & Hermans, H. J. M. (2006). The Self-Confrontation Method: Theory, Research, and Practical Utility. Counselling Psychology Quarterly, 19(1), 27-43. doi:10.1080/09515070600589719

Lyle, J. (2003). Stimulated Recall: A Report on Its Use in Naturalistic Research. British *Educational Research Journal*, 29(6), 861-878. doi:10.1080/0141192032000137349

MacKay, D. G., Shafto, M., Taylor, J. K., Marian, D. E., Abrams, L., & Dyer, J. R. (2004). Relations between Emotion, Memory, and Attention: Evidence from Taboo Stroop, Lexical Decision, and Immediate Memory Tasks. *Memory & Cognition*, 32(3), 474-488.

Maia, C. L. B., & Furtado, E. S. (2016). A Study About Psychophysiological Measures in User Experience Monitoring and Evaluation.

Mandryk, R. L., Inkpen, K. M., & Calvert, T. W. (2006). Using Psychophysiological Techniques to Measure User Experience with Entertainment Technologies. *Behaviour & Information Technology*, 25(2), 141-158. doi:10.1080/01449290500331156

Markey, R., & Reichheld, F. (2011). Introducing : The Net Promoter System.

Mehu, M., & Scherer, K. R. (2012). A Psycho-Ethological Approach to Social Signal Processing. *Cognitive Processes*, 13 Suppl 2, 397-414. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/22328016. doi:10.1007/s10339-012-0435-2

Miron-Shatz, T., Stone, A., & Kahneman, D. (2009). Memories of Yesterday's Emotions: Does the Valence of Experience Affect the Memory-Experience Gap? Emotion, 9(6), 885-891. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/20001131. doi:10.1037/a0017823

Murdock, B. B. (1962). The Serial Position Effect of Free Recall. *Journal of experimental psychology*, 64(5), 482-488.

Nisbett, R. E., & Wilson, T. D. (1977). Telling More Than We Can Know: Verbal Reports on Mental Processes. *Psychological Review*, 84(3), 231.

Oh, I. S., Wang, G., & Mount, M. K. (2011). Validity of Observer Ratings of the Five-Factor Model of Personality Traits: A Meta-Analysis. Journal of Applied Psychology, 96(4), 762-773. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/21142341. doi:10.1037/a0021832

Omodei, M. M., & McLennan, J. (1994). Studying Complex Decision Making in Natural Settings: Using a Head-Mounted Video Camera to Study Competitive Orienteering. *Perceptual and Motor Skills* (79), 1411-1425.

Picard, R. W. (1995). Affective Computing. Perceptual Computing Section Technical Report No. 321, Cambridge, Massachusetts, M.I.T. Media Lab

Riedl, R., & Léger, P.-M. (2016). Fundamentals of Neurois. Studies in Neuroscience, *Psychology and Behavioral Economics*. Springer, Berlin, Heidelberg.

Robins, R. W., & John, O. P. (1997). Effects of Visual Perspective and Narcissism on Self-Perception: Is Seeing Believing? *Psychological Science*, 8(1), 37-42.

Robins, R. W., & John, O. P. (1997). The Quest for Self-Insight: Theory and Research on Accuracy and Bias in Self-Perception. In N. Y. A. Press (Ed.), Handbook of Personality Psychology (pp. 649-679).

Rosenthal, R. (1966). Expetimenter Effects in Behavioral Research. Appleton-Century-Crofts. New York.

Russell, J. A. (1980). A Circumplex Model of Affect. *Journal of Personality and Social Psychology*, 39(6), 1161-1178.

Scherer, K. R. (1984). Emotion as a Multicomponent Process: A Model and Some Cross-Cultural Data. *Review of personality & social psychology*.

Schooler, J. W. (2011). Introspecting in the Spirit of William James: Comment on Fox, Ericsson, and Best (2011). *Psychological Bulletin*, 137(2), 345-350. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/21355632. doi:10.1037/a0022390

Schwarz, N. (2007). Retrospective and Concurrent Self-Reports: The Rationale for Real-Time Data Capture. The science of real-time data capture: Self-reports in health research, 11, 26.

Smith, A. (1997). Human Computer Factors : A Study of Users and Information Systems.

Stern, R. M., Stern, R. M., Ray, W. J., & Quigley, K. S. (2001). Psychophysiological Recording: Oxford University Press, USA.

Tesser, A., & Campbell, J. (1983). Self-Definition and Self-Evaluation Maintenance. Psychological perspectives on the self, 2, 1-31.

Thomas, D. L., & Diener, E. (1990). Memory Accuracy in the Recall of Emotions. *Journal of Personality and Social Psychology*, 59(2), 291-297.

Uyl, M. J. d., & Kuilenburg, H. v. (2005). The Facereader : Online Facial Expression Recognition. *Proceedings of Measuring Behavior*.

Van Gog, T., Paas, F., van Merrienboer, J. J., & Witte, P. (2005). Uncovering the Problem-Solving Process: Cued Retrospective Reporting Versus Concurrent and Retrospective Reporting. *Journal of Experimental Psychology Applied*, 11(4), 237-244. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/16393033. doi:10.1037/1076- 898X.11.4.237

Vermeeren, A. P. O. S., Roto, V., & Väänänen, K. (2015). Design-Inclusive UX Research: Design as a Part of Doing User Experience Research. *Behaviour & Information Technology*, 35(1), 21-37. doi:10.1080/0144929x.2015.1081292

Vicente, K. J., Thornton, D. C., & Moray, N. (1987). Spectral Analysis of Sinus Arrhythmia: A Measure of Mental Effort. *Human Factors*, 29(2), 171-182.

Ward, R. D., & Marsden, P. H. (2003). Physiological Responses to Different Web Page Designs. *International Journal of Human-Computer Studies*, 59(1-2), 199-212. doi:10.1016/s1071-5819(03)00019-3

Watkins, M. J. (1974). Concept and Measurement of Primary Memory. *Psychological Bulletin*, 81(10), 695-711.

Waugh, N. C., & Norman, D. A. (1965). Primary Memory. Psychological Review, 72.

White, P. A., & Younger, D. P. (1988). Differences in the Ascription of Transient Internal States to Self and Other. *Journal of Experimental Social Psychology*, 24(4), 292-309.

Wilson, G. M., & Sasse, M. A. (2004). From Doing to Being: Getting Closer to the User Experience. *Interacting with Computers*, 16, 697-705. doi:10.1016/j.intcom.2004.06.001

Zanting, A., Verloop, N., & Vermunt, J. D. (2001). Student Teachers Eliciting Mentors' Practical Knowledge and Comparing It to Their Own Beliefs. *Teaching and Teacher Education*, 17, 725-740.

Zijlstra, F. R. H., Roe, R. A., Leonora, A. B., & Krediet, I. (1999). Temporal Factors in Mental Work: Effects of Interrupted Activities. *Journal of Occupational and Organizational Psychology*, 72, 163-1

Chapitre 2 : Deuxième Essai

Visualizing a User's Cognitive and Emotional Journeys: A FinTech Case

Marianne Veilleux[1], Sylvain Sénécal[1], Bertrand Demolin[1], Frédérique Bouvier[1], Marie-Laure Di Fabio[2], Constantinos Coursaris[1] and Pierre-Majorique Léger[1],

[1] HEC Montréal, Montréal, H3T 2A7, Canada

[2] Mouvement Desjardins, Montréal, H5B 1B4, Canada

marianne.veilleux@hec.ca

Abstract. In this article, we propose a visualization approach that presents the user's cognitive and emotional states in conjunction with the actual journey of the user on a web interface. Specifically, we have designed a new visualization method which contextualizes the user's physiological and behavioral data while interacting with a web-based information system in the financial services industry. The proposed approach brings together the user's behavior with his/her cognitive and emotional states to produce a rich overview of his/her experience. Combining these methods produces key insights into the user experience and facilitates an understanding of the evolution of the experience since it highlights where the user was on the interface when s/he experienced a given cognitive and emotional state. Results from an illustrative case suggest that the proposed visualization method is useful in conveying where participants deviate from the optimal path and facilitates the identification of usability issues on web interfaces.

Keywords: Interface design; Web design; Cognitive effort; Affective responses; Consumer behavior

1 Introduction

Customer experience (CX) has become a central concept for many businesses. Experience became a critical element when purchasing a product or service (Rust & Lemon, 2001; Gopalani, 2011). Hence, creating positive experiences allows brands to attract and retain customers (Pine & Gilmore, 2011). The measurement of customer experience plays a key role in making insights actionable for firms and helps them remain competitive. Practitioners and academics have developed several tools and methods to understand and manage the customer experience (De Keyser et al., 2015). These tools and methods focus mainly on the development of visualization and prototyping techniques as well as the implementation of measurement tools (Stickdorn et al., 2011; De Keyser et al., 2015). The practice of CX management is characterized by the processes used to monitor and organize a series of interactions between the organization and its customers. Thus, visualization and prototyping methods allow identifying key moments of the customer experience that, once addressed, will provide a more valuable global experience.

Customer journey mapping (Browne, 2012; Rawson et al., 2013), service blueprinting (Bitner et al., 2008; Patrício et al., 2008) and customer experience mapping (Verma et al., 2012) are among the most frequently used visualization methods. These techniques are used to visualize and structure the process that the consumer works through during their interaction with the firm (Bitner et al., 2008). Most studies use the service blueprint methodology, and few go beyond this approach in order to analyze the customer journey (Bitner et al., 2008; Lemon & Verhoef, 2016; Rosenbaum et al., 2017). There is a need to improve customer experience measurement tools by focusing on qualitative techniques, such as narrative data collection, to seek more detailed insights (Helkkula et al., 2012).All in all, progress is needed in customer journey mapping in order to identify key opportunities to influence customer experience.

By deepening the understanding of the customer experience, it strengthens the understanding of overall customer satisfaction. New technologies and data represent interesting assets to be integrated in the mapping process. Researchers have begun to incorporate neuroscientific measures (e.g., emotions and cognitive load) in order to track more precisely concurrent measures of customer experience (Venkatraman et al., 2012; Lewinski, 2015; Plassmann et al., 2015). Scholars and practitioners commonly agree that the customer experience includes cognitive, emotional, behavioral, sensorial, and social dimensions (Schmitt, 1999; Verhoef et al., 2009; vom Brocke et al., 2020). Measures such as biometrics, eye-tracking and electrocardiography (ECG) help to better understand how the customer experience is formed and the series of events that led to the overall level of satisfaction (Venkatraman et al., 2012). By understanding how affective and cognitive variables influence the customer in his/her journey, it contributes to identifying key drivers to overall customer satisfaction. Our study aims to propose an approach that presents, in conjunction with the actual journey of a user on a web interface, his/her implicit cognitive and emotional states along this journey. Users can deviate from the predicted path, and identifying these deviations can help select the source of intervention or influence needed on a given interface to prevent users from deviating (Lemon & Verhoef, 2016). To illustrate the relevance of the developed method, we present a methodology based on a case study of a given task and the results obtained using the visualization technique. The proposed method is discussed and validated by a group of User Experience (UX) experts to ensure its relevance and usefulness.

2 Literature Review

2.1 The Customer Journey

The analysis of a customer's journey was first used in the areas of service management and multichannel management (Neslin et al., 2006; Bitner et al., 2008). Customer journeys are the most popular visualization techniques within the domain of service design (Segelström, 2013). It refers to a series of events that the consumer goes through in order to be informed, to buy or to interact with a given organization (Norton & Pine, 2013). It is a visual and chronological representation of the events experienced by the user (Halvorsrud et al., 2016). The purpose of such an approach is to simulate a "walk in the customer's shoes" (Holmlid & Evenson, 2008). The customer journey analysis allows identifying critical touch points throughout the customer journey that have the most significant impact on customer outcomes (Temkin, 2010; Lemon & Verhoef, 2016). The goal of such technique is to improve customer interaction by enhancing each touch point between the customer and the organization. A study by Lemon & Verhoed (2016) states that the understanding the customer journey leads to better insights about the customer experience. Indeed, the customer experience encompasses the customer journey with a given organization over time through multiple touch points (Lemon & Verhoef, 2016). Thus, the customer journey approach allows analyzing the touch points between the organizations and its customers.

Customer journey mapping (CJM) is among the most frequently used visualization methods of the customer experience (Browne, 2012; Rawson et al., 2013). This method is a visual representation of the customer experience using a given service (Stickdorn et al., 2011; Marquez et al., 2015). This map is a consolidated overview of all sequences of events through which customers must pass through to complete their purchase (Rosenbaum et al., 2017). CJM establishes all touch points between the firm, and its customer during a given purchasing process. The aim of such technique is to improve the quality of the overall customer experience by improving the customer experience associated with each touch point (Rosenbaum et al., 2017). Mapping the customer experience offers information from the customer's perspective and represents a great way towards operational improvements (Tseng et al., 1999).

Mapping is a commonly used tool in the service design methodology (Marquez et al., 2015). However, a study by Rosenbaum (2017) mentions that CJM assumes that all customers of a given organization experience the same touch points and assigns the same importance to each of these touch points. Hence, the understanding of the participants' overall experience does not necessarily reflect discreet experiences with specific phases, events, or activities. Currently, there is a lack of methods that combine the users' behaviors with their affective and cognitive states. To date, some studies have incorporated neuroscientific data into the analysis of the customer journey in order to have a richer understanding (Venkatraman et al., 2012; Lewinski, 2015; Plassmann et al., 2015).

2.2 Measuring Emotional and Cognitive Responses

Psychophysiological data are physical signals measured in real time, which are generated due to psychological changes (Dirican & Göktürk, 2011). These measures allow researchers to assess the user's reaction so a specific stimulus (Riedl & Léger, 2016). Several tools are available to measure psychophysiological measures, such as electrocardiography (ECG), skin-based measures including electrodermal activity (EDA), ocular measures, brain measures including electroencephalogram (EEG), respiration rates, and blood pressure (Charles & Nixon, 2019). Furthermore, psychophysiological measures are unobtrusive, which allows collecting data about the user's experience without affecting their decision-making (Dirican & Göktürk, 2011; Guinea & Webster, 2013; Guinea et al., 2014; Leger et al., 2014). These measures therefore allow for more natural reactions from the user while also offering uninterrupted reports of emotions (Kahneman & Riis, 2005; Guinea et al., 2014; Lourties et al., 2018). The advantage of having an uninterrupted report of emotions is that it gives access to the unconscious emotional reactions of users (Ivonin et al., 2014). Indeed, research has shown that interrupting users while they are completing a task often leads to biased results (Zijlstra et al., 1999; Bailey et al., 2006; Bailey & Konstan, 2006; Lourties et al., 2018). Psychophysiological measures offer a more complete view of the human-computer interaction and allow assessing events of cognitive and emotional relevance to the users (Picard, 1995; Ward & Marsden, 2003; Bentley et al., 2005; Giroux-Huppé et al., 2019).

In the UX literature, emotions are often defined according to two complementary dimensions, namely valence and arousal (Guinea & Markus, 2009; Léger et al., 2014). The circumplex model of affect is a very popular dimensional approach to defining several emotional states according to these two psychophysiological constructs (Russell, 1980). In addition, valence varies from pleasure to displeasure (Boucsein, 2012; Forne, 2012; Guinea & Webster, 2013), which is equivalent to a variation from positive to negative emotions (Colombetti, 2005; Maia & Furtado, 2016). Valence can be defined as "what a user feels" and the most reliable way to measure such constructs is with facial coding (Burton-Jones & Gallivan, 2007). Individuals tend to express their emotions with micro-movements of the facial muscles (Uyl & Kuilenburg, 2005). Arousal is the second and complementary dimension to the circumplex model of affect. This construct refers to “level of arousal which ranges from calm to excited (Boucsein, 2012; Maia & Furtado, 2016). Thus, psychophysiological measures translate emotions into measurable constructs.

Moreover, cognition, which refers to the process of reasoning and the mental effort required to understand and complete a given task (Haapalainen et al., 2010) is also a measurable physiological construct (Charland et al., 2015). Several studies have shown that pupil dilation is a significant measure of cognitive load (Beatty & Lucero-Wagoner, 2000; Wilson, 2002; Ikehara & Crosby, 2005; Iqbal et al., 2005; Léger et al., 2018; Desrochers et al., 2019). Indeed, the degree of the pupil’s dilation correlates with the workload of the task (Beatty & Lucero-Wagoner, 2000). Hence, pupil size offers an immediate measure of cognitive effort.

A study by Gentile et al. (2007) suggests that customers interpret information from the interface from a cognitive and affective point of view which leads to a personal impression of the given website. Epstein’s Cognitive Experiential Self-Theory (CEST) suggests that there are two systems which operate in parallel when exposed to a stimulus event. These two systems are affective and cognitive. Thus, combining cognitive and emotional measures provides a comprehensive understanding of the user's experience. The customer journey map illustrates every touch point the consumer has with the company in order to complete a given task (Richardson, 2010). Hence, by adding affective and cognitive measures to the customer journey, it allows UX researchers to identify critical interface elements influencing the user’s journey. Thus, the convergence of these three perspectives (i.e., emotions, cognition and behavior) affords a complete view, which few methods allow (Coursaris & Kim, 2011). Therefore, we pose the following proposition:

P1: Enriching customer journey visualizations with information about the cognitive and emotional states of the user facilitates the identification of critical touch points shaping customer experience.

3 Method

The study is presented in two phases and consists of a collection of user tests and a focus group. The user tests were performed to collect data for the purpose of developing an illustrative case of the customer experience of all users. The focus group was conducted to validate the research proposal. First, we collected psychophysiological data about each user's experience in order to track their cognitive and emotional responses during an e-commerce interaction. The goal was to develop a method which triangulates behavioral data with psychophysiological data in order to gain key insights into the user journey and to produce a complete overview of the user's experience. Thus, we present a new visualization method that contextualizes the user's physiological and behavioral data while interacting with a web-based information system in the financial services industry, an example of FinTech. We illustrate the methodology by presenting in detail one task and the results obtained using the visualization method. Second, the new visualization method is reviewed by a group of experts in order to confirm whether it enriches contemporary understanding of critical touch points shaping customer experience, and therefore offers support for the research proposition.

Phase 1

3.1 Participants

A total of 38 participants (42% female) were recruited in this study and their age ranged from 23 to 62 years old. Prior to the study, participants were asked to rate the financial institution on the Net Promoter Score (NPS) Scale (Reichheld & Markey, 2011). NPS is a one-question metric used to assess a customer's overall perception of a brand using a Likert scale from 0 to 10, with 0 being not at all likely to recommend the brand, and 10 being extremely likely to. To reduce the potential influence of brand equity (either extremely high or low) on the participants' experience, only participants who rated the financial institution between a score of 3 to 8 were invited to participate in our study. Each participant received a moderate financial compensation to participate. To participate, users needed normal vision and were screened out for glasses, laser eye surgery, astigmatism, epilepsy, neurological and psychiatric diagnoses.

3.2 User Test Procedure

The test sessions took place at a university usability laboratory based on standard practice for UX enriched with psychophysiological measures (Alvarez et al., 2019). Upon arrival, participants were asked to complete a consent form, after which they were informed about the purpose of the study and fitted with physiological sensors. Participants were told that their participation was requested in order to evaluate a financial website. Each key segment of the financial institution (i.e., early retirees, midlife adults, youth, and entrepreneurs) was assigned a set of tasks specific to their group. To control the task type, each participant was assigned to a total of nine tasks completed frequently on the website. The nine tasks were grouped evenly into three main categories representing frequent actions taken by users: searching for specific information on the website, using a tool to find answers, and find ways to contact the financial institution. This approach minimizes the effect of the task type by having a varied number of different tasks. Participants were asked to complete tasks for which an optimal navigation path was first established by the authors vis-à-vis the smallest number of clicks necessary to complete a task.

3.3 Measures

For the user test, data were collected using a variety of non-intrusive tools. Behavioral data were captured from the recording of the screen interface allowing the user's actions to be monitored. A Tobii X-60 eye-tracker (Stockholm, Sweden) sampled at 60 Hz was used to capture participant eye movements and Tobii Studio was used to record the experience. The screen recording along with the eye-tracking device allows having the user’s perspective by identifying precisely where the participant was looking at every second (Vasseur et al., 2019). Moreover, pupil size measurement allowed for tracking each user’s cognitive load (Van Gerven et al., 2004; Hyönä et al., 2007). Arousal was measured using electrodermal activity (EDA) (Hassenzahl & Tractinsky, 2006; Boucsein, 2012; Braithwaite et al., 2013) with the AcqKnowledge software at MP150 sampled at 500 Hz (BIOPAC, Goleta, USA). EDA represents an indication of the variation of physiological arousal (Hassenzahl et al., 2003). In fact, it measures the skin conductance, which measures current flow between two points of skin contact after an electrical potential has been applied to them (Braithwaite et al., 2013). Hence, it indicates cognitive and emotional stimulation throughout the experience (Stern et al., 2001). Furthermore, valence was measured using facial expressions (Ekman & Friesen, 2003; Uyl & Kuilenburg, 2005) performed with the FaceReaderTM Software (Noldus, Wageningen, Netherlands). Indeed, this tool is used to observe facial reaction and assess emotional valence with a score ranging from -1 (most negative emotions) to +1 (most positive emotions) and a temporal precision of 30 inferences per second (Ekman & Friesen, 2003; Uyl & Kuilenburg, 2005). More specifically, the valence score is calculated as the intensity of a happy score minus the intensity of the highest negative responses (Lojiens & Krips, 2018). Then, Observer XT and CubeHX (Courtemanche et al., 2019; Léger et al., 2019) were used to synchronize all the signals between Acknowledge and FaceReader (Noldus, Wageningen, Netherlands) as recommended by Léger et al. (2014).

3.4 Preparation of Visualization Models

In order to triangulate collected data, webpages visited are first grouped and assembled to construct each user's journey consisting of its various navigation paths. Using the eye-tracking device and Tobii Studio, these tools help contextualize the behavioral data of users and group all the webpages used for a given task. This compilation leads to a visual representation of all paths used by the sample for a given task. Moreover, using the complete psychophysiological dataset generated by all users, an average of each measure (i.e., arousal, valence, and cognitive load) was calculated for each webpage; this generated an overview of the page-by-page evolution of the users' cognitive and emotional states. Moreover, the averages for each measure are grouped by webpage, tasks and segments to get an overview of an average customer journey for a given segment.

In order to facilitate the understanding of all collected data, the averages obtained from both cognitive and emotional measures are translated into symbols and visual codes according to pre-established thresholds. Arousal, emotional valence and cognitive load are rescaled (-1 to +1) and adjusted to the baseline. The baseline allows distinguishing the user's psychophysiological sensitivity since each participants' emotional reaction to a stimulus does not have the same direction and intensity (Lourties et al., 2018). Hence, the baseline allows participants' emotional responses to be standardized on the same scale.

First, the user's emotional responses are translated into four main categories of the circumplex model: enthusiasm, frustration, serenity and tolerance (Posner et al., 2005). Indeed, all affective states are characterized by two fundamental neurophysiological constructs which are valence and arousal (Russell, 1980). Consequently, Russell (1989) proposed the Affect Grid, which is a scale based on the circumplex model of affect in order to describe emotions according to the crossing of the dimensions of valence on the horizontal axis and arousal on the vertical axis (see Fig. 1). The center of the grid represents a neutral valence and a medium level of arousal. For example, when the valence value is below 0, it refers to the lower part of the Affect Grid (i.e., tolerance and serenity). Moreover, when the valence is below 0, it refers to the left side of the Affect Grid (i.e., frustration and tolerance). Thus, when the arousal and emotional valence are below 0, the user ends up in the quadrant of tolerance.

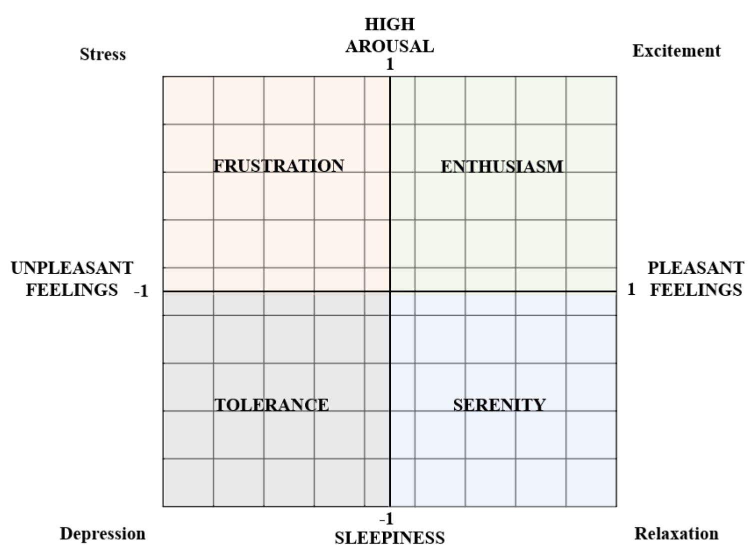

Fig. 1. The Affect Grid adapted from Russell et al. (1989)

With the Affect Grid, it is possible to analyze whether the participant feels pleasant/unpleasant or active/inactive. Hence, the symbolization of emotions on the visualization models respects the coding of the Affect Grid. Therefore, each quadrant (i.e., frustration, enthusiasm, tolerance and serenity) is represented by an icon of facial expression. Hence, it is easy to visually recognize what emotion is felt on average by users for each webpage.

Second, cognitive loads are categorized into three levels such as easy, moderate or difficult and these levels are represented using a gear icon (e.g., one gear symbolizes a relatively easy task whereas three gears represent a difficult task). Since the arousal variable varies from -1 to 1, the difference is divided into three equal parts to distinguish the different levels of difficulty. Given that the arousal values are rescaled between -1 and 1, a cognitive load value of less than zero (i.e., inferior to the baseline) means that cognitive effort is low. Thus, cognitive loads were divided into three main categories which are low cognitive load (< -0.3), moderate cognitive load (between 0.3 and -0.3) and high cognitive load (>0.3) (See Table 1).

The thresholds for cognitive loads and emotional responses are presented in the table below (see Table 1). Such visual codes allow simplifying the visualization of the page-by-page evolution of the user’s cognitive and emotional states.

Table 1. Thresholds for Psychological Data.

Cognitive load	**Low cognitive load**	**Medium cognitive load**		**High cognitive load**
	< -0.3	Between -0.3 and 0.3		> 0.3
Circumplex Model of Affect	**Frustration**	**Enthusiasm**	**Tolerance**	**Serenity**
	EDA > 0 Valence < 0	EDA > 0 Valence > 0	EDA < 0 Valence < 0	EDA < 0 Valence > 0

Phase 2

3.5 Focus Group Procedure

Once all tasks were completed, two visualization models were created using psychophysiological data with the objective of representing the overall experience of users. These two models were then discussed in a focus group with UX experts. Experts are impartial individuals selected on a voluntary basis, who understand usability and have experiences in the analysis of the user journey. The focus group is a collective interview where participants meet to discuss a defined topic (Lallemand & Gronier, 2016). Since our approach aims at facilitating the UX expert's analysis of the customer journey, the focus is on the acceptance of this system by ensuring it meets UX experts' needs. The type of focus group selected is called "prioritizing functionalities", as the main objective is to identify the most attractive functionalities for UX experts in order to guide and optimize the method design.

The focus group procedure lasted 90 minutes from the reception to the closing of the interview. A total of six (6) UX experts participated. The number of participants in a focus group is recommended to be between six and 10 individuals (Debus, 2007), as a smaller number of participants promote interdependence among members (Anzieu & Martin, 1968). First, participants were informed about the purpose of the focus group, the activities planned and their roles as UX experts. Then, UX experts had to evaluate two visualization models for which they were asked a series of questions. UX experts used the technique of associating ideas to express what first comes to mind when looking as the visualization model (Debus, 2007). UX experts were then asked to state their analysis of the customer journey. By stating their understanding, it would reveal whether the model represents an appropriate and relevant analysis tool. Afterwards, UX experts were asked to elicit both strengths and weaknesses for each visualization scheme in order to identify which features to keep or remove. Lastly, following the evaluation of both visualization models, UX experts were asked to define what an ideal visualization scheme should be according to the two models shown previously. Once the focus group was finished, the visualization model chosen by UX experts as the best option is modified according to their recommendations and sent back to UX experts in order to obtain their final impressions.

4 Method

4.1 Customer Journey Visualization

Two models were designed and discussed in this study. Each model is a distinctive visual representation of a task using a variety of pre-established visual codes and symbols. This study aims at comparing both models in order to define the optimal approach to understand the concurrent user experience. Through the representation process, there are seven key elements found in each model which support the representation of customer journeys. These elements are presented in the table below and the visual representation of some elements vary between both models (see Table 2). Indeed, the representation of the optimal path, which refers to the optimal navigation vis-à-vis the smallest number of clicks necessary to complete a task, is represented differently between the two models. The optimal path layout varies across models. It is not aligned or centered in the first model but is in the second model. Also, individual customer journeys are displayed on the second model while the first model does not distinguish between each customer journey.

Table 2. Elements Found in Each Visualization Model.

	Model 1 (see Figure 2)	**Model 2** (see Figure 3)
Sunny Path	The desired path is **coloured** to highlight the path that users are supposed to follow.	The desired path is **coloured, aligned and centered** to highlight the path that users are supposed to follow.
Physiological data	physiological data are represented by icons and placed at the top of the webpage.	
Action	The place where the clicked user is highlighted with a dot.	
Traffic	Traffic is highlighted and the size of the circle is proportional to the number of users who have used this avenue.	
Success and Failures	The endpoints are illustrated with **green** (when the user ends on the desired page) and **red** (when he does not end on this page) circles.	
Customer Journey	No distinction is made for individual paths, only sees all webpages visited.	Coloured pastilles allow to distinguish individual customer journeys.

The first visualization scheme is presented in Figure 2. All pages are aligned in order to provide a structured representation of the different pages used. The most problematic pages can be identified with the indicators of emotional value and cognitive effort. Indeed, these pages are shown in red and represent frustration (key located in top-left box) and high cognitive effort (top-right box) respectively. Moreover, traffic and used paths are indicated in blue to emphasize this information. In short, this model aims to highlight travel and psychophysiological data.

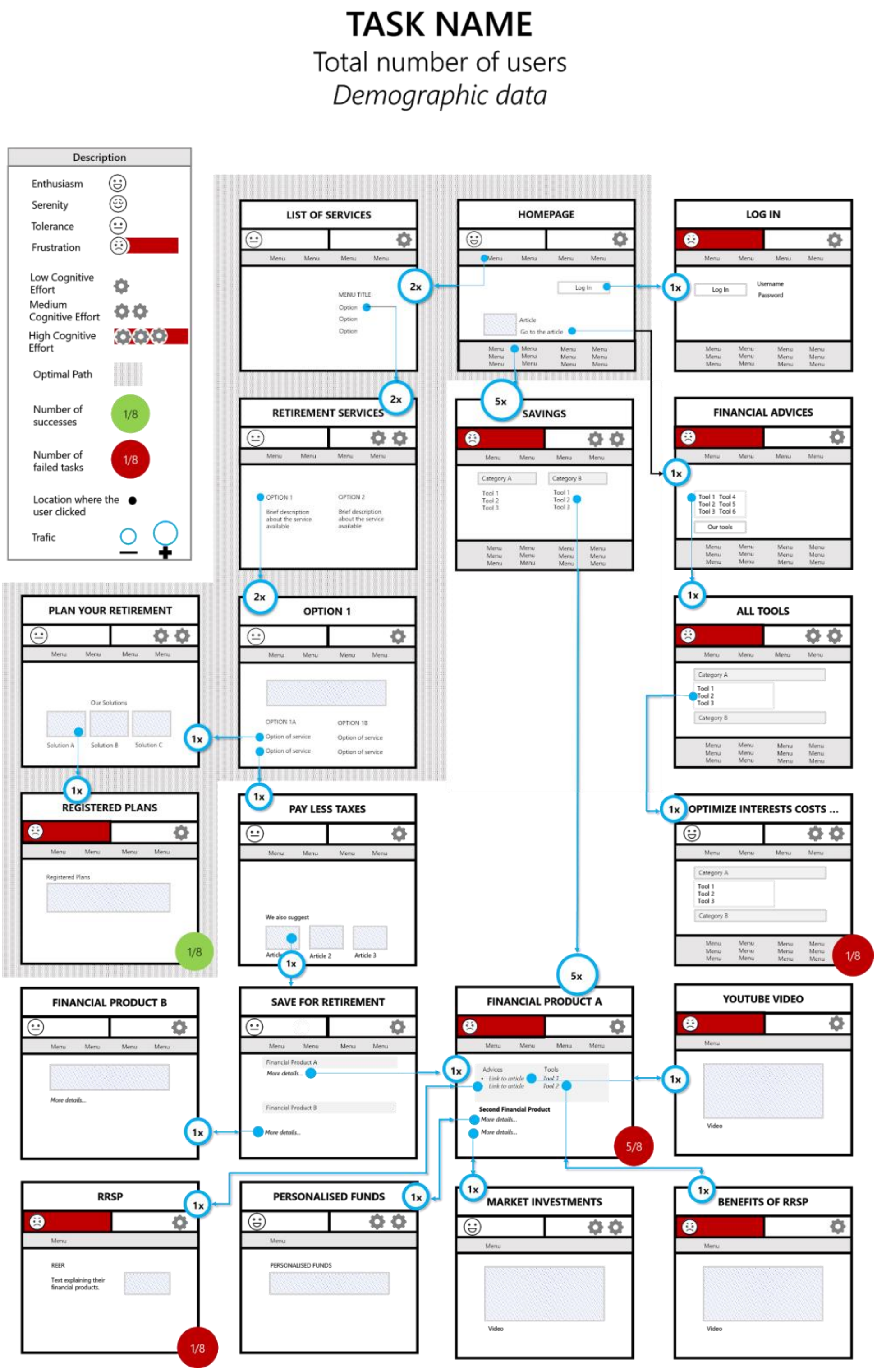

Fig. 2. First Visualization Model.

The second visualization model is presented in Figure 3. Unlike the first model, the journeys are represented in the form of swim lanes, where psychophysiological data (i.e., cognitive load, emotional valence, and arousal) are associated with each webpage visited. The most problematic areas are identified using the same indicators of emotional valence and cognitive effort. Indeed, the problematic areas are illustrated by the red zones and represent frustration (left zone) and high cognitive effort (right zone). In Figure 3, several pages seem to be problematic regarding negative emotional valence. However, no page seems to require high cognitive effort. Since the paths are linear, it is possible to see where participants have deviated from the original goal and where problems have emerged. For example, most users used the link at the bottom of the homepage to go to the "Savings" page, deviating from the expected route.

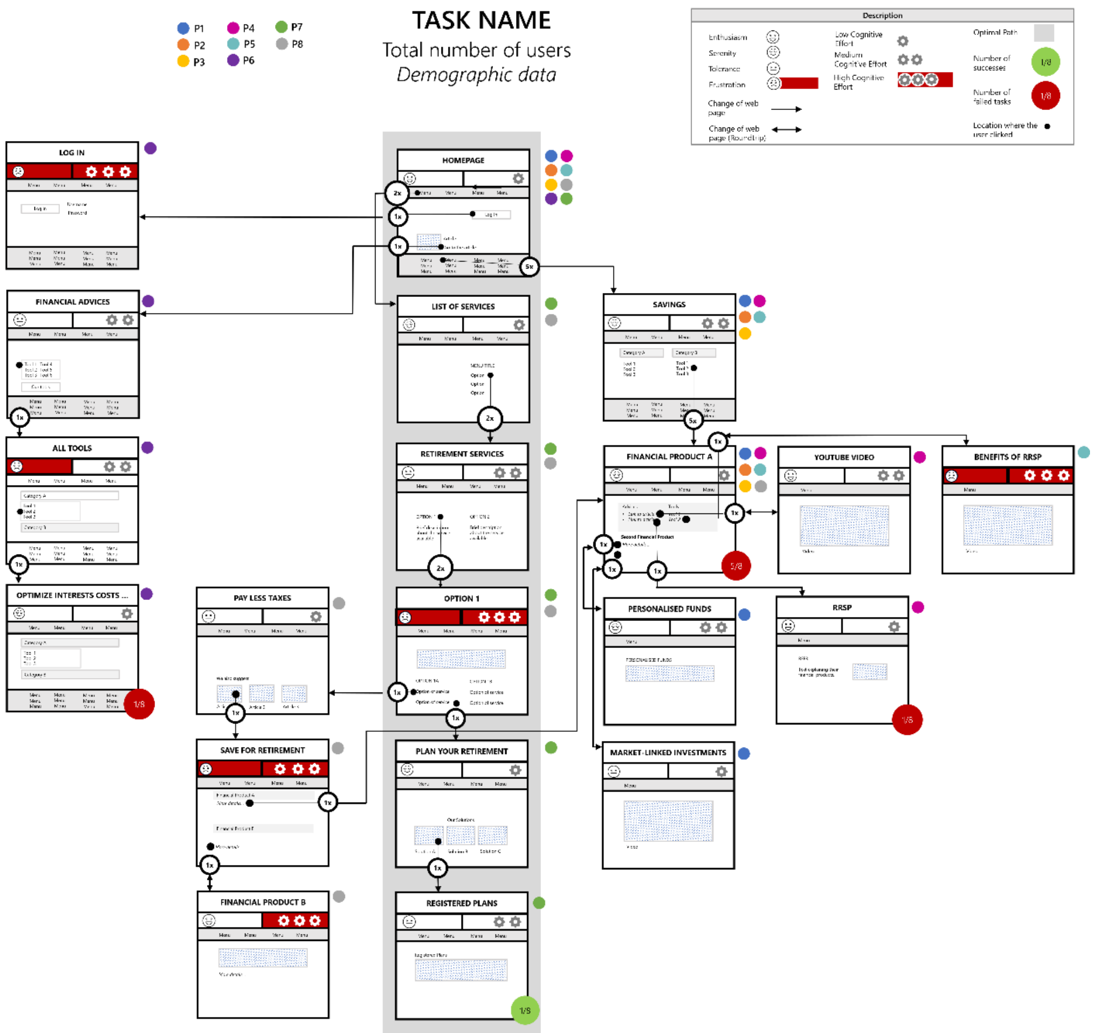

Fig. 3. Second Visualization Model.

4.2 Results of the Focus Group

Regarding the first visualization model, the UX experts mentioned that the beginning of the experiment is not clearly indicated and thus, the starting point is difficult to identify. Moreover, the fact that the optimal path is not aligned and centered makes the reading of the customer journey less intuitive. In general, UX experts find this model not sufficiently refined and the understanding of the users' experience is not obvious. According to the participating UX experts, this model seems suitable for client reporting through a matrix format. This first visualization model allows identifying problematic areas on the interface.

Regarding the second visualization model, the UX experts mention that the model allows for a better understanding of the sequence experienced by users. Indeed, as mentioned previously, the first model does not distinguish between each journey. Thus, UX experts are more likely to be interested in physiological data than in the context surrounding the use of the webpage. In short, the second method allows us to have more context around the physiological data in order to prioritize the source of intervention on the interface to improve the overall experience. For example, the "log in" page seems problematic according to the first model, because it indicates frustration. However, looking through the second model, it is possible to see that this page has only been viewed by one participant. Hence, this frustration can be caused by several factors and the “log in” page may not be a priority when updating the website. Therefore, UX experts suggest that model 2 represents a good tool to have a summary of the situation and can also serve as a benchmark to measure progress through improvements made on the website.

The results of the Focus Group suggest that the research proposition be accepted since the experts mention that the visualization models facilitate the identification of problematic areas on the interface. Indeed, by bringing together the cognitive and emotional data of the users, it allows for an enriched customer journey and a more complete understanding of the associated experience.

5 Method

Our results suggest, through an illustrative case, that the use of the proposed method facilitates the identification of critical touch points to customer experience. This model is useful in order to have an overview of the experience and different paths taken by users. It allows one to see the problematic webpages when diagnosing the customer journey. Following the focus group, the UX experts agreed that the second model is the more appropriate tool for customer journey analysis. The visual representation has been described as clear and simple to use. A UX expert mentioned that "the red color makes it easier to read problematic pages and the optimal path is well highlighted without cluttering up the parallel paths" (UX expert 5). Moreover, this model has been selected because it has several strengths. First, this model is self-supporting, which means that it presents a "simple and quick visualization, without the need for explanation" (UX expert 3). Secondly, the information presented is clear, thanks to "a clear legend and the use of icons and colors that are easy to understand" (UX expert 2). It is easy to identify successes and failures as well as areas of frustration or high cognitive load. Finally, this model also allows following the customer journey of each user thanks to the color tablets. It gives an overview of the experience while differentiating the various paths. This model is also useful for understanding the main errors made during experience sequence (i.e., to understand where participants deviated from the optimal path). Simply put, the model allows to quickly identify where participants are getting lost.

However, the model also has a limitation. Indeed, it is easy to analyze for a certain number of participants. The visual representation represents the paths of 8 participants. However, the larger the number of participants, the more complex the visual representation becomes. The use of color tablets allows the distinction between participants, but leads to limits in terms of sample size.

Our study contributes to the UX literature by presenting a comprehensive method allowing for the visualization of a user’s emotions and behaviors during his/her navigation on a website. Thus, this study adds to the literature which focuses on the modelling of the consumer decision journey (Rosenbaum et al., 2017).

Additionally, it provides more precision into the analysis and the interpretation of results, aiding to reveal problematic areas on the interface. This method serves as a complementary approach to other methods available such as questionnaires and interviews that enable collection of self-reported data and adding the convergence of these three perspectives (i.e., emotions, cognition, and behaviors), which few methods allow (Coursaris & Kim, 2011). The relative simplicity of this method, which enables the visual representation of the evolution of a user's cognitive and emotional states throughout their online journey and experience, should be particularly useful to both UX researchers and practitioners. Modelling service delivery from the customer's perspective is an important topic for service providers seeking to improve their services (Halvorsrud et al., 2016).

Our method allows contrasting several user journeys against the planned journey. We contend that most customer journey maps can potentially be critically flawed. They assume all customers of an organization experience the same organizational touch points and view these touch points as equally important (Rosenbaum et al., 2017). In contrast, our method provides an accurate report of several user journeys. The results also pose managerial implications. First, this new method allows both practitioners and researchers to identify psychophysiological pain points on a webpage easily and the visualization helps to analyze and interpret results more efficiently (Lamontagne et al., 2019). Moreover, our method is useful for comparing user experiences on various interfaces, which can be used to compare the user experience of a specific task on competing interfaces.

Furthermore, some limitations need to be acknowledged. First, the method focuses on one section of the website. Hence, the evaluation of the customer experience does not reflect that of the entire website but depends on the selected task. Second, this model is limited to a certain number of users. As there were only 8 individuals user journey studied, this is not a large-scale study, mostly due to the high cost of obtaining the data. As the number of users studied increases, there will be different user journeys. In this way, the visualization becomes more complex to analyze due to the large amount of information. Thus, it would be interesting to study the possibility of creating segments from user navigation when the sample size is too large to distinguish each path using a colored pad.

6 Conclusion

Our results show that the models proposed adequately identified and interpreted psychophysiological data supporting an understanding of the user's experience. By representing the various customer journeys through the webpages visited, it makes it possible to have a real report of the experience lived by users. Using this new visualization method generates a complete overview of users' experience and produces key insights. Indeed, it facilitates the understanding of the evolution of the experience since it shows critical touch points of the interface where the user experienced a given cognitive and emotional state. It also helps to identify the main differences between the planned customer journey and the user's decision-making. This method serves as a complementary approach to other methods available such as questionnaires and interviews that enable the collection of self-reported data and adds the convergence of these three perspectives (i.e., emotions, cognition, and behaviors), which few methods allow (Coursaris & Kim, 2011).

References

Alvarez, J., Brieugne, D., Léger, P.-M., Sénécal, S., & Frédette, M. (2019). Towards Agility and Speed in Enriched UX Evaluation Projects. *Human-Computer Interaction*: IntechOpen.

Anzieu, D., & Martin, J.-Y. (1968). La Dynamique Des Groupes Restreints: Presses Universitaires de France.

Bailey, B. P., Adamczyk, P. D., Chang, T. Y., & Chilson, N. A. (2006). A Framework for Specifying and Monitoring User Tasks. *Computers in Human Behavior*, 22(4), 709-732. doi:10.1016/j.chb.2005.12.011

Bailey, B. P., & Konstan, J. A. (2006). On the Need for Attention-Aware Systems: Measuring Effects of Interruption on Task Performance, Error Rate, and Affective State. *Computers in Human Behavior*, 22(4), 685-708. doi:10.1 016/j.chb.2005.12.009

Beatty, J., & Lucero-Wagoner, B. (2000). The Pupillary System. Handbook of *psychophysiology*, 2(142-162).

Beauregard, R., & Corriveau, P. (2007). User Experience Quality: A Conceptual Framework for Goal Setting and Measurement. Paper presented at the *International Conference on Digital Human Modeling*.

Bentley, T., Johnson, L., & Baggo, K. v. (2005). Evaluation Using Cued-Recall Debrief to Elicit Information About a User's Affective Experiences. Paper presented at the *OZCHI*, Canberra, Australia.

Bitner, M. J., Ostrom, A. L., & Morgan, F. N. (2008). Service Blueprinting: A Practical Technique for Service Innovation. *California Management Review*, 50(3), 66-94.

Boucsein, W. (2012). Electrodermal Activity: Springer.

Bradley, M. M., & Lang, P. J. (1994). Measuring Emotion: The Self-Assessment Manikin and the Semantic Differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25(1), 49-59. doi:10.1016/0005-7916(94)90063-9

Braithwaite, J. J., Watson, D. G., Jones, R., & Rowe, M. (2013). A guide for analysing electrodermal activity (EDA) & skin conductance responses (SCRs) for psychological experiments. *Psychophysiology*, 49(1), 1017-1034.

Browne, J. (2012). Customer Journey Mapping : What Is It For? Retrieved from *Forrester Research Inc*.

Burton-Jones, A., & Gallivan, M. J. (2007). Toward a Deeper Understanding of System Usage in Organizations : A Multilevel Perspective. *MIS Quarterly*, 31(4), 657-679.

Charland, P., Léger, P.-M., Sénécal, S., Courtemanche, F., Mercier, J., Skelling, Y., & Labonté-Lemoyne, E. (2015). Assessing the Multiple Dimensions of Engagement to Characterize Learning: A Neurophysiological Perspective. *JoVE (Journal of Visualized Experiments)* (101), e52627.

Colombetti, G. (2005). Appraising Valence. *Journal of Consciousness Studies*, 12(8-10), 103-126.

Coursaris, C. K., & Kim, D. J. (2011). A Meta-Analytical Review of Empirical Mobile Usability Studies. *Journal of Usability Studies*, 6(3), 117-171.

Courtemanche, F., Fredette, M., Senecal, S., Leger, P.-m., Dufresne, A., Georges, V., & Labonte-lemoyne, E. (2019). Method of and System for Processing Signals Sensed from a User. In: Google Patents.

De Keyser, A., Lemon, K. N., Klaus, P., & Keiningham, T. L. (2015). A Framework for Understanding and Managing the Customer Experience. *Marketing Science Institute working paper series*, 15(121), 1-48.

Debus, M. (2007). Handbook for Excellence in Focus Group Research: Academy for Educational Development.

Desrochers, C., Léger, P.-M., Fredette, M., Mirhoseini, S., & Sénécal, S. (2019). The Arithmetic Complexity of Online Grocery Shopping: The Moderating Role of Product Pictures. *Industrial Management & Data Systems*.

Dirican, A. C., & Göktürk, M. (2011). Psychophysiological Measures of Human Cognitive States Applied in Human Computer Interaction. *Procedia Computer Science*, 3, 1361-1367. doi:10.1016/j.procs.2011.01.016

Ekman, P., & Friesen, W. V. (2003). In M. Books (Ed.), Unmasking the Face : A Guide to Recognizing Emotions from Facial Clues.

Forne, M. (2012). Physiology as a Tool for UX and Usability Testing. (Media Technology Master), Royal Institute of technology,

Giroux-Huppé, C., Sénécal, S., Fredette, M., Chen, S. L., Demolin, B., & Léger, P.-M. (2019). Impact D'une Expérience Psychophysiologique Négative Sur La Satisfaction Des Consommateurs Dans Un Contexte D'épicerie En Ligne. (Master), HEC Montréal, Montréal.

Gopalani, A. (2011). The Service-Enabled Customer Experience: A Jump-Start to Competitive Advantage. *Journal of Business Strategy*, 32(3), 4-12. Retrieved from https://doi.org/10.1108/02756661111121947. doi:10.1108/02756661111121947

Guinea, A. O. d., & Markus, L. (2009). Why Break the Habit of a Lifetime? Rethinking the Roles of Intention, Habit, and Emotion in Continuing Information Technology Use. *MIS Quarterly*, 33(3), 433-444.

Guinea, A. O. d., Titah, R., & Léger, P.-M. (2014). Explicit and Implicit Antecedents of Users' Behavioral Beliefs in Information Systems: A Neuropsychological Investigation. *Journal of Management Information Systems*, 30(4), 179-210. doi:10.2753/mis0742-1222300407

Guinea, A. O. d., & Webster, J. (2013). An Investigation of Information Systems Use Patterns : Technological Events as Triggers, the Effect of Time and Consequences for Performance. *MIS Quarterly*, 37(4), 1165-1188.

Haapalainen, E., Kim, S., Forlizzi, J. F., & Dey, A. K. (2010). Psycho-Physiological Measures for Assessing Cognitive Load. Paper presented at the Proceedings of the *12th ACM international conference on Ubiquitous computing*.

Halvorsrud, R., Kvale, K., & Følstad, A. (2016). Improving Service Quality through Customer Journey Analysis. *Journal of Service Theory and Practice*, 26(6), 840-867. doi:10.1108/jstp-05-2015-0111

Hassenzahl, M., Burmester, M., & Koller, F. (2003). Attrakdiff: Ein Fragebogen Zur Messung Wahrgenommener Hedonischer Und Pragmatischer Qualität. *Mensch & Computer 2003* (pp. 187-196): Springer.

Hassenzahl, M., & Tractinsky, N. (2006). User Experience - a Research Agenda. *Behaviour & Information Technology*, 25(2), 91-97. doi:10.1080/01449290500330331

Helkkula, A., Kelleher, C., & Pihlström, M. (2012). Characterizing Value as an Experience: Implications for Service Researchers and Managers. *Journal of Service Research*, 15(1), 59-75.

Holmlid, S., & Evenson, S. (2008). Bringing Service Design to Service Sciences, Management and Engineering. In Service Science, *Management and Engineering Education for the 21st Century* (pp. 341-345): Springer.

Hyönä, J., Tommola, J., & Alaja, A.-M. (2007). Pupil Dilation as a Measure of Processing Load in Simultaneous Interpretation and Other Language Tasks. The Quarterly *Journal of Experimental Psychology Section A*, 48(3), 598-612. doi:10.1080/14640749508401407

Ikehara, C. S., & Crosby, M. E. (2005). Assessing Cognitive Load with Physiological Sensors. Paper presented at the Proceedings of the *38th annual hawaii international conference on system sciences*.

Iqbal, S. T., Adamczyk, P. D., Zheng, X. S., & Bailey, B. P. (2005). Towards an Index of Opportunity: Understanding Changes in Mental Workload During Task Execution. Paper presented at the Proceedings of the *SIGCHI conference on Human factors in computing systems*.

Ivonin, L., Chang, H.-M., Díaz, M., Català, A., Chen, W., & Rauterberg, M. (2014). Beyond Cognition and Affect: Sensing the Unconscious. *Behaviour & Information Technology*, 34(3), 220-238. doi:10.1080/0144929x.2014.912353

Kahneman, D., & Riis, J. (2005). Living, and Thinking About It : Two Perspectives on Life. *The Science of Well-Being* (pp. 285-304): Oxford University Press.

Lallemand, C., & Gronier, G. (2016). Focus Group. In Eyrolles (Ed.), Méthodes De Design UX.

Lamontagne, C., Sénécal, S., Fredette, M., Chen, S. L., Pourchon, R., Gaumont, Y., P.M, Léger, P.-M. (2019). User Test: How Many Users Are Needed to Find the Psychophysiological Pain Points in a Journey Map? Paper presented at the *International Conference on Human Interaction and Emerging Technologies*.

Léger, P.-M., Charland, P., Sénécal, S., & Cyr, S. (2018). Predicting Properties of Cognitive Pupillometry in Human–Computer Interaction: A Preliminary Investigation. *Information Systems and Neuroscience* (pp. 121-127): Springer.

Léger, P.-M., Courtemanche, F., Fredette, M., & Sénécal, S. (2019). A Cloud-Based Lab Management and Analytics Software for Triangulated Human-Centered Research, Cham.

Léger, P.-M., Davis, F. D., Cronan, T. P., & Perret, J. (2014). Neurophysiological Correlates of Cognitive Absorption in an Enactive Training Context. *Computers in Human Behavior*, 34, 273-283. doi:10.1016/j.chb.2014.02.011

Léger, P.-M., Riedl, R., & vom Brocke, J. (2014). Emotions and Erp Information Sourcing: The Moderating Role of Expertise. *Industrial Management & Data Systems*.

Lemon, K. N., & Verhoef, P. C. (2016). Understanding Customer Experience Throughout the Customer Journey. *Journal of Marketing*, 80(6), 69-96. doi:10.1509/jm.15.0420

Lewinski, P. (2015). Automated Facial Coding Software Outperforms People in 6, 1386. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/26441761. doi:10.3389/fpsyg.2015.01386

Lojiens, L., & Krips, O. (2018). Facereader Methodology Note.

Lourties, S., Léger, P.-M., Sénécal, S., Fredette, M., & Chen, S. L. (2018). Testing the Convergent Validity of Continuous Self-Perceived Measurement Systems : An Exploratory Study. Paper presented at the *HCI International*, Las Vegas, NV, USA.

Maia, C. L. B., & Furtado, E. S. (2016). A Study About Psychophysiological Measures in User Experience Monitoring and Evaluation.

Marquez, J. J., Downey, A., & Clement, R. (2015). Walking a Mile in the User's Shoes: Customer Journey Mapping as a Method to Understanding the User Experience. *Internet Reference Services Quarterly*, 20(3-4), 135-150.

Neslin, S. A., Grewal, D., Leghorn, R., Shankar, V., Teerling, M. L., Thomas, J. S., & Verhoef, P. C. (2006). Challenges and Opportunities in Multichannel Customer Management. *Journal of Service Research*, 9(2), 95-112. doi:10.1177/1094670506293559

Norton, D. W., & Pine, B. J. (2013). Using the Customer Journey to Road Test and Refine the Business Model. *Strategy & Leadership*, 41(2), 12-17. doi:10.1108/10878571311318196

Patrício, L., Fisk, R. P., & Falcão e Cunha, J. (2008). Designing Multi-Interface Service Experiences: The Service Experience Blueprint. *Journal of Service Research*, 10(4), 318-334.

Picard, R. W. (1995). Affective Computing. Perceptual Computing Section Technical Report No. 321, Cambridge, Massachusetts, M.I.T. Media Lab

Pine, B. J., & Gilmore, J. H. (2011). The Experience Economy: Harvard Business Press.

Plassmann, H., Venkatraman, V., Huettel, S., & Yoon, C. (2015). Consumer Neuroscience : Applications, Challenges, and Possible Solutions. *Journal of marketing research*, 427-435.

Posner, J., Russell, J. A., & Peterson, B. S. (2005). The Circumplex Model of Affect : An Integrative Approach to Affective Neuroscience, Cognitive Development, and Psychopathology. *Development and Psychopathology*, 17(3).

Rawson, A., Duncan, E., & Jones, C. (2013). The Truth About Customer Experience.

Reichheld, F. F., & Markey, R. (2011). The Ultimate Question 2.0: How Net Promoter Companies Thrive in a Customerdriven World. Retrieved from Boson, MA:

Richardson, A. (2010). Using Customer Journey Maps to Improve Customer Experience. *Harvard business review*, 15(1), 2-5.

Riedl, R., & Léger, P.-M. (2016). Fundamentals of Neurois. Studies in Neuroscience, *Psychology and Behavioral Economics*. Springer, Berlin, Heidelberg.

Rosenbaum, M. S., Otalora, M. L., & Ramírez, G. C. (2017). How to Create a Realistic Customer Journey Map. *Business Horizons*, 60(1), 143-150. doi:10.1016/j.bushor.2016.09.010

Russell, J. A. (1980). A Circumplex Model of Affect. *Journal of Personality and Social Psychology*, 39(6), 1161-1178.

Russell, J. A., Weiss, A., & Mendelsohn, G. A. (1989). Affect Grid: A single-item scale of pleasure and arousal. *Journal of Personality and Social Psychology*, *57*(3), 493-502. http://dx.doi.org/10.1037/0022-3514.57.3.493

Rust, R. T., & Lemon, K. N. (2001). E-Service and the Consumer. *International journal of electronic commerce*, 5(3), 85-101.

Schmitt, B. (1999). Experiential Marketing. *Journal of marketing management*, 15(1-3), 53-67.

Segelström, F. (2013). Stakeholder Engagement for Service Design: How Service Designers Identify and Communicate Insights. Linköping University Electronic Press.

Stern, R. M., Stern, R. M., Ray, W. J., & Quigley, K. S. (2001). Psychophysiological Recording: Oxford University Press, USA.

Stickdorn, M., Schneider, J., Andrews, K., & Lawrence, A. (2011). This Is Service Design Thinking: Basics, Tools, Cases (Vol. 1): Wiley Hoboken, NJ.

Temkin, B. D. (2010). Mapping the Customer Journey : Best Practices for Using an Important Customer Experience Tool. *Forrester Research, Inc*.

Tseng, M. M., Qinhai, M., & Su, C. J. (1999). Mapping Customers' Service Experience for Operations Improvement. *Business Process Management Journal*.

Uyl, M. J. d., & Kuilenburg, H. v. (2005). The Facereader : Online Facial Expression Recognition. *Proceedings of Measuring Behavior*.

Van Gerven, P. W., Paas, F., Van Merriënboer, J. J., & Schmidt, H. G. (2004). Memory Load and the Cognitive Pupillary Response in Aging. *Psychophysiology*, 41(2), 167-174.

Vasseur, A., Léger, P.-M., & Sénécal, S. (2019). Eye-Tracking for Information Systems Research : A Literature Review. *SIGHCI*, 8.

Venkatraman, V., Clithero, J. A., Fitzsimons, G. J., & Huettel, S. A. (2012). New Scanner Data for Brand Marketers: How Neuroscience Can Help Better Understand Differences in Brand Preferences. *Journal of consumer psychology*, 22(1), 143-153.

Verhoef, P. C., Lemon, K. N., Parasuraman, A., Roggeveen, A., Tsiros, M., & Schlesinger, L. A. (2009). Customer Experience Creation: Determinants, Dynamics and Management Strategies. *Journal of retailing*, 85(1), 31-41.

Verma, R., Teixeira, J., Patrício, L., Nunes, N. J., Nóbrega, L., Fisk, R. P., & Constantine, L. (2012). Customer Experience Modeling: From Customer Experience to Service Design. *Journal of Service Management*.

Vom Brocke, J., Hevner, A., Léger, P. M., Walla, P., & Riedl, R. (2020). Advancing a Neurois Research Agenda with Four Areas of Societal Contributions. *European Journal of Information Systems*, 1-16.

Ward, R. D., & Marsden, P. H. (2003). Physiological Responses to Different Web Page Designs. *International Journal of Human-Computer Studies*, 59(1-2), 199-212. doi:10.1016/s1071-5819(03)00019-3

Wilson, G. F. (2002). An Analysis of Mental Workload in Pilots During Flight Using Multiple Psychophysiological Measures. *The International Journal of Aviation Psychology*, 12(1), 3-18.

Zijlstra, F. R. H., Roe, R. A., Leonora, A. B., & Krediet, I. (1999). Temporal Factors in Mental Work : Effects of Interrupted Activities. *Journal of Occupational and Organizational Psychology*, 72, 163-185.

Conclusion

Ce mémoire aborde la visualisation de l'expérience utilisateur ainsi que différentes approches et méthodes permettant d'approfondir la compréhension de l'expérience vécue. Un premier objectif vise à tester, via une approche méthodologique, la contribution de l'utilisateur lors de l'évaluation de son expérience psychophysiologique. Plus précisément, le premier essai permet de comparer la précision des perceptions de l'utilisateur face à des événements intenses de son expérience avec les données psychophysiologiques provenant de l'expérience vécue. Ces interprétations sont par la suite comparées avec la performance d'un groupe d'experts en UX afin d'évaluer les performances de chaque groupe. Un deuxième objectif vise à présenter une nouvelle méthode de visualisation permettant de contextualiser l'expérience émotionnelle et cognitive des utilisateurs au fil de leur navigation. Cette méthode, présentée dans le cadre du deuxième essai, facilite le processus d'analyse d'une expérience puisqu'elle montre où se trouvait l'utilisateur sur l'interface lorsqu'il a vécu un état cognitif et émotionnel donné. Bref, ce mémoire a permis une meilleure compréhension de la contribution de l'utilisateur dans l'évaluation de son expérience en plus de mettre en lumière l'utilité des données psychophysiologiques dans le suivi de l'état émotionnel et cognitif de l'utilisateur.

Une expérience en laboratoire a été réalisée à l'hiver 2019 auprès de 38 participants. Chaque participant devait accomplir trois séries de tâches sélectionnées à partir d'un ensemble de tâches ayant un même niveau de difficulté. Les tâches consistaient à accomplir une série d'actions sur le site internet d'une institution financière pouvant être résumée en trois grandes catégories : recherche d'informations sur le site internet, utilisation d'outils financiers et contact avec l'institution financière pour plus d'informations. Les mesures utilisées lors de l'expérience étaient un oculomètre afin de suivre le regard de l'utilisateur en ligne, le logiciel FaceReader pour identifier les émotions faciales et BIOPAC pour mesurer l'intensité émotionnelle. Les mesures psychophysiologiques permettent d'identifier par la suite les pics émotionnels positifs (activation élevée et valence positive) et négatifs (activation élevée et valence négative). L'expérience était suivie d'une entrevue où l'utilisateur devait interpréter des séquences vidéo provenant des pics émotionnels de son expérience. Les interprétations de l'utilisateur consistaient en une évaluation des concepts de valence et d'activation à partir du Affective Slider (Betella & Verschure, 2016) ainsi qu'une description qualitative de ses intentions. Cette même interprétation a été également complétée par des experts en UX, qui représentent des individus non engagés émotionnellement à l'expérience, mais ayant de l'expérience en expérience utilisateur. Ces données ont été collectées afin de mettre en lumière la contribution et l'utilité d'intégrer l'utilisateur dans l'interprétation de l'expérience vécue. Les données psychophysiologiques collectées ont également été utilisées dans le développement d'une nouvelle méthode permettant de visualiser l'évolution de l'expérience des utilisateurs au fil de leur navigation. Les données psychophysiologiques ont permis d'illustrer l'utilisation de cette méthode pour un ensemble de tâches définies. La collecte a donc permis la rédaction des deux essais.

Les prochaines sections de ce chapitre présentent un rappel des questions de recherche ainsi que de principaux résultats observés. Ensuite, les contributions de ce mémoire ainsi que les implications seront énoncées suivies des limites ainsi que des avenues possibles de la recherche.

Rappel des questions de recherche et principaux résultats

Ce mémoire avait pour but d'explorer différents concepts entourant l'expérience utilisateur afin de répondre aux questions de recherche. En effet, les résultats ont tout d'abord permis de répondre à la question suivante : Lorsque les utilisateurs interprètent eux-même rétrospectivement leur expérience en ligne, sont-ils plus précis que les experts en UX?

Des réponses à cette question ont été abordées dans le cadre du premier essai. En effet, les résultats des interprétations des utilisateurs ont été comparés aux résultats des experts en UX afin de mettre en lumière la contribution de l'utilisateur lors du processus d'interprétation. Afin de répondre plus précisément à la question Q1, deux hypothèses ont été formulées, basées sur la littérature existante. L'hypothèse 1 (H1) postulait que les experts en UX allaient mieux performer que les utilisateurs lors de l'identification de l'état émotionnel de l'utilisateur pour un événement donné.

L’hypothèse 2 (H2) stipule que les moments négatifs seront mieux évalués par l’utilisateur que les moments positifs.

Les résultats suggèrent que, pour les événements sans ambiguïté, les utilisateurs ne sont pas en mesure de correctement évaluer les émotions vécues rétrospectivement tandis que les experts en UX sont en mesure de les évaluer correctement. Dans le cadre d’un événement sans ambiguïté, les experts en UX arrivent à la même interprétation de l’intention de l’utilisateur. En effet, lorsqu’une séquence vidéo est montrée à l’utilisateur afin de favoriser sa ré-immersion lors d’un événement précis de leur expérience, les utilisateurs ne sont pas en mesure d’identifier correctement l’émotion vécue. Cependant, les experts en UX ont une plus grande capacité à identifier l’émotion vécue par l’utilisateur pour une même séquence vidéo. Par ailleurs, en comparant la performance de chacun des groupes (utilisateurs vs experts en UX), les experts semblent davantage performants à identifier l’émotion vécue, et cette tendance est encore plus marquée pour les moments négatifs. Ces résultats démontrent donc que lors de la contextualisation des données psychophysiologiques, des observateurs externes semblent davantage en mesure de comprendre l’expérience vécue que les utilisateurs eux-mêmes.

Dans un deuxième temps, ce mémoire visait également à répondre à la question suivante : Comment présenter visuellement l’évolution de l’expérience vécue par les utilisateurs pour faciliter la compréhension de leur navigation ?

Les résultats démontrent qu’en superposant les données émotionnelles et cognitives à une représentation de la navigation des utilisateurs en ligne, il est possible d’avoir une vue d’ensemble résumant l’évolution de l’expérience pour l’ensemble des utilisateurs. Dans le cadre du deuxième essai, nous proposons une approche de visualisation qui présente, en conjonction avec le parcours réel d'un utilisateur sur une interface web, ses états cognitifs et émotionnels implicites au cours de ce parcours. L'approche, présentée au chapitre 2, triangule le comportement de l'utilisateur avec ses états cognitifs et émotionnels pour produire un aperçu complet de son expérience. La combinaison de ces méthodes permet d'obtenir des informations clés sur l'expérience de l'utilisateur et facilite la compréhension de l'évolution de l'expérience puisqu'elle montre où se trouvait l'utilisateur sur l'interface lorsqu'il a vécu un état cognitif et émotionnel donné. Cette méthode représente donc un outil intéressant pour les concepteurs puisqu'il permet d'identifier rapidement les sections nécessitant une amélioration d'interface.

Contributions

D'un point de vue théorique, les résultats contribuent à la littérature existante dans plusieurs domaines. En effet, les deux essais contribuent principalement à la littérature en expérience utilisateur ainsi qu'au domaine du commerce électronique.

Dans un premier temps, les résultats du premier essai contribuent, d'un point de vue théorique, à la littérature en expérience utilisateur en explorant la contribution réelle de l'utilisateur et de l'observateur externe lors de l'analyse de l'expérience utilisateur. De cette façon, les résultats ajoutent de la précision quant aux habiletés de chaque groupe. Plusieurs études explorent la congruence entre les données psychophysiologiques et ce que l'utilisateur se remémore de son expérience (Bentley et al., 2005; Bruun et al., 2016; Lourties et al.,2018) ainsi que l'efficacité d'une méthode d'évaluation comme outil complémentaire aux mesures psychophysiologiques (Lim, 2002). Cependant, peu d'études abordent la qualité d'interprétation de l'utilisateur à propos de leur expérience en comparaison à la qualité d'interprétation d'un groupe d'experts en UX.

Dans un deuxième temps, le second essai contribue également à la littérature en UX en présentant une méthode complète permettant de visualiser les émotions et les comportements d'un utilisateur lors de sa navigation sur un site web. De plus, cette méthode apporte davantage de précision quant à l'analyse et l'interprétation des résultats, permettant ainsi d'identifier les zones problématiques de l'interface. La méthode sert d'approche complémentaire à d'autres méthodes disponibles telles que les questionnaires et les entrevues qui permettent de recueillir des données autodéclarées et ainsi de permettre la convergence de ces trois perspectives (émotions, cognition et comportements), ce que peu de méthodes permettent de faire (Coursaris & Kim, 2011).

Implications

Le premier essai comprend également des implications managériales, car les résultats suggèrent que les experts en UX, caractérisés comme des individus impartiaux étant familier avec le UX, tels des praticiens en UX, sont davantage en mesure d'évaluer l'expérience émotionnelle d'un utilisateur que l'utilisateur lui-même. Ce résultat peut donc potentiellement influencer le type d'approche utilisé lors de l'analyse de l'expérience utilisateur. Toutefois, les experts nécessitent plus de temps pour analyser les résultats.

D'un point de vue managérial, la simplicité de la méthode présentée dans le deuxième essai permet de représenter visuellement l'évolution des états cognitifs et émotionnels d'un utilisateur tout au long de son parcours et de son expérience en ligne. Cette méthode devrait être particulièrement utile aux chercheurs et aux praticiens UX ainsi qu'en commerce électronique. Enfin, cette nouvelle méthode est utile pour comparer les expériences des utilisateurs sur différentes interfaces, qui peuvent être utilisées pour comparer l'expérience utilisateur d'une tâche spécifique sur des interfaces concurrentes.

Limites et recherches futures

La recherche suivante présente quelques limites à la généralisation de l'étude. Tout d'abord, le nombre total d'experts en UX peut être considéré comme une quantité très faible. En effet, la littérature abordant le nombre d'experts nécessaires lors d'une évaluation varie. Une étude par Lynn (1986) suggère un minimum de trois tandis que d'autres recommandent entre 2 et 20 experts (Gable & Wolf, 1993; Waltz, Strickland, & Lenz, 1991). Cependant, utiliser un plus grand nombre d'experts permet de générer plus d'informations à propos de la mesure souhaitée (Rubio et al., 2003). Cette étude pourrait être répliquée avec un plus grand nombre d'experts en UX afin d'augmenter la validité des résultats. Cependant, dû aux conditions d'analyse des experts en UX, davantage de ressources sont nécessaires afin d'atteindre un échantillon plus large.

Ensuite, l'expérience nécessite que l'utilisateur cherche de l'information concernant divers sujets et donc, l'expérience en ligne était d'une longue durée. Par conséquent, il est possible que certains pics émotionnels soient liés à un élément de fatigue plutôt que par une problématique spécifique au site internet. Des études de plus courte durée permettraient potentiellement de mitiger ce risque.

Finalement, les résultats présentés dans ce mémoire sont uniquement liés au domaine bancaire et donc lié aux tâches utilitaires. En effet, les tâches de nature utilitaire génèrent une certaine variation d'activation et de valence émotionnelle, mais ne présentent pas d'aussi grande variation en comparaison au contexte hédonique. Conséquemment, appliquer les mêmes études auprès d'un ensemble de tâches hédoniques permettrait de généraliser les résultats.

Bibliographie

Adobe, EConsultancy (2019). 2019 Digital trends.

Alvarez, J., Brieugne, D., Léger, P.-M., Sénécal, S., & Frédette, M. (2019). Towards Agility and Speed in Enriched UX Evaluation Projects. *Human-Computer Interaction*: IntechOpen.

Ambert, A.-M., Adler, P. A., Adler, P., & Detzner, D. F. (1995). Understanding and Evaluating Qualitative Research. *Journal of Marriage and the Family*, 879-893.

Anzieu, D., & Martin, J.-Y. (1968). La Dynamique Des Groupes Restreints: Presses Universitaires de France.

Ariely, D. (1998). Combining Experiences over Time: The Effects of Duration, Intensity Changes and on-Line Measurements on Retrospective Pain Evaluations. *Journal of Behavioral Decision Making*, 11, 11-45.

Bagdare, S., & Jain, R. (2013). Measuring Retail Customer Experience. International *Journal of Retail & Distribution Management*, 41(10), 790-804. doi:10.1108/ijrdm-08-2012-0084

Bailey, B. P., Adamczyk, P. D., Chang, T. Y., & Chilson, N. A. (2006). A Framework for Specifying and Monitoring User Tasks. *Computers in Human Behavior*, 22(4), 709-732. doi:10.1016/j.chb.2005.12.011

Bailey, B. P., & Konstan, J. A. (2006). On the Need for Attention-Aware Systems: Measuring Effects of Interruption on Task Performance, Error Rate, and Affective State. *Computers in Human Behavior*, 22(4), 685-708. doi:10.1 016/j.chb.2005.12.009

Barbier, C., Guyonvarc'h, L., Guillaume, A., & Tattegrain, H. (2018). Is the Self-Confrontation Method Applicable to Naturalistic Driving Studies? *Safety Science*. doi:10.1016/j.ssci.2018.11.005

Bargas-Avila, J. A., & Hornbaek, K. (2011). Old Wine in New Bottles or Novel Challenges? A Critical Analysis of Empirical Studies of User Experience. Paper presented at the *CHI*, Vancouver BC, Canada.

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad Is Stronger Than Good. *Review of General Psychology*, 5(4), 323-370. doi:10.1037/1089-2680.5.4.323

Beatty, J., & Lucero-Wagoner, B. (2000). The Pupillary System. Handbook of psychophysiology, 2(142-162).

Beauregard, R., & Corriveau, P. (2007). User Experience Quality: A Conceptual Framework for Goal Setting and Measurement. Paper presented at the *International Conference on Digital Human Modeling*.

Bentley, T., Johnson, L., & Baggo, K. v. (2005). Evaluation Using Cued-Recall Debrief to Elicit Information About a User's Affective Experiences. Paper presented at the *OZCHI*, Canberra, Australia.

Bernardin, H. J., & Buckley, M. R. (1981). Strategies in Rater Training. *Academy of Management Review*, 6(2), 205-212.

Betella, A., & Verschure, P. F. (2016). The Affective Slider: A Digital Self-Assessment Scale for the Measurement of Human Emotions. *PLoS ONE*, 11(2), e0148037. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/26849361. doi:10.1371/journal.pone.0148037

Bitner, M. J., Ostrom, A. L., & Morgan, F. N. (2008). Service Blueprinting: A Practical Technique for Service Innovation. *California Management Review*, 50(3), 66-94.

Bloom, L. C., & Watkins, M. J. (1999). Two-Component Theory of the Suffix Effect : Contrary Findings. *Journal of experimental psychology*, 25(6), 1452-2474.

Borkenau, P., & Ostendorf, F. (1987). Retrospective Estimates of Act Frequencies: How Accurately Do They Reflect Reality? *Journal of Personality and Social Psychology*, 52(3), 626-638. doi:10.1037/0022-3514.52.3.626

Boucsein, W. (2012). Electrodermal Activity: Springer.

Bradley, M. M., Greenwald, M. K., Petry, M. C., & Lang, P. J. (1992). Remembering Pictures: Pleasure and Arousal in Memory. *Journal of experimental psychology*, 18(2), 379-390.

Bradley, M. M., & Lang, P. J. (1994). Measuring Emotion: The Self-Assessment Manikin and the Semantic Differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25(1), 49-59. doi:10.1016/0005-7916(94)90063-9

Braithwaite, J. J., Watson, D. G., Jones, R., & Rowe, M. (2013). A guide for analysing electrodermal activity (EDA) & skin conductance responses (SCRs) for psychological experiments. *Psychophysiology*, 49(1), 1017-1034.

Browne, J. (2012). Customer Journey Mapping : What Is It For? Retrieved from *Forrester Research Inc*.

Bruun, A., & Ahm, S. (2015). Mind the Gap! Comparing Retrospective and Concurrent Ratings of Emotion in User Experience Evaluation. In Human-Computer Interaction – Interact 2015 (pp. 237-254).

Bruun, A., Law, E. L.-C., Heintz, M., & Alkly, L. H. A. (2016). Understanding the Relationship between Frustration and the Severity of Usability Problems. Paper presented at the Proceedings of the 2016 *CHI Conference on Human Factors in Computing Systems - CHI '16*.

Burton-Jones, A., & Gallivan, M. J. (2007). Toward a Deeper Understanding of System Usage in Organizations: A Multilevel Perspective. *MIS Quarterly*, 31(4), 657-679.

Cacioppo, J. T., & Tassinary, L. G. (1990). Principles of Psychophysiology: Physical, Social, and Inferential Elements: Cambridge University Press.

Charland, P., Léger, P.-M., Sénécal, S., Courtemanche, F., Mercier, J., Skelling, Y., & Labonté- Lemoyne, E. (2015). Assessing the Multiple Dimensions of Engagement to Characterize Learning: A Neurophysiological Perspective. *JoVE (Journal of Visualized Experiments)* (101), e52627.

Charles, R. L., & Nixon, J. (2019). Measuring Mental Workload Using Physiological Measures: A Systematic Review. *Applied Ergonomics*, 74, 221-232. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/30487103. doi:10.1016/j.apergo.2018.08.028

Chittenden, A. H. R. (2002). A Pastoral Care Teacher's Theory of Action, Interactive Thinking and Effective Teaching Practice. *Pastoral Care in Education*, 20(1), 3-10. doi:10.1111/1468-0122.00214

Cockburn, A., Quinn, P., & Gutwin, C. (2017). The Effects of Interaction Sequencing on User Experience and Preference. *International Journal of Human-Computer Studies*, 108, 89-104. doi:10.1016/j.ijhcs.2017.07.005

Colombetti, G. (2005). Appraising Valence. *Journal of Consciousness Studies*, 12(8-10), 103-126.

Coursaris, C. K., & Kim, D. J. (2011). A Meta-Analytical Review of Empirical Mobile Usability Studies. *Journal of Usability Studies*, 6(3), 117-171.

Courtemanche, F., Fredette, M., Senecal, S., Leger, P.-m., Dufresne, A., Georges, V., & Labonte-lemoyne, E. (2019). Method of and System for Processing Signals Sensed from a User. In: Google Patents.

De Gelder, B., & Vroomen, J. (1997). Modality Effects in Immediate Recall of Verbal and Non-Verbal Information. *European Journal of Cognitive Psychology*, 9(1), 97-110. doi:10.1080/713752541

De Grave, W. S., Boshuizen, H. P. A., & Schmidt, H. G. (1996). Problem Based Learning: Cognitive and Metacognitive Processes During Problem Analysis. *Instructional Science*, 24(5), 321-341. Retrieved from https://doi.org/10.1007/BF00118111. doi:10.1007/bf00118111

De Keyser, A., Lemon, K. N., Klaus, P., & Keiningham, T. L. (2015). A Framework for Understanding and Managing the Customer Experience. *Marketing Science Institute working paper series*, 15(121), 1-48.

Debus, M. (2007). Handbook for Excellence in Focus Group Research: Academy for Educational Development.

Desrochers, C., Léger, P.-M., Fredette, M., Mirhoseini, S., & Sénécal, S. (2019). The Arithmetic Complexity of Online Grocery Shopping: The Moderating Role of Product Pictures. *Industrial Management & Data Systems*.

Dimoka, A. (2010). What Does the Brain Tell Us About Trust and Distrust? Evidence from a Functional Neuroimaging Study. *MIS Quarterly*, 34(2), 373-396.

Dimoka, A., Pavlou, P. A., & Davis, F. D. (2011). Neurois: The Potential of Cognitive Neuroscience for Information Systems Research. *Information Systems Research*, 22(4), 687-702. doi:10.1287/isre.1100.0284

Dirican, A. C., & Göktürk, M. (2011). Psychophysiological Measures of Human Cognitive States Applied in Human Computer Interaction. *Procedia Computer Science*, 3, 1361-1367. doi:10.1016/j.procs.2011.01.016

Dmello, S. K. (2016). On the Influence of an Iterative Affect Annotation Approach on Inter-Observer and Self-Observer Reliability. *IEEE Transactions on Affective Computing*, 7(2), 136-149. doi:10.1109/taffc.2015.2457413

Duric, Z., Gray, W. D., Heishman, R., Li, F., Rosenfeld, A., Schoelles, M. J., Wechsler, H. (2002). Integrating Perceptual and Cognitive Modeling for Adaptive and Intelligent Human-Computer Interaction. Proceedings of the *IEEE*, 90(7).

Ekman, P. (1965). Differential Communication of Affect by Head and Body Cues. *Journal of Personality and Social Psychology*, 2(5), 726.

Ekman, P. (1973). Cross-Cultural Studies of Facial Expression. Darwin and facial expression: A century of research in review, 169222(1).

Ekman, P., & Friesen, W. V. (2003). In M. Books (Ed.), Unmasking the Face: A Guide to Recognizing Emotions from Facial Clues.

Finkenauer, C., & Rimé, B. (1998). Socially Shared Emotional Experiences Vs. Emotional Experiences Kept Secret: Differential Characteristics and Consequences. *Journal of Social and Clinical Psychology*, 17(3), 295-318.

Forne, M. (2012). Physiology as a Tool for UX and Usability Testing. (Media Technology Master), Royal Institute of technology.

Fredrickson, B. L., & Kahneman, D. (1993). Duration Neglect in Retrospective Evaluations of Affective Episodes. *Journal of Personality and Social Psychology*, 65(1), 45-55.

Gable, R. K., & Wolf, M. B. (2012). Instrument Development in the Affective Domain: Measuring Attitudes and Values in Corporate and School Settings (Vol. 36): Springer Science & Business Media.

Giroux-Huppé, C., Sénécal, S., Fredette, M., Chen, S. L., Demolin, B., & Léger, P.-M. (2019). Impact D'une Expérience Psychophysiologique Négative Sur La Satisfaction Des Consommateurs Dans Un Contexte D'épicerie En Ligne. (Master), HEC Montréal, Montréal.

Glanzer, M., & Cunitz, A. R. (1966). Two Storage Mechanisms in Free Recall. *Journal of verbal learning and verbal behavior*, 5(4), 351-360.

Goldman, R., Pea, R., Barron, B., & Derry, S. (2007). Video Research in the Learning Sciences. Mahwah, NJ: Lawrence Erlbaum Associates.

Gomez. (2010). Why Web Performance Matters: Is Your Site Driving Customer Away? Retrieved from http://www.mcrinc.com/Documents/Newsletters/201110_why_web_performance_matters.pdf

Gopalani, A. (2011). The Service-Enabled Customer Experience: A Jump-Start to Competitive Advantage. *Journal of Business Strategy*, 32(3), 4-12. Retrieved from https://doi.org/10.1108/02756661111121947. doi:10.1108/02756661111121947

Gosling, S. D., John, O. P., Craik, K. H., & Robins, R. W. (1998). Do People Know How They Behave? Self-Reported Act Frequencies Compared with on-Line Coding by Observers. *Journal of Personality and Social Psychology*, 74(5), 1337-1349.

Guérin, J., Riff, J., & Testevuide, S. (2004). Étude De L'activité « Située » De Collégiens En Cours D'eps : Une Opportunité Pour Examiner Les Conditions De Validité Des Entretiens D'autoconfrontation. *Revue française de pédagogie*, 147(1), 15-26. doi:10.3406/rfp.2004.3116

Guinea, A. O. d., & Markus, L. (2009). Why Break the Habit of a Lifetime? Rethinking the Roles of Intention, Habit, and Emotion in Continuing Information Technology Use. *MIS Quarterly*, 33(3), 433-444.

Guinea, A. O. d., Titah, R., & Léger, P.-M. (2014). Explicit and Implicit Antecedents of Users' Behavioral Beliefs in Information Systems: A Neuropsychological Investigation. *Journal of Management Information Systems*, 30(4), 179-210. doi:10.2753/mis0742-1222300407

Guinea, A. O. d., & Webster, J. (2013). An Investigation of Information Systems Use Patterns: Technological Events as Triggers, the Effect of Time and Consequences for Performance. *MIS Quarterly*, 37(4), 1165-1188.

Haapalainen, E., Kim, S., Forlizzi, J. F., & Dey, A. K. (2010). Psycho-Physiological Measures for Assessing Cognitive Load. Paper presented at the Proceedings of the *12th ACM international conference on Ubiquitous computing*.

Halford, W. K., & Sanders, M. R. (1990). The Relationship of Cognition and Behavior During Marital Interaction. *Journal of Social and Clinical Psychology*, 9(4), 489-510.

Halvorsrud, R., Kvale, K., & Følstad, A. (2016). Improving Service Quality through Customer Journey Analysis. *Journal of Service Theory and Practice*, 26(6), 840-867. doi:10.1108/jstp-05-2015-0111

Hamann, S. (2001). Cognitive and Neural Machanisms of Emotional Memory. *TRENDS in Cognitive Sciences*, 5(9).

Harré, R., & Secord, P. F. (1972). The Explanation of Social Behaviour. Oxford, England.

Harrigan, J. A., Rosenthal, R., & R.Scherer, K. (2005). The New Handbook of Methods in Nonverbal Behavior Research.

Hartson, R., & Pyla, P. (2012). The UX Book: Process and Guidelines for Ensuring a Quality User Experience.

Hassenzahl, M., Burmester, M., & Koller, F. (2003). Attrakdiff: Ein Fragebogen Zur Messung Wahrgenommener Hedonischer Und Pragmatischer Qualität. *Mensch & Computer 2003* (pp. 187-196): Springer.

Hassenzahl, M. (2008). User Experience (UX): Towards an Experiential Perspective on Product Quality. Paper presented at the *IHM*.

Hassenzahl, M., & Tractinsky, N. (2006). User Experience - a Research Agenda. *Behaviour & Information Technology*, 25(2), 91-97. doi:10.1080/01449290500330331

Helkkula, A., Kelleher, C., & Pihlström, M. (2012). Characterizing Value as an Experience: Implications for Service Researchers and Managers. *Journal of Service Research*, 15(1), 59-75.

Hermans, H. J. M. F., Riet;de Groot, Roel;Nauta, Jan F. (1990). Self-Confrontation as a Method for Assessment and Intervention. *Journal of Counseling and Development*, 69(2), p.156.

Hetland, A., Vitterso, J., Oscar Bo Wie, S., Kjelstrup, E., Mittner, M., & Dahl, T. I. (2018). Skiing and Thinking About It: Moment-to-Moment and Retrospective Analysis of Emotions in an Extreme Sport. *Frontiers Psychology*, 9, 971. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/29973894. doi:10.3389/fpsyg.2018.00971

Hockenbury, D. H., & Hockenbury, S. E. (2016). Discovering Psychology: Worth Publishers.

Hollan, J. D., & Hutchins, E. L. (2009). Opportunities and Challenges for Augmented Environments: A Distributed Cognition Perspective. *Designing User Friendly Augmented Work Environments* (pp. 237-259).

Holmlid, S., & Evenson, S. (2008). Bringing Service Design to Service Sciences, Management and Engineering. In Service Science, *Management and Engineering Education for the 21st Century* (pp. 341-345): Springer.

Hyönä, J., Tommola, J., & Alaja, A.-M. (2007). Pupil Dilation as a Measure of Processing Load in Simultaneous Interpretation and Other Language Tasks. The *Quarterly Journal of Experimental Psychology Section A*, 48(3), 598-612. doi:10.1080/14640749508401407

IDG (2018). State of Digital Business Transformation.

Ikehara, C. S., & Crosby, M. E. (2005). Assessing Cognitive Load with Physiological Sensors. Paper presented at the Proceedings of the *38th annual hawaii international conference on system sciences*.

Iqbal, S. T., Adamczyk, P. D., Zheng, X. S., & Bailey, B. P. (2005). Towards an Index of Opportunity: Understanding Changes in Mental Workload During Task Execution.

Paper presented at the Proceedings of the *SIGCHI* conference on Human factors in computing systems.

Ivonin, L., Chang, H.-M., Díaz, M., Català, A., Chen, W., & Rauterberg, M. (2014). Beyond Cognition and Affect: Sensing the Unconscious. *Behaviour & Information Technology*, 34(3), 220-238. doi:10.1080/0144929x.2014.912353

Jarodzka, H., Scheiter, K., Gerjets, P., & van Gog, T. (2010). In the Eyes of the Beholder: How Experts and Novices Interpret Dynamic Stimuli. *Learning and Instruction*, 20(2), 146-154. doi:10.1016/j.learninstruc.2009.02.019

Javetski, Bill. Fanderl, Harald. Maechler, Nicolas (2016). Why the customer experience matters, *McKinsey & Company*. Récupéré le 19 août 2019 de https://www.mckinsey.com/business-functions/marketing-and-sales/our-insights/why-the-customer-experience-matters

John, O. P., & Robins, R. W. (1994). Accuracy and Bias in Self-Perception: Individual Differences in Self-Enhancement and the Role of Narcissim. *Journal of Personality and Social Psychology*, 66(1), 206-219.

Kahneman, D., Fredrickson, B. L., Schreiber, C. A., & Redelmeier, D. A. (1993). When more pain is preferred to less: Adding a better end. *Psychological science*, 4(6), 401-405.

Kahneman, D., & Riis, J. (2005). Living, and Thinking About It: Two Perspectives on Life. *The Science of Well-Being* (pp. 285-304): Oxford University Press.

Kaplan, B., & Maxwell, J. A. (2005). Qualitative Research Methods for Evaluating Computer Information Systems. *Evaluating the Organizational Impact of Healthcare Information Systems* (pp. 30-55): Springer.

Kecklund, G., & Akerstedt, T. (2004). Apprehension of the Subsequent Working Day Is Associated with a Low Amount of Slow Wave Sleep. *Biological Psychology*, 66(2), 169-176. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/15041138. doi:10.1016/j.biopsycho.2003.10.004

Kelly, D. (2009). Methods for Evaluating Interactive Information Retrieval Systems with Users. *Foundations and Trends® in Information Retrieval*, 3(1–2), 1-224.

Kim, H.-W., Chan, H. C., & Gupta, S. (2007). Value-Based Adoption of Mobile Internet: An Empirical Investigation. *Decision support systems*, 43(1), 111-126.

Kolar, D. W., Funder, D. C., & Colvin, C. R. (1996). Comparing the Accuracy of Personality Judgments by the Self and Knowledgeable Others. *Journal of Personality*, 64(2), 311-337. doi:10.1111/j.1467-6494.1996.tb00513.x

Krosnick, J. A. (1999). Survey Research. Annual Review of Psychology, 50, 537-567.

Laeng, B., Sirois, S., & Gredeback, G. (2012). Pupillometry: A Window to the Preconscious? *Perspectives on Psychological Science*, 7(1), 18-27. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/26168419. doi:10.1177/1745691611427305

Lahlou, S. (2010). Digitization and Transmission of Human Experience. *Social Science Information*, 49(3), 291-327. doi:10.1177/0539018410372020

Lahlou, S. (2011). How Can We Capture the Subject's Perspective? An Evidence-Based Approach For the Social Scientist. *Social Science Information*, 50(3-4).

Lallemand, C., & Gronier, G. (2016). Focus Group. In Eyrolles (Ed.), Méthodes De Design UX.

Lamontagne, C., Sénécal, S., Fredette, M., Chen, S. L., Pourchon, R., Gaumont, Y., P-M, Léger, P.-M. (2019). User Test: How Many Users Are Needed to Find the Psychophysiological Pain Points in a Journey Map? Paper presented at the *International Conference on Human Interaction and Emerging Technologies*.

Léger, P.-M., Charland, P., Sénécal, S., & Cyr, S. (2018). Predicting Properties of Cognitive Pupillometry in Human–Computer Interaction: A Preliminary Investigation. *Information Systems and Neuroscience* (pp. 121-127): Springer.

Léger, P.-M., Courtemanche, F., Fredette, M., & Sénécal, S. (2019). A Cloud-Based Lab Management and Analytics Software for Triangulated Human-Centered Research, Cham.

Léger, P.-M., Davis, F. D., Cronan, T. P., & Perret, J. (2014). Neurophysiological Correlates of Cognitive Absorption in an Enactive Training Context. *Computers in Human Behavior*, 34, 273-283. doi:10.1016/j.chb.2014.02.011

Leger, P.-M., Riedl, R., & vom Brocke, J. (2014). Emotions and Erp Information Sourcing: The Moderating Role of Expertise. *Industrial Management & Data Systems*.

Lemon, K. N., & Verhoef, P. C. (2016). Understanding Customer Experience Throughout the Customer Journey. *Journal of Marketing*, 80(6), 69-96. doi:10.1509/jm.15.0420

Lewinski, P. (2015). Automated Facial Coding Software Outperforms People in Recognizing Neutral Faces as Neutral from Standardized Datasets. *Frontiers Psychology*, 6, 1386. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/26441761. doi:10.3389/fpsyg.2015.01386

Lewis, S., Pea, R., & Rosen, J. (2010). Beyond Participation to Co-Creation of Meaning: Mobile Social Media in Generative Learning Communities. *Social Science Information*, 49(3), 351-369. doi:10.1177/0539018410370726

Lim, S. (2002). The Self-Confrontation Interview: Towards an Enhanced Understanding of Human Factors in Web-Based Interaction for Improved Website Usability. *Journal of Electronic Commerce Research*, 3(3).

Lindgren, E. (2002). The Effect of Stimulated Recall on 14-Year-Olds' L1 Swedish and Efl Writing and Revision. *Language Teaching Research*, 6(3), 267-268.

Lipsman, Andrew (2019). Global Ecommerce 2019. *eMarketer*. Récupéré le 19 août 2019 de https://www.emarketer.com/content/global-ecommerce-2019.

Lojiens, L., & Krips, O. (2018). Facereader Methodology Note.

Lourties, S., Léger, P.-M., Sénécal, S., Fredette, M., & Chen, S. L. (2018a). Testing the Convergent Validity of Continuous Self-Perceived Measurement Systems: An Exploratory Study. Paper presented at the *HCI International*, Las Vegas, NV, USA.

Lourties, S., Léger, P.-M., Sénécal, S., Fredette, M., & Chen, S. L. (2018b). Testing the Convergent Validity of Continuous Self-Perceived Measurement Systems: An Exploratory Study. Paper presented at the *HCI International*, Las Vegas, NV, USA.

Lyddon, W. J., Yowell, D. R., & Hermans, H. J. M. (2006). The Self-Confrontation Method: Theory, Research, and Practical Utility. *Counselling Psychology Quarterly*, 19(1), 27-43. doi:10.1080/09515070600589719

Lyle, J. (2003). Stimulated Recall: A Report on Its Use in Naturalistic Research. British *Educational Research Journal*, 29(6), 861-878. doi:10.1080/0141192032000137349

Lynn, M. R. (1986). Determination and Quantification of Content Validity. *Nursing research*.

MacKay, D. G., Shafto, M., Taylor, J. K., Marian, D. E., Abrams, L., & Dyer, J. R. (2004). Relations between Emotion, Memory, and Attention: Evidence from Taboo Stroop, Lexical Decision, and Immediate Memory Tasks. *Memory & Cognition*, 32(3), 474-488.

Maia, C. L. B., & Furtado, E. S. (2016). A Study About Psychophysiological Measures in User Experience Monitoring and Evaluation.

Maklan, S., & Klaus, P. (2011). Customer Experience: Are We Measuring the Right Things? *International Journal of Market Research*, 53(6), 771-772.

Mandryk, R. L., Inkpen, K. M., & Calvert, T. W. (2006). Using Psychophysiological Techniques to Measure User Experience with Entertainment Technologies. *Behaviour & Information Technology*, 25(2), 141-158. doi:10.1080/01449290500331156

Markey, R., & Reichheld, F. (2011). Introducing: The Net Promoter System.

Marquez, J. J., Downey, A., & Clement, R. (2015). Walking a Mile in the User's Shoes: Customer Journey Mapping as a Method to Understanding the User Experience. *Internet Reference Services Quarterly*, 20(3-4), 135-150.

Mehu, M., & Scherer, K. R. (2012). A Psycho-Ethological Approach to Social Signal Processing. *Cognition Process*, 13 Suppl 2, 397-414. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/22328016. doi:10.1007/s10339-012-0435-2

Miles, M. B., & Huberman, A. M. (1984). Qualitative Data Analysis: A Sourcebook of New Methods. In Qualitative Data Analysis: A Sourcebook of New Methods: Sage publications.

Miron-Shatz, T., Stone, A., & Kahneman, D. (2009). Memories of Yesterday's Emotions: Does the Valence of Experience Affect the Memory-Experience Gap? *Emotion*, 9(6), 885-891. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/20001131. doi:10.1037/a0017823

Murdock, B. B. (1962). The Serial Position Effect of Free Recall. *Journal of experimental psychology*, 64(5), 482-488.

Neslin, S. A., Grewal, D., Leghorn, R., Shankar, V., Teerling, M. L., Thomas, J. S., & Verhoef, P. C. (2006). Challenges and Opportunities in Multichannel Customer Management. *Journal of Service Research*, 9(2), 95-112. doi:10.1177/1094670506293559

Nisbett, R. E., & Wilson, T. D. (1977). Telling More Than We Can Know: Verbal Reports on Mental Processes. *Psychological Review*, 84(3), 231.

Norton, D. W., & Pine, B. J. (2013). Using the Customer Journey to Road Test and Refine the Business Model. *Strategy & Leadership*, 41(2), 12-17. doi:10.1108/10878571311318196

Oh, I. S., Wang, G., & Mount, M. K. (2011). Validity of Observer Ratings of the Five-Factor Model of Personality Traits: A Meta-Analysis. Journal of Applied Psychology, 96(4), 762-773. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/21142341. doi:10.1037/a0021832

Omodei, M. M., & McLennan, J. (1994). Studying Complex Decision Making in Natural Settings: Using a Head-Mounted Video Camera to Study Competitive Orienteering. *Perceptual and Motor Skills* (79), 1411-1425.

Patrício, L., Fisk, R. P., & Falcão e Cunha, J. (2008). Designing Multi-Interface Service Experiences: The Service Experience Blueprint. *Journal of Service Research*, 10(4), 318-334.

Picard, R. W. (1995). Affective Computing. Perceptual Computing Section Technical Report No. 321, Cambridge, Massachusetts, M.I.T. Media Lab

Pine, B. J., & Gilmore, J. H. (2011). The Experience Economy: Harvard Business Press.

Plassmann, H., Venkatraman, V., Huettel, S., & Yoon, C. (2015). Consumer Neuroscience: Applications, Challenges, and Possible Solutions. *Journal of marketing research*, 427-435.

Posner, J., Russell, J. A., & Peterson, B. S. (2005). The Circumplex Model of Affect: An Integrative Approach to Affective Neuroscience, Cognitive Development, and Psychopathology. *Development and Psychopathology*, 17(3).

Rawson, A., Duncan, E., & Jones, C. (2013). The Truth About Customer Experience.

Redelmeier, D. A., & Kahneman, D. (1996). Patients' Memories of Painful Medical Treatments: Real-Time and Retrospective Evaluations of Two Minimally Invasive Procedures. *Pain*, 66(1), 3-8.

Reichheld, F. F., & Markey, R. (2011). The Ultimate Question 2.0: How Net Promoter Companies Thrive in a Customerdriven World. Retrieved from Boson, MA:

Richardson, A. (2010). Using Customer Journey Maps to Improve Customer Experience. *Harvard business review*, 15(1), 2-5.

Riedl, R., & Léger, P.-M. (2016). Fundamentals of Neurois. Studies in Neuroscience, Psychology and Behavioral Economics. Springer, Berlin, Heidelberg.

Robins, R. W., & John, O. P. (1997). Effects of Visual Perspective and Narcissism on Self-Perception: Is Seeing Believing? *Psychological Science*, 8(1), 37-42.

Robins, R. W., & John, O. P. (1997). The Quest for Self-Insight: Theory and Research on Accuracy and Bias in Self-Perception. In N. Y. A. Press (Ed.), *Handbook of Personality Psychology* (pp. 649-679).

Rosenbaum, M. S., Otalora, M. L., & Ramírez, G. C. (2017). How to Create a Realistic Customer Journey Map. *Business Horizons*, 60(1), 143-150. doi:10.1016/j.bushor.2016.09.010

Rosenthal, R. (1966). Experimenter Effects in Behavioral Research. Appleton-Century-Crofts. New York.

Rubio, D. M., Berg-Weger, M., Tebb, S. S., Lee, E. S., & Rauch, S. (2003). Objectifying Content Validity: Conducting a Content Validity Study in Social Work Research. *Social Work Research*, 27(2).

Russell, J. A. (1980). A Circumplex Model of Affect. *Journal of Personality and Social Psychology*, 39(6), 1161-1178.

Russell, J. A., Weiss, A., & Mendelsohn, G. A. (1989). Affect Grid: A single-item scale of pleasure and arousal. *Journal of Personality and Social Psychology, 57*(3), 493-502. http://dx.doi.org/10.1037/0022-3514.57.3.493

Rust, R. T., & Lemon, K. N. (2001). E-Service and the Consumer. *International journal of electronic commerce*, 5(3), 85-101.

Scherer, K. R. (1984). Emotion as a Multicomponent Process: A Model and Some Cross-Cultural Data. Review of personality & social psychology.

Schooler, J. W. (2011). Introspecting in the Spirit of William James: Comment on Fox, Ericsson, and Best (2011). *Psychological Bulletin*, 137(2), 345-350. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/21355632. doi:10.1037/a0022390

Schmitt, B. (1999). Experiential Marketing. *Journal of marketing management*, 15(1-3), 53-67.

Schwarz, N. (2007). Retrospective and Concurrent Self-Reports: The Rationale for Real-Time Data Capture. The science of real-time data capture: *Self-reports in health research*, 11, 26.

Segelström, F. (2013). Stakeholder Engagement for Service Design: How Service Designers Identify and Communicate Insights. Linköping University Electronic Press.

Smith, A. (1997). Human Computer Factors: A Study of Users and Information Systems.

Stern, R. M., Stern, R. M., Ray, W. J., & Quigley, K. S. (2001). Psychophysiological Recording: Oxford University Press, USA.

Stickdorn, M., Schneider, J., Andrews, K., & Lawrence, A. (2011). This Is Service Design Thinking: Basics, Tools, Cases (Vol. 1): Wiley Hoboken, NJ.

Temkin, B. D. (2010). Mapping the Customer Journey: Best Practices for Using an Important Customer Experience Tool. *Forrester Research, Inc*.

Tesser, A., & Campbell, J. (1983). Self-Definition and Self-Evaluation Maintenance. *Psychological perspectives on the self*, 2, 1-31.

Thomas, D. L., & Diener, E. (1990). Memory Accuracy in the Recall of Emotions. *Journal of Personality and Social Psychology*, 59(2), 291-297.

Tseng, M. M., Qinhai, M., & Su, C. J. (1999). Mapping Customers' Service Experience for Operations Improvement. *Business Process Management Journal*.

Uyl, M. J. d., & Kuilenburg, H. v. (2005). The Facereader: Online Facial Expression Recognition. Proceedings of Measuring Behavior.

Van Gerven, P. W., Paas, F., Van Merriënboer, J. J., & Schmidt, H. G. (2004). Memory Load and the Cognitive Pupillary Response in Aging. *Psychophysiology*, 41(2), 167-174.

Van Gog, T., Paas, F., van Merrienboer, J. J., & Witte, P. (2005). Uncovering the Problem-Solving Process: Cued Retrospective Reporting Versus Concurrent and Retrospective Reporting. *Journal of Experimental Psychology Applied*, 11(4), 237-244. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/16393033. doi:10.1037/1076- 898X.11.4.237

Vasseur, A., Léger, P.-M., & Sénécal, S. (2019). Eye-Tracking for Information Systems Research: A Literature Review. *SIGHCI*, 8.

Venkatraman, V., Clithero, J. A., Fitzsimons, G. J., & Huettel, S. A. (2012). New Scanner Data for Brand Marketers: How Neuroscience Can Help Better Understand Differences in Brand Preferences. *Journal of consumer psychology*, 22(1), 143-153.

Verhoef, P. C., Lemon, K. N., Parasuraman, A., Roggeveen, A., Tsiros, M., & Schlesinger, L. A. (2009). Customer Experience Creation: Determinants, Dynamics and Management Strategies. *Journal of retailing*, 85(1), 31-41.

Verma, R., Teixeira, J., Patrício, L., Nunes, N. J., Nóbrega, L., Fisk, R. P., & Constantine, L. (2012). Customer Experience Modeling: From Customer Experience to Service Design. *Journal of Service Management*.

Vermeeren, A. P. O. S., Roto, V., & Väänänen, K. (2015). Design-Inclusive UX Research: Design as a Part of Doing User Experience Research. *Behaviour & Information Technology*, 35(1), 21-37. doi:10.1080/0144929x.2015.1081292

Vicente, K. J., Thornton, D. C., & Moray, N. (1987). Spectral Analysis of Sinus Arrhythmia: A Measure of Mental Effort. *Human Factors*, 29(2), 171-182.

Vom Brocke, J., Hevner, A., Léger, P. M., Walla, P., & Riedl, R. (2020). Advancing a Neurois Research Agenda with Four Areas of Societal Contributions. *European Journal of Information Systems*, 1-16.

Waltz, C. F., Strickland, O., & Lenz, E. R. (1991). Measurement in Nursing Research: FA Davis Company.

Ward, R. D., & Marsden, P. H. (2003). Physiological Responses to Different Web Page Designs. *International Journal of Human-Computer Studies*, 59(1-2), 199-212. doi:10.1016/s1071-5819(03)00019-3

Watkins, M. J. (1974). Concept and Measurement of Primary Memory. *Psychological Bulletin*, 81(10), 695-711.

Waugh, N. C., & Norman, D. A. (1965). Primary Memory. *Psychological Review*, 72.

White, P. A., & Younger, D. P. (1988). Differences in the Ascription of Transient Internal States to Self and Other. *Journal of Experimental Social Psychology*, 24(4), 292-309.

Wilson, G. F. (2002). An Analysis of Mental Workload in Pilots During Flight Using Multiple Psychophysiological Measures. *The International Journal of Aviation Psychology*, 12(1), 3-18.

Wilson, G. M., & Sasse, M. A. (2004). From Doing to Being: Getting Closer to the User Experience. *Interacting with Computers*, 16, 697-705. doi:10.1016/j.intcom.2004.06.001

Zanting, A., Verloop, N., & Vermunt, J. D. (2001). Student Teachers Eliciting Mentors' Practical Knowledge and Comparing It to Their Own Beliefs. *Teaching and Teacher Education*, 17, 725-740.

Zijlstra, F. R. H., Roe, R. A., Leonora, A. B., & Krediet, I. (1999). Temporal Factors in Mentel Work: Effects of Interrupted Activities. *Journal of Occupational and Organizational Psychology*, 72, 163-185.