



**HAL**  
open science

# Interaction de la pertinence intrinsèque et de la pertinence au but dans l'attention émotionnelle et la tendance à l'action : Evidences expérimentales et propositions théorique et méthodologique

Hippolyte Fournier

► **To cite this version:**

Hippolyte Fournier. Interaction de la pertinence intrinsèque et de la pertinence au but dans l'attention émotionnelle et la tendance à l'action : Evidences expérimentales et propositions théorique et méthodologique. Psychologie. Université Lumière - Lyon II, 2023. Français. NNT : 2023LYO20025 . tel-04227056

**HAL Id: tel-04227056**

**<https://theses.hal.science/tel-04227056>**

Submitted on 3 Oct 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



N° d'ordre NNT : 2023LYO20025

# THÈSE de DOCTORAT DE L'UNIVERSITÉ LUMIÈRE LYON 2

**École Doctorale : ED 476 Neurosciences et Cognition**

Discipline : Sciences cognitives

Soutenue publiquement le 12 juin 2023, par :

Hippolyte FOURNIER

---

## **Interaction de la pertinence intrinsèque et de la pertinence au but dans l'attention émotionnelle et la tendance à l'action.**

*Évidences expérimentales et propositions théorique et méthodologique.*

---

Devant le jury composé de :

David SANDER, Professeur des Universités, Université de Genève, Président

Gilles POURTOIS, Professeur des Universités, Université de Gent, Rapporteur

Pascal HOT, Professeur des Universités, Université de Savoie, Rapporteur

Hanna CHAINAY, Professeure des Universités, Université Lumière Lyon 2, Examinatrice

Eva POOL, Maître assistant, Université de Genève, Examinatrice

Olivier KOENIG, Professeur des Universités, Université Lumière Lyon 2, Directeur de thèse

# Contrat de diffusion

Ce document est diffusé sous le contrat *Creative Commons* « [Paternité – pas d'utilisation commerciale - pas de modification](#) » : vous êtes libre de le reproduire, de le distribuer et de le communiquer au public à condition d'en mentionner le nom de l'auteur et de ne pas le modifier, le transformer, l'adapter ni l'utiliser à des fins commerciales.

UNIVERSITÉ LUMIÈRE LYON 2

École doctorale Neurosciences et Cognition (NSCo – ED 476)

Laboratoire d'Étude des Mécanismes Cognitifs (EMC)

---

**Interaction de la pertinence intrinsèque et de la pertinence au  
but dans l'attention émotionnelle et la tendance à l'action :**

*Évidences expérimentales et propositions théorique et  
méthodologique*

---

Hippolyte FOURNIER

Thèse de doctorat en Sciences Cognitives

Réalisée sous la direction du Pr. Olivier KOENIG

Présentée et soutenue publiquement le 12 juin 2023

Composition du jury :

David SANDER, Professeur, Université de Genève, Président

Gilles POURTOIS, Professeur, Ghent University, Rapporteur

Pascal HOT, Professeur, Université Savoie Mont Blanc, Rapporteur

Eva POOL, Maître Assistante, Université de Genève, Examinatrice

Hanna CHAINAY, Professeur, Université Lyon 2, Examinatrice

Olivier KOENIG, Professeur, Université Lyon 2, Directeur de thèse

*Today's posterior is tomorrow's prior.*

Dennis Lindley

## Résumé

Selon les théories de l'appraisal, les interactions des individus avec leur environnement pourraient être définies sur la base d'un processus d'évaluation cognitive des stimuli selon plusieurs critères. Dans cette thèse, nous nous sommes intéressés à la manière dont le processus d'évaluation cognitive impacte les sous-systèmes cognitifs à travers l'attention par le biais de l'évaluation des stimuli selon le critère de pertinence. Notre intérêt s'est porté sur l'impact combiné de deux sous-critères du critère de pertinence sur l'attention, à savoir : la pertinence intrinsèque, définie par la nature plutôt déplaisante ou plaisante des stimuli de manière générale ; et la pertinence aux buts, définie par le niveau d'obstruction ou de facilitation des stimuli pour la réalisation des buts et besoins momentanément importants. Tandis que le caractère déplaisant et obstructif pour les buts et besoins des stimuli générerait probablement une tendance à les éviter, leur caractère plaisant et facilitateur générerait probablement une tendance à l'approche. Quatre expériences ont été réalisées afin d'apporter des données à cette thématique. Les résultats de ces études ont montré que : 1) la pertinence intrinsèque et la pertinence aux buts et besoins ont des effets multiplicatifs sur l'attention ; 2) la capture attentionnelle vers les stimuli est augmentée lorsque leurs pertinences génèrent à la fois approche et évitement ; 3) la capture attentionnelle vers des stimuli intrinsèquement pertinents peut être annulée en rendant ces stimuli non pertinents pour la réalisation d'un but induit auprès des participants ; 4) une mesure plus fine du biais attentionnel vers la menace, exacerbé dans l'anxiété, peut être obtenue au moyen d'une tâche que nous avons développée, dite « d'eye-dot-probe », comparativement à une tâche de « dot-probe » classique. Enfin, un essai théorique a été réalisé afin d'intégrer les hypothèses des théories de l'appraisal à celles de la théorie de l'inférence active (i.e., définissant le cerveau comme une machine à prédire). Cet essai a permis de proposer un modèle computationnel de l'attention émotionnelle dérivé de ces deux théories.

Mots-clés : attention, émotion, pertinence intrinsèque, pertinence aux buts

## Abstract

According to appraisal theories, individuals' interactions with their environment could be defined on the basis of a cognitive appraisal process of stimuli according to several criteria. In this thesis, we were interested in how the cognitive appraisal process impacts cognitive subsystems through attention by evaluating stimuli according to the criterion of relevance. Our interest was in the combined impact of two sub-criteria of the relevance criterion on attention, namely: intrinsic relevance, defined by the rather unpleasant or pleasant nature of the stimuli in general; and goals and needs relevance, defined by the level of obstruction or facilitation of the stimuli for the realization of momentarily important goals and needs. While the unpleasantness and obstructiveness of the stimuli for goals and needs would likely generate an avoidance tendency, their pleasantness and facilitation would likely generate an approach tendency. Four experiments were conducted providing the following results: 1) intrinsic relevance and goals and needs relevance have multiplicative effects on attention; 2) the attentional capture toward stimuli is increased when their relevance generates both approach and avoidance; 3) the attentional capture toward intrinsically relevant stimuli can be negated by making these stimuli irrelevant to the achievement of a goal induced in participants; 4) a finer measure of the attentional bias toward threat, that is exacerbated in anxiety, can be obtained using a task we developed, called the "eye-dot-probe" task, compared to a classic dot-probe task. Finally, a theoretical essay was carried out in order to integrate the assumptions of appraisal theories with those of the active inference theory (i.e., a theory defining the brain as a prediction machine). This essay allowed us to propose a computational model of emotional attention derived from these two theories.

Keywords: attention, emotion, intrinsic relevance, goal relevance

## Remerciements

*Difficile de savoir par où commencer au sujet des remerciements que j'aimerais adresser. Aussi, cela sera fait par ordre chronologique. En matière de tissu narratif, la chronologie a le mérite d'illustrer la manière dont les fils sont agencés.*

*Premièrement, j'aimerais remercier mes parents, sans qui, je n'aurais certainement pas suivi cette trajectoire de vie. Tout en me laissant faire mes propres choix – et erreurs, vous avez su me guider vers la bonne direction. J'imagine que cela n'a pas toujours été facile, surtout pour toi maman, qui a dû faire face à mes entêtements concernant des choix dont la portée m'échappait. Il ne m'appartient pas de vous dire que vous pouvez être fiers de vous, cela étant, j'ai le sentiment d'être là où je dois être, au moment où je le suis, en partie grâce à vous.*

*Je souhaite ensuite remercier mes amis d'enfance qui m'ont permis de maintenir une certaine forme d'équilibre durant la période du doctorat, ce qui me semble être une condition nécessaire au fait de bien vivre cette aventure – ainsi que toutes les autres. Aussi, merci à Clément, Pierre, Greg, Joris, Laura, Gouv, Bounab, Jonath, Victor, Lio et Moiroud. I would especially like to thank Malik for his constant positivity, despite the difficult pandemic context of the last few years, which he did not let affect him. As you often say, you are a nurse and that's it. Pour les mêmes raisons, merci à Sara, qui m'a accompagné depuis la licence, le master et enfin le doctorat, ainsi qu'à Shannon, dont le caillou a veillé sur moi pendant cette dernière année de thèse.*

*Je tiens également à remercier l'ensemble des membres du laboratoire EMC et particulièrement les doctorants qui m'ont accompagné au quotidien durant la thèse. Élise, j'ai hésité jusqu'au dernier moment à savoir si je devais t'intégrer à ce paragraphe ou au précédent. Dans la mesure où vous ne faites qu'une personne avec Shannon, je me suis dit que la mentionner avant me permettrait de te citer aux deux endroits. Plus sérieusement, merci pour*



*ta présence ces dernières années, tant du point de vue du travail que de la vie sociale. Tu m'as beaucoup apporté sur les deux plans et j'espère que l'avenir nous permettra de continuer à collaborer à la fois en dehors et dans le travail. Merci à Blublu et Maxi avec qui il a parfois été compliqué de s'arrêter de discuter, tant il est difficile de les arrêter. Merci également à Théo pour nos discussions scientifiques qui m'ont challengé et qui m'ont permis de mieux comprendre que le cadre théorique que j'avais choisi était le bon. Enfin, merci à Fede ainsi qu'à Brice qui prenait parfois le temps de nous rendre visite entre deux arbitrages.*

*Je souhaite évidemment remercier Olivier, qui m'a donné la chance de pouvoir travailler à ses côtés depuis le Master jusqu'au Doctorat. Je pense que chaque thèse est différente en partie pour l'encadrement qui l'accompagne et je n'aurais pu espérer en avoir un meilleur. La liberté guidée dont votre encadrement a fait preuve m'a permis de trouver ma voie scientifique et de m'armer pour la suite.*

*Merci aux membres du consortium THERADIA avec qui j'ai pu travailler ces dernières années sur un projet vraiment très intéressant. En particulier, je tiens à remercier Isabella pour sa présence et ses précieux conseils. Merci à Fabien qui m'a permis de finir de rédiger cette thèse dans de bonnes conditions et qui m'a offert la chance de continuer à travailler à ses côtés pour les mois qui viennent.*

*Eléonore, ton nom aurait pu figurer dans plusieurs des paragraphes précédents, aussi, je t'en dédie un à toi seule. Merci pour ton accompagnement durant ces dernières années et particulièrement ces derniers mois. Ta présence m'a beaucoup apporté et je n'aurais pas vécu les choses aussi sereinement sans ton soutien. Merci également pour ta relecture de la présente thèse. Quelque part, les fautes restantes dans le manuscrit sont aussi un peu les tiennes.*

*Je remercie également tous les participants aux études qui ont été réalisées, ainsi que les étudiants de Master qui ont pu participer au recueil de données.*

*Enfin, je tiens à remercier l'ensemble du jury, David Sander, Eva Pool, Gilles Pourtois, Hanna Chainay et Pascal Hot, d'avoir accepté d'évaluer ce travail de thèse et d'en débattre lors de la soutenance. Merci également à Yves Rossetti d'avoir participé aux différentes réunions du comité de suivi de thèse durant ces dernières années. Vos conseils et retours, en complément de ceux de David, m'ont été d'une aide précieuse.*

## Liste des abréviations

<b>Abréviation</b>	<b>Signification</b>
FACS	Facial Action Coding System
MPC	Modèle des Processus Composants
IBA	Indice de Biais Attentionnel
VBA	Variabilité du Biais Attentionnel
TL-BS	Trial Level-Bias Score
ECI	Entraînement Cognitif Informatisé
4P	Préventive, Personnalisée, Participative et Prédicative
PANAS	Positive and Negative Affect Schedule
SAM	Self Assessment Manikin
BMIS	Brief Mood Introspection Scale
STAI	State-Trait Anxiety Inventory
TIPI	Ten-Item Personality Inventory
GMS-28	Global Motivation Scale
MMSE	Mini-Mental State Examination
TMT	Trial Making Test
FAB	Frontal Assessment Battery
5MD	5 Mots de Dubois
TEPE	Tripartite d'Expression et de Perception de l'Émotion

# Table des matières

<i>Résumé</i> .....	3
<i>Abstract</i> .....	4
<i>Remerciements</i> .....	5
<i>Liste des abréviations</i> .....	8
<i>Table des matières</i> .....	9
<i>Avant-propos</i> .....	12
<b>Chapitre 1 - Introduction générale</b> .....	<b>20</b>
<b>1. Les théories modernes de l'émotion</b> .....	<b>20</b>
a. Le débat de la séquence .....	20
b. La théorie bifactorielle.....	21
c. La théorie des émotions de base .....	22
d. La théorie bi-dimensionnelle .....	23
e. Les théories de l'appraisal.....	23
<b>2. Focus sur les théories de l'appraisal</b> .....	<b>25</b>
a. Le Modèle des Processus Composants .....	25
b. La nature des épisodes : non-affectifs ou émotionnels ? .....	38
c. La différenciation des émotions .....	40
<b>3. L'attention émotionnelle</b> .....	<b>41</b>
a. La théorie du module de la peur.....	42
b. La théorie de l'arousal .....	43
c. Les théories de l'appraisal et pertinence .....	44
<b>4. Attention émotionnelle et trouble affectif : le cas de l'anxiété</b> .....	<b>50</b>
a. Capture attentionnelle vers la menace et troubles anxieux .....	50
b. Une littérature en crise.....	50

	10
c. Les pistes d'amélioration .....	52
<b>5. La théorie de l'inférence active .....</b>	<b>53</b>
a. L'hypothèse du codage prédictif.....	54
b. Le principe d'énergie libre.....	55
c. L'inférence active .....	57
<b>6. Conclusion de l'introduction générale .....</b>	<b>58</b>
 <i>Chapitre 2 - Article : "Combined effects of intrinsic and goal relevances on attention and action tendency during the emotional episode" .....</i>	 <i>60</i>
 <i>Chapitre 3 - Article : "From emotion to action tendency: Time course of attentional effects of agonistic and antagonistic overlay of intrinsic and goal relevances" .....</i>	 <i>100</i>
 <i>Chapitre 4 - Article : "Modulation of attentional capture by combining intrinsic pleasantness and goal relevance or unavoidable goal obstruction" .....</i>	 <i>141</i>
 <i>Chapitre 6 - Article : "The eye-dot-probe task: An innovative lead to overcome the crisis of the attentional bias toward threat" .....</i>	 <i>179</i>
 <i>Chapitre 7 - Article : "Integrated Theory of Appraisal under Causal Hierarchical Inference: an attempt to map appraisal theories to active inference theory" .....</i>	 <i>201</i>
 <b>Chapitre 8 - Projet THERADIA .....</b>	 <b>242</b>
<b>1. Introduction .....</b>	<b>242</b>
a. Altération des fonctions cognitives et état des traitements actuels.....	242
b. L'ECI en réponse à l'altération des fonctions cognitives : des perspectives intéressantes .....	243
c. Les problèmes inhérents à l'ECI.....	244
<b>2. Présentation du projet THERADIA.....</b>	<b>245</b>
a. L'organisation du projet et ses partenaires .....	247
b. Les problématiques liées au développement de l'assistant virtuel .....	249
c. Phase expérimentale répondant à ces problématiques .....	252

<b>3. Méthode de la phase expérimentale.....</b>	<b>253</b>
a. Participants.....	253
b. Matériel.....	254
c. Design.....	256
d. Procédure.....	258
<b>4. Usage du corpus de données récoltées durant la phase d'expérimentation : focus sur la partie détection d'émotion.....</b>	<b>262</b>
a. Segmentation des séquences audio-visuelles.....	262
b. Annotation des extraits audiovisuels selon l'émotion.....	263
c. État d'avancement actuel du projet et suite.....	277
<b><i>Chapitre 9 - Discussion générale et conclusion.....</i></b>	<b>279</b>
<b>1. Discussion générale.....</b>	<b>279</b>
a. Effets multiplicatifs des pertinences.....	280
b. Effets combinés des pertinences dans le cas d'antagoniste et leur temporalité.....	282
c. Annulation des effets de la pertinence intrinsèque par la pertinence aux buts.....	286
d. Considérations méthodologiques de la mesure du biais attentionnel dans le cadre de l'anxiété.....	289
e. ITACHI : vers une théorie intégrée de l'appraisal et de l'inférence active.....	291
<b>2. Conclusion.....</b>	<b>293</b>
<b><i>Bibliographie.....</i></b>	<b>296</b>

## Avant-propos

La présente thèse a été financée par le projet THERADIA, un projet pluri-équipes soutenu par Bpifrance, dans le cadre de projets structurants des pôles de compétitivité. Ce projet a pour objectif de développer un assistant virtuel sous la forme d'un avatar qui pourra accompagner des patients dans la réalisation d'entraînements cognitifs informatisés visant à remédier/maintenir leur fonctionnement cognitif. Notre apport dans ce projet s'est porté sur la partie visant à créer un corpus de données qui permettra, entre autres, à doter l'avatar d'un modèle de détection des émotions des patients. Pour ce faire, une partie expérimentale a été menée à l'aide du paradigme du magicien d'Oz. Plus précisément, des participants ont réalisé des entraînements cognitifs informatisés accompagnés d'un assistant virtuel qui était en réalité piloté par une expérimentatrice. L'enregistrement audio-visuel des sessions des participants a ensuite été évalué par des annotateurs, du point de vue de l'émotion, afin d'obtenir un corpus permettant d'entraîner des modèles de détection d'émotion. Un travail de réflexion a été mené pour identifier la manière dont les extraits devaient être annotés. Ce travail nous a conduit à inscrire l'annotation dans le cadre des théories de l'appraisal. Ces théories postulent que les émotions émergeraient selon un processus d'évaluation cognitive de la signification des événements et stimuli de l'environnement selon différents critères. Aussi, ces critères ont été proposés aux annotateurs sous la forme de dimensions. En d'autres termes, il a été demandé aux annotateurs d'estimer l'évaluation cognitive que les participants semblaient effectuer dans les extraits audiovisuels sur la base de ces dimensions. De plus, les extraits ont également été évalués selon des labels issus du cadre des émotions d'accomplissement d'activité, i.e., les émotions engagées dans les processus de réalisation de tâche. Dans la mesure où le projet THERADIA a une durée de 5 ans et est toujours en cours, il n'a pas été possible d'obtenir des résultats suffisamment tôt pour les incorporer dans ma thèse.

Le travail réalisé dans cette thèse comporte donc une partie expérimentale visant à étudier les relations entre émotion, attention et tendance à l'action dans le cadre des théories de l'appraisal. À la suite de cette partie expérimentale, une réflexion théorique a été menée dans laquelle nous avons inscrit les postulats des théories de l'appraisal dans ceux de la théorie de l'inférence active. Cette réflexion a été guidée par notre souhait d'intégrer le fonctionnement du processus d'évaluation cognitive proposé par les théories de l'appraisal dans un paradigme dit « prédictif » (ou « Bayésien ») du cerveau, paradigme dans lequel évolue la théorie de l'inférence active. Cette réflexion nous a amené à repenser certaines hypothèses des théories de l'appraisal et à articuler les résultats obtenus dans cette thèse au sein d'un modèle computationnel. Le manuscrit de cette thèse s'articule de la manière suivante :

Un premier chapitre introductif est consacré à dresser un état de la littérature actuelle par rapport aux questions qui sont abordées dans cette thèse. Tout d'abord, les théories contemporaines de l'émotion sont rapidement présentées afin de montrer les atouts des théories de l'appraisal pour décrire les phénomènes affectifs. Ensuite, les hypothèses des théories de l'appraisal sont décrites à travers un modèle émergent de ce cadre théorique, le modèle des processus composants. Sur la base de ce modèle, les critères impliqués dans le processus d'évaluation cognitive sont présentés. Un état de l'art de la littérature portant sur les liens entre attention et émotion est ensuite proposé dans lequel sont décrites les propositions récentes des théories de l'appraisal. Nous présentons notamment l'hypothèse selon laquelle l'attention est capturée en fonction de l'évaluation réalisée par les individus de la pertinence des stimuli selon deux sous-critères : la pertinence intrinsèque (i.e., leur nature plaisante de manière générale) et la pertinence aux buts (i.e., leur possible impact sur la réalisation de buts momentanément importants pour les individus). Un intérêt particulier est ensuite porté à la thématique du biais attentionnel pour les stimuli menaçants, biais exacerbé chez les individus souffrant d'anxiété, ainsi qu'aux questionnements méthodologiques relatifs à la mesure de la capture attentionnelle



vers ces stimuli. Enfin, les concepts manipulés dans la théorie de l'inférence active sont définis, et présentés en relation avec l'hypothèse du codage prédictif et le principe d'énergie libre.

Le second chapitre correspond au premier article que nous avons publié au sujet des liens entre attention et émotion (Fournier & Koenig, 2023). Dans cet article, nous avons mené une étude pour confronter plusieurs hypothèses au sujet de la manière dont l'attention serait capturée par les stimuli. Une première hypothèse propose que l'attention est capturée par les stimuli seulement en fonction de leur niveau de pertinence aux buts, tandis qu'une seconde propose que l'attention est capturée selon les effets combinés de la pertinence intrinsèque et de la pertinence aux buts. Pour modéliser les effets combinés des pertinences, nous avons testé deux types de relation, l'une additive, l'autre multiplicative. Pour ce faire, des participants ont réalisé une tâche de dot-probe dans laquelle des boissons et des aliments étaient mis en compétition. Avant la tâche, les participants ont renseigné le niveau de pertinence intrinsèque, de pertinence aux buts, et de tendance à l'approche associé à chaque stimulus à l'aide d'un questionnaire. Les résultats ont montré que le niveau de capture attentionnelle par les stimuli était mieux expliqué en tenant compte des effets multiplicatifs de la pertinence intrinsèque et de la pertinence aux buts. De plus, le niveau de tendance à l'approche associé aux stimuli était mieux expliqué en tenant compte à la fois de leur niveau de pertinence intrinsèque et de pertinence aux buts.

Le troisième chapitre du manuscrit décrit la deuxième étude que nous avons réalisée, dont l'article est actuellement révisé par les pairs suite à soumission dans une revue. Dans cette étude, nous avons tenté de comparer l'impact attentionnel des effets combinés de la pertinence intrinsèque et de la pertinence aux buts selon que la combinaison des pertinences était agoniste ou antagoniste. Les notions de combinaison agoniste et antagoniste signifient que les pertinences génèrent, respectivement, le même type de tendance à l'action (e.g., l'approche) et des tendances à l'action contraires (i.e., d'approche et d'évitement). Dans cette étude, la

combinaison agoniste des pertinences a été modélisée dans le cas d'une tendance à l'approche générée par le fait que les stimuli soient à la fois plaisants et facilitateurs pour la réalisation d'un but. La combinaison antagoniste des pertinences quant à elle a été modélisée par une tendance à l'évitement générée par le caractère déplaisant des stimuli et une tendance à l'approche générée par le caractère facilitateur des stimuli pour la réalisation d'un but. Pour ce faire, des carrés de différentes couleurs associés aux différentes pertinences étaient mis en compétition dans une tâche de dot-probe. La pertinence aux buts était manipulée par le fait qu'un carré permettait de gagner des points dans une tâche d'induction proposée parallèlement à la tâche de dot-probe. La pertinence intrinsèque était manipulée par le fait qu'un carré générait un bruit plaisant pour un groupe de participants et déplaisant pour un second groupe. De plus, cette étude avait pour objectif de tester l'hypothèse selon laquelle la pertinence intrinsèque a un impact plus fort sur l'attention à courte temporalité, alors que la pertinence aux buts a un impact plus fort sur l'attention à temporalité longue. À cette fin, le temps de présentation des stimuli a été manipulé dans la tâche de dot-probe. Avant, au milieu, et à la fin de la tâche, les participants ont évalué les différents carrés selon leur niveau d'agrément, de désagrément, de facilitation pour le but, de tendance à l'approche et de tendance à l'évitement. D'une part, les résultats ont montré que la tendance à l'approche rapportée à propos des stimuli était mieux prédite en tenant compte à la fois de leur niveau d'agrément, de désagrément et de facilitation pour le but. Plus précisément, des relations positives ont été trouvées entre niveau d'agrément et tendance à l'approche et entre niveau de facilitation pour le but et tendance à l'approche. Au contraire, une relation négative a été mise en évidence entre tendance à l'approche et niveau de désagrément. D'autre part, les résultats ont montré que la tendance à l'évitement à l'encontre des stimuli était également mieux prédite en tenant compte de leur niveau d'agrément, de désagrément et de facilitation pour le but. Cette fois-ci, des relations opposées ont été observées, de sorte que la tendance à l'évitement était reliée négativement au niveau d'agrément et de facilitation pour le

but alors qu'une relation positive a été trouvée entre tendance à l'évitement et niveau de désagrément.

Concernant la relation entre pertinence et attention, les résultats de l'étude ont mis en évidence que les combinaisons agoniste et antagoniste de la pertinence intrinsèque et de la pertinence aux buts avaient des effets combinés sur l'attention, et ce, de manière équivalente. Du point de vue de la temporalité, nos résultats n'ont pas permis de mettre en évidence des impacts plus forts de la pertinence intrinsèque et de la pertinence aux buts à temporalité courte et longue, respectivement.

Le quatrième chapitre décrit la troisième étude que nous avons réalisée, dont le manuscrit est actuellement soumis pour publication. Cette étude avait pour objectif de tester à nouveau l'impact attentionnel des combinaisons agoniste et antagoniste de la pertinence intrinsèque et de la pertinence aux buts lorsque les stimuli peuvent faciliter un but. Cela étant, cette étude visait également à tester l'impact attentionnel des combinaisons agoniste et antagoniste des pertinences dans le cas où la pertinence aux buts était modélisée par une obstruction inévitable. Pour ce faire, des stimuli intrinsèquement pertinents (i.e., plaisant : une poitrine de femme nue ; et déplaisant : une araignée) étaient mis en compétition avec des stimuli neutres (i.e., un champignon et une plante) dans une tâche de dot-probe. En parallèle, les participants réalisaient une tâche d'induction dans laquelle deux stimuli parmi les quatre précédents pouvaient permettre de gagner des points en appuyant sur une touche lors de leur apparition ou faire perdre des points de manière inévitable. Les deux autres stimuli présentés dans la tâche d'induction n'avaient aucun impact sur le score. Les participants étaient répartis par groupes dans lesquels les deux rôles de la tâche d'induction étaient attribués à des stimuli différents. Un groupe a également réalisé la tâche sans que les stimuli n'apparaissent dans la tâche d'induction (en y présentant un autre set de stimuli) afin de mesurer le niveau de capture attentionnelle généré par la pertinence intrinsèque des stimuli sans induction. À l'issue des tâches, les participants

ont évalué le niveau de pertinence intrinsèque des stimuli. Concernant l'évaluation des stimuli, les résultats ont montré que les stimulus définis comme plaisant et déplaisant étaient effectivement évalués comme plus plaisants et déplaisants que les autres, respectivement. De plus, il a été mis en évidence que dans les groupes où les stimuli conduisaient à la facilitation du but (i.e., au gain de points), ces stimuli étaient évalués comme plus plaisants que pour les groupes sans induction. Au contraire, dans les groupes où les stimuli conduisaient à la perte automatique (i.e., à l'obstruction inévitable), ces stimuli étaient évalués comme plus déplaisants que pour les groupes sans induction.

En ce qui concerne les hypothèses relatives aux impacts des combinaisons de pertinence intrinsèque et de pertinence aux buts sur l'attention, des résultats mitigés ont été observés. Conformément à ce qui était attendu, les stimuli intrinsèquement pertinents capturaient plus fortement l'attention que les stimuli neutres en l'absence d'induction. De plus, nous avons observé que lorsque les stimuli conduisaient à la facilitation du but, leur niveau de capture attentionnelle était augmenté et ce, à la fois pour les stimuli intrinsèquement pertinents et les stimuli neutres. Quant à l'impact attentionnel des stimuli conduisant à une obstruction inévitable, des résultats surprenants ont été mis en évidence. En effet, il est apparu que lorsque les stimuli conduisaient à une obstruction inévitable, tous les stimuli captaient l'attention de manière équivalente. Ce résultat a été observé lorsque l'induction d'obstruction inévitable était attribuée aux stimuli neutres et aux stimuli intrinsèquement pertinents.

Le cinquième chapitre du manuscrit est consacré à la quatrième étude que nous avons réalisée, dont le manuscrit a également été soumis à une revue. Cette étude s'est intéressée à des aspects méthodologiques de la mesure du biais attentionnel à l'aide de la tâche de dot-probe dans le cadre de l'anxiété, un champ d'étude dans lequel des divergences sont apparues dans les résultats des travaux de cette dernière décennie. L'objectif de cette étude était de déterminer la meilleure manière de mesurer le niveau de biais attentionnel vers la menace de participants.

Pour ce faire, nous avons : 1) comparé les résultats d'une tâche de dot-probe classique et d'une tâche dite « d'eye-dot-probe » (i.e., une tâche développée dans le cadre de cette thèse dont le côté d'apparition de la cible devait être indiquée avec une saccade oculaire) ; 2) comparé différents indices attentionnels parmi ceux proposés dans la littérature; 3) considéré seulement l'état ou le trait d'anxiété des participants ou un score représentant les effets combinés de l'état et du trait d'anxiété.

Des participants ont ainsi réalisé une tâche de dot-probe et une tâche d'eye-dot-probe dans laquelle des visages menaçants étaient mis en compétition avec des visages neutres. Avant l'étude, les participants ont rempli un questionnaire STAI afin d'évaluer leur niveau de trait et d'état d'anxiété. Les résultats ont montré que la tâche d'eye-dot-probe permettait de mieux décrire le niveau de capture attentionnelle des participants vers les stimuli menaçants selon leur niveau d'anxiété que la tâche de dot-probe. Concernant la question des indices attentionnels, il a été montré que seul un indice attentionnel, « l'Indice de Biais Attentionnel » permettait d'expliquer le niveau de biais attentionnel en direction de la menace mesuré dans les deux tâches. Enfin, dans la tâche de dot-probe, le score qui expliquait le mieux le niveau de capture attentionnelle des participants envers les stimuli menaçants était celui tenant compte des effets combinés du trait et de l'état d'anxiété. Dans la tâche d'eye-dot-probe, ce score était celui tenant compte uniquement du trait d'anxiété des participants.

Le sixième chapitre de la thèse consiste en un essai théorique visant à confronter les théories de l'appraisal et la théorie de l'inférence active, dont le manuscrit a également été soumis pour publication. Dans un premier temps, les hypothèses des théories de l'appraisal ont été décrites. La nature et l'organisation du processus d'évaluation cognitive ont été explicitées. Ensuite, un intérêt particulier a été porté au fonctionnement et à l'organisation du critère de pertinence et à son lien avec l'attention. Après quoi, la question des frontières permettant de distinguer les différents types de phénomènes affectifs (e.g., émotion, humeur) a été abordée ainsi la question

de la différenciation des émotions. Dans un second temps, les hypothèses de la théorie de l'inférence active ont été détaillées. À ce sujet, la proposition de l'architecture et l'organisation du système cérébral a été explicitée selon la proposition de l'hypothèse du codage prédictif et du principe d'énergie libre. L'extension de cette hypothèse dans la théorie de l'inférence active pour décrire l'ensemble des interactions entre les individus et leur environnement a par la suite été abordée. Après quoi, la définition de la notion d'affect dans le cadre de la théorie de l'inférence active a été décrite. Enfin, les liens entre les théories de l'appraisal et la théorie de l'inférence active ont été établis permettant de proposer de nouvelles perspectives concernant les postulats des théories de l'appraisal en tenant compte de ceux de la théorie de l'inférence active.

Le septième chapitre est consacré au projet THERADIA. Dans un premier temps, l'état de la littérature sur les entraînements cognitifs informatisés est présenté. Ainsi, l'intérêt du projet, à savoir, implémenter un assistant virtuel à ces entraînements pour en proposer une version efficiente à domicile, est décrit. La phase expérimentale visant à obtenir un corpus de données pour développer l'assistant virtuel est ensuite présentée, ainsi que la partie d'annotation des extraits audiovisuels issus de la phase expérimentale. Plus précisément, nous décrivons la manière dont les labels et dimensions à annoter au sujet des extraits audiovisuels ont été sélectionnés ainsi que le protocole d'annotation.

Enfin, dans le huitième et dernier chapitre, les résultats empiriques de cette thèse ont été synthétisés pour être articulés dans le cadre d'une proposition d'intégration des hypothèses des théories de l'appraisal dans celles de la théorie de l'inférence active.

# Chapitre 1 - Introduction générale

## 1. Les théories modernes de l'émotion

La première section de ce chapitre a pour objectif de retracer l'histoire des théories modernes de l'émotion dans le but de situer la manière dont les questionnements de cette thèse ont été amenés. Bien qu'elle n'illustre que de manière fragmentaire l'ensemble des théories qui ont contribué à l'état de la recherche actuelle sur l'émotion, cette partie permet de justifier le choix d'avoir inscrit cette thèse dans le cadre des théories de l'appraisal. La structure de cette section est inspirée d'une revue plus exhaustive des théories modernes de l'émotion proposée par Sander et Scherer (2014).

### a. Le débat de la séquence

Les théories modernes de l'émotion trouvent leur origine dans un débat célèbre questionnant la primauté du corps ou de l'esprit dans le déclenchement de l'émotion. Ce débat se résume simplement par la question suivante : « Est-ce que j'ai peur parce que je tremble, ou bien, est-ce que je tremble parce que j'ai peur ? ». Cette question a donné lieu à deux traditions théoriques opposées, soulignant respectivement la primauté du corps ou de l'esprit dans le déclenchement de l'émotion, à savoir, la théorie périphéraliste et la théorie centraliste.

La paternité de la théorie périphéraliste est attribuée à James (1890) et Lange (1885). Leur théorie défendait l'idée que les réactions corporelles suivraient directement la perception des événements et que l'émotion émergerait en réponse à la perception de ces modifications. Cette théorie est liée à plusieurs hypothèses. Premièrement, les émotions ne pourraient être déclenchées sans l'apparition de réactions corporelles. Deuxièmement, les émotions pourraient être différenciées sur la base du schéma de réactions corporelles activé. Enfin, l'activation d'un schéma spécifique de réactions corporelles déclencherait l'émotion qui lui est associée, quel que soit la manière dont ce schéma a été activé. Ce dernier point pourrait être illustré par le fait

que produire un sourire de manière non-consciente (e.g., en demandant à des individus de tenir un stylo dans leur bouche) augmenterait le ressenti d'émotion positive (Strack et al., 1988).

Plus tard, la théorie centraliste de Cannon (1927) et Bard (1928) a été proposée, supposant qu'une analyse des événements suite à leur perception déclencherait l'émotion, ce qui entraînerait ensuite des réactions physiologiques. Naturellement, cette théorie s'est fermement opposée à la théorie périphéraliste. À son sujet, Cannon et Bard ont critiqué le fait que des schémas corporels identiques étaient communs à certaines émotions telles que la peur et la colère, mais aussi à des états non-émotionnels tels que la digestion ou la fièvre. De plus, les auteurs ont avancé que les réactions corporelles ne seraient pas nécessaires au déclenchement de l'émotion. Pour illustrer cette hypothèse, les auteurs ont mis en évidence que des chats ayant subi une sympathectomie (i.e., déconnexion des viscères du système nerveux central) pouvaient toujours exprimer des réactions émotionnelles.

À propos de ce débat, il est à noter que la théorie périphéraliste a continué d'influencer la recherche jusqu'à récemment. En effet, ses postulats ont été affinés au sein de théories dites « néo-jamesienne » (Lavazza, 2009). Il convient aussi de mentionner que les hypothèses formulées par James dans ses travaux originaux ont pu être parfois caricaturées dans la littérature. En effet, les nuances apportées par James dans ses ouvrages se sont quelque peu estompées au profit de sa notoriété (voir Ellsworth, 1994 pour une discussion).

## **b. La théorie bifactorielle**

Ce débat a ensuite donné naissance à l'une des premières théories de l'émotion plaçant la cognition au cœur des émotions, la théorie bifactorielle (Schachter & Singer, 1962). Cette théorie avait pour objectif de réconcilier la théorie périphéraliste et la théorie centraliste en proposant un rôle central à la fois du corps et de l'esprit (i.e., la cognition), dans le déclenchement de l'émotion. Plus précisément, selon la théorie bifactorielle, l'émotion émergerait de l'interaction de deux facteurs : une activité physiologique et une analyse



cognitive de la cause du déclenchement de cette activité. Dans ce contexte, l'activité physiologique répondrait directement aux événements (comme proposée par la théorie périphéraliste) mais déterminerait seulement l'intensité de l'émotion. En effet, une analyse cognitive serait nécessaire à l'émergence de l'émotion (comme proposé par la théorie centraliste), visant à donner un sens à cette activité en rapport avec l'événement qui l'aura déclenchée. Cette théorie a beaucoup inspiré la littérature de l'émotion dans la deuxième moitié du 20<sup>e</sup> siècle, avant d'être peu à peu abandonnée du fait d'un certain nombre de critiques à son sujet (Reisenzein, 1983). De plus, la théorie bifactorielle (au même titre que la théorie périphéraliste) ne rend pas compte de la nature du déclenchement de l'activité physiologique face aux événements, point crucial pour donner un cadre à l'étude des phénomènes émotionnels.

### **c. La théorie des émotions de base**

L'une des théories les plus influentes de la recherche sur l'émotion, la théorie des émotions de base, a ensuite été portée par Ekman (1992). Cette théorie prend ses racines dans les travaux de Darwin (1872/1998) selon lesquels les émotions sont des phénomènes universels résultant de l'évolution. Plus précisément, la théorie des émotions de base suppose que tout être humain serait doté d'un petit nombre d'émotions dites de base : la joie, la colère, la peur, la tristesse, le dégoût et la surprise. Chacune de ces émotions serait associée à des fonctions, des conditions d'émergence, des patterns physiologiques, expressifs, et neuronaux spécifiques et universels issus de l'évolution. Si la vision trop rigide de cette théorie a conduit la recherche actuelle à se tourner vers des approches plus flexibles, il n'en reste pas moins que la théorie des émotions de base a permis de grandes avancées dans le champ de l'émotion. On peut par exemple citer le développement du Facial Action Coding System (FACS ; Ekman & Friesen, 1978) encore largement utilisé dans la recherche actuelle, permettant de codifier toutes les unités d'action de l'expression faciale indépendamment d'un parti pris idéologique.

#### **d. La théorie bi-dimensionnelle**

La première version des théories dites « dimensionnelles », visant à différencier les émotions selon un plan à plusieurs dimensions, a été proposée à la fin du 19<sup>e</sup> siècle par Wundt (1897). Selon cette version, l'état émotionnel pourrait être représenté par les coordonnées d'un point sur un plan à trois dimensions selon le niveau de plaisir ou d'ennui, le niveau de tension ou de relaxation et le niveau d'état mental, plutôt actif ou passif. Bien qu'un certain nombre de théories dimensionnelles aient vu le jour par la suite, la plus célèbre est sans doute celle de Russel (1980). Dans sa théorie bidimensionnelle, Russel a proposé d'inscrire les états émotionnels autour du « Circumplex », un plan cartésien défini par la valence, allant de « négatif » à « positif », et l'arousal (ou éveil), allant de « désactivé » à « activé ». Cette approche est encore très répandue actuellement, particulièrement dans le domaine de la modélisation computationnelle (e.g., Garg & Verma, 2020; Rutkowski et al., 2020). L'engouement à son égard vient probablement du fait de sa simplicité de compréhension et d'application. Par exemple, cette approche permet de distinguer deux émotions négatives comme la peur, associée à un fort arousal, et la tristesse, associée à un faible arousal. Cependant, le niveau de différenciation des émotions du Circumplex a été vivement critiqué, car celui-ci ne permet pas de distinguer la peur de la colère, par exemple, toutes deux définies comme des émotions fortement intenses et négatives (Fontaine et al., 2007).

#### **e. Les théories de l'appraisal**

Le cœur des théories de l'appraisal, dont la parenté est généralement attribuée à Arnold (1969), réside dans le fait que les émotions seraient déclenchées et différenciées sur la base d'un processus d'évaluation cognitive basé sur la signification de l'environnement (Lazarus & Folkman, 1984). L'une des forces de ces théories est qu'elle permet d'intégrer – dans une certaine mesure – l'ensemble des idées proposées par les théories présentées précédemment.

Premièrement, la place de la cognition (i.e., le processus d'évaluation cognitive mentionné ci-dessus) est centrale dans l'émotion, comme proposée dans la théorie centraliste et la théorie bifactorielle. Deuxièmement, les théories de l'appraisal proposent que les sous-systèmes de l'organisme (comprenant les réactions physiologiques) impacteraient le processus d'évaluation cognitive (Grandjean & Scherer, 2014), comme proposée dans la théorie périphéraliste et la théorie bifactorielle. Troisièmement, les théories de l'appraisal intègrent les postulats de la théorie des émotions de base selon une interprétation nuancée. Dans ce contexte, les émotions de base ne seraient pas différentes des autres émotions. Comme toute émotion, elles seraient associées à un certain type de fonctions, de conditions d'émergence (i.e., des patterns d'évaluation cognitive) et de patterns physiologiques, expressifs et neuronaux (Ellsworth, 1991). Cependant, le caractère universel de ces émotions ne reflèterait pas seulement un héritage phylogénétique rigide commun à tout être humain, comme proposé par la théorie des émotions de base. En effet, l'universalité des émotions de base émergerait également de traits communs dans l'expérience culturelle des individus. L'idée est que des individus, certes, issus de cultures très éloignées, seraient finalement sujets aux mêmes besoins (e.g., se nourrir, sociabiliser, éviter les menaces) au sein d'environnements aux caractéristiques semblables (e.g., présence de nourriture, d'autres êtres humains, de menaces). Face au nombre infini d'interactions environnementales imaginables, la loi forte des grands nombres postule que l'occurrence répétée d'un même événement qui semble improbable devient évident face à l'immensité des possibles. Aussi, le caractère universel des émotions de base serait en partie une coïncidence, résultant d'une probabilité d'occurrence plus forte pour ces émotions que les autres. En d'autres termes, du fait du nombre important d'émotions expérimentées par les individus, il serait peu probable que certaines ne se retrouvent pas dans toutes les cultures. Paulos écrivait d'ailleurs à ce sujet : « En réalité, la coïncidence la plus étonnamment incroyable que l'on puisse imaginer serait l'absence totale de coïncidences. » (cité dans Myers, 2002,

p.238). Enfin, les théories de l'appraisal peuvent être elles aussi comprises en tant que théorie dimensionnelle, dans laquelle chaque dimension reflèterait un des critères du processus d'évaluation cognitif, critères qui seront détaillés dans la section suivante (Scherer, 2009).

## **2. Focus sur les théories de l'appraisal**

Tout d'abord, les théories de l'appraisal ne doivent pas être réduites à un cadre théorique visant simplement à expliquer l'émotion. Elles s'inscrivent dans le courant épistémologique de l'affectivisme qui souligne l'importance de penser tous les modèles du comportement, du cerveau et de l'esprit en tenant compte du rôle de l'affect (Dukes et al., 2021). La notion d'affect fait ici référence à l'ensemble des phénomènes à composantes multiples qui intensifient ou atténuent les réponses corporelles, façonnant ainsi les interactions environnementales des individus (Massumi, 2002). L'idée forte du courant affectiviste invite à considérer la frontière entre les phénomènes non-affectifs et affectifs, tels que les émotions ou les humeurs, comme plus fines que ce qui a pu être proposé dans la littérature de ces dernières décennies. Aussi, l'un des objectifs principaux des théories de l'appraisal est de définir précisément les limites de ces différents types de phénomènes, afin de pouvoir les séparer de manière satisfaisante. Pour ce faire, il s'agit d'abord de définir la nature des interactions entre les individus et leur environnement ainsi que les processus qui impliqués dans ces interactions.

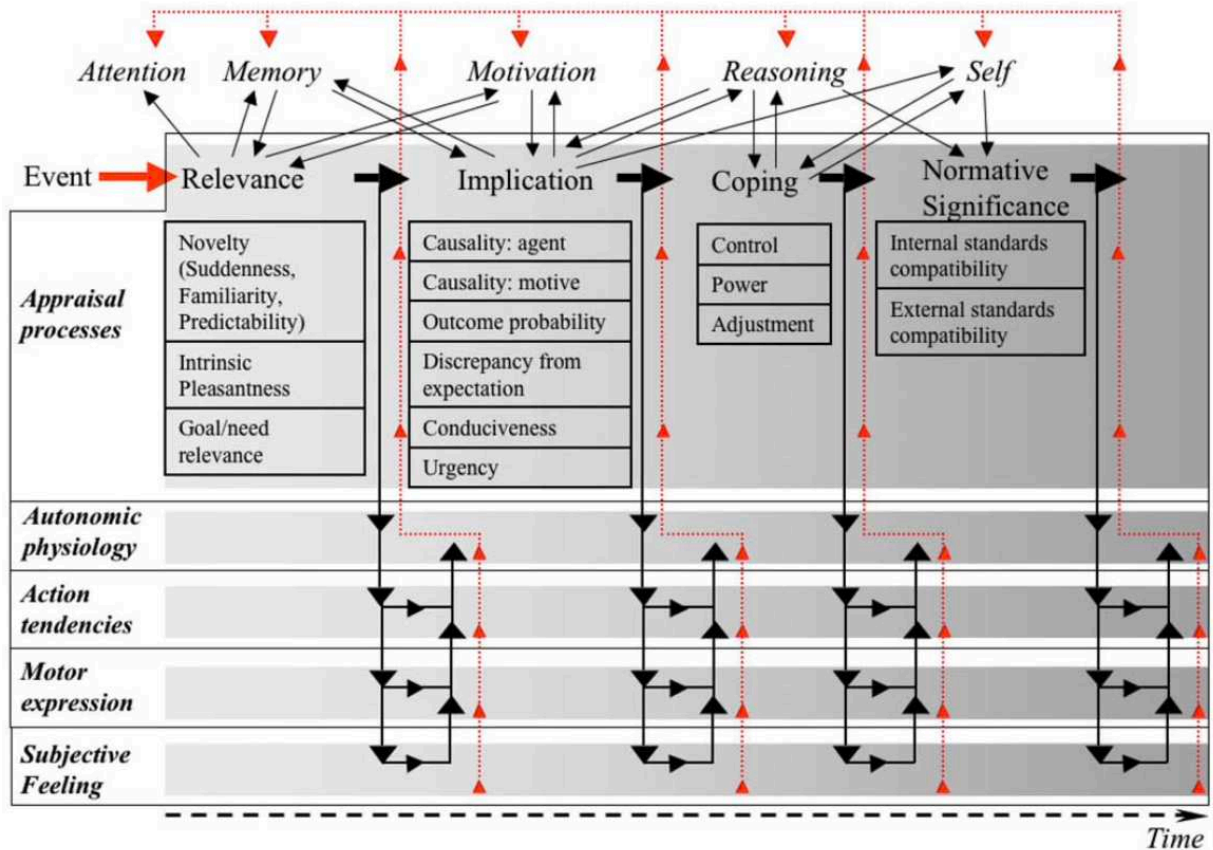
### **a. Le Modèle des Processus Composants**

Le modèle des Processus Composants (MPC) propose justement un cadre explicatif précis de la manière dont les épisodes d'interaction entre les individus et leur environnement sont régis. Selon le MPC (Grandjean & Scherer, 2014), comprendre ces épisodes nécessite de s'intéresser à l'imbrication de cinq sous-systèmes de l'organisme : la cognition, la physiologie, la tendance à l'action, l'expression motrice, et le sentiment subjectif (voir Figure 1). Dans ce modèle, l'accent est mis sur le sous-système cognitif, divisé en deux parties : les critères et sous-critères du processus d'évaluation cognitive (i.e., les *Stimulus Evaluation Checks* dans

leur première forme ; Scherer, 1984) et les fonctions cognitives qui participent à la mise en place de ce processus.

**Figure 1**

*Le Modèle des Processus Composants.*



*Note. Les colonnes représentent les différents critères et sous-critères du processus d'évaluation cognitive réalisé par les individus à propos de l'environnement. Les flèches représentent la dynamique temporelle de ces critères et sous-critères ainsi que la manière dont ils impactent et sont impactés par les autres sous-systèmes de l'organisme représentés en lignes. Figure tirée de "A systems approach to appraisal mechanisms in emotion" par Sander et al., 2005.*

### **i. Le processus d'évaluation cognitive**

Le processus d'évaluation cognitive occupe une place centrale au sein du modèle car ce processus serait le chef d'orchestre des sous-systèmes de l'organisme au sein des épisodes d'interactions environnementales. Plus précisément, leur rôle serait d'évaluer la signification des stimuli par rapport au bien-être des individus, afin de synchroniser de manière adaptée les autres sous-systèmes de l'organisme dans le but de faciliter l'accomplissement des réponses adaptées. Le processus d'évaluation cognitive serait divisé en quatre critères d'évaluation, eux-mêmes divisés en sous-critères (voir Figure 1).

#### 1. La pertinence

Le critère de pertinence a pour objectif d'évaluer dans quelle mesure les stimuli sont pertinents pour les individus. Étant donné que les ressources attentionnelles sont limitées (Kahneman, 1973), cette fonction est capitale pour focaliser les ressources des individus de manière adaptée dans leur perception de l'environnement. La pertinence serait établie grâce à l'évaluation de trois sous-critères, la nouveauté, la pertinence intrinsèque et la pertinence aux buts :

- 1) L'évaluation de la nouveauté consisterait en un premier filtre (de plutôt bas niveau) concernant la pertinence des stimuli. Cette évaluation des stimuli serait réalisée selon trois dimensions : leur niveau de soudaineté, leur niveau de familiarité et leur niveau de prédictibilité. Du point de vue de la soudaineté, un stimulus sera considéré comme pertinent si celui-ci est plutôt intense et apparaît de manière abrupte dans le champ perceptif des individus. Cela peut faire référence à un son, une odeur, ou bien une couleur intense qui témoignerait d'une différence dans l'environnement par rapport à une perception antérieure.
- 2) L'évaluation de la pertinence intrinsèque permettrait d'établir si les stimuli sont plutôt plaisants ou déplaisants de manière générale, et ce, indépendamment des buts et besoins

actuels des individus. Ce sous-critère est en lien direct avec les préférences des individus (Scherer, 2005b), qu'elles aient pu être acquises de manière phylogénétique (i.e., héritées par les gènes de génération en génération) ou ontogénétique (i.e., acquises par expérience). L'indépendance soulignée de la pertinence intrinsèque avec les buts et besoins actuels des individus met en avant une certaine forme de rigidité dans son évaluation. Pour illustrer cela, l'exemple du gâteau au chocolat a beaucoup été utilisé. Imaginons qu'un individu évalue la pertinence intrinsèque de ce gâteau comme plaisante. L'idée est que dans une situation donnée, même si l'individu est rassasié après avoir mangé 8 parts de ce gâteau, il évaluera toujours sa pertinence intrinsèque comme plaisante. La terminologie concernant ce sous-critère a évolué à plusieurs reprises depuis sa première proposition (Scherer, 1984). En effet, celle-ci a pu par exemple être dénommée « agrément intrinsèque » ou « plaisir intrinsèque ». En accord avec Scherer (2013) et dans un but de clarté, la terminologie au sujet de ce sous-critère sera définie comme « pertinence intrinsèque » dans l'ensemble du présent manuscrit.

- 3) L'évaluation de la pertinence aux buts viserait à établir la pertinence des stimuli selon qu'ils peuvent faciliter ou obstruer la réalisation des buts et/ou besoins momentanément importants pour l'individu. Ce sous-critère est la deuxième pièce du puzzle concernant la pertinence totale du gâteau au chocolat dans l'exemple précédent. À ce sujet, trois cas de figure pourraient être imaginés : le gâteau au chocolat est pertinent du point de vue de la pertinence aux buts car il peut faciliter le besoin de recherche de satiété si l'individu a faim. Mais sa pertinence aux buts peut aussi émerger de son caractère obstructif si l'individu est rassasié et que le gâteau peut engendrer du dégoût. Enfin, sa pertinence aux buts pourra être neutre si l'individu est dans un état homéostatique et que le fait de satisfaire sa faim ou éviter le dégoût n'est pas une préoccupation actuelle. Aussi, la pertinence aux buts permet de rendre compte de différences intra-individuelles dans

l'évaluation de la pertinence des stimuli selon l'état de l'organisme. Il est à noter que dans les trois cas de figure, la pertinence intrinsèque du gâteau restera inchangée par rapport aux préférences initiales de l'individu. Comme pour la pertinence intrinsèque, la terminologie concernant la pertinence aux buts a évolué au cours du temps (Scherer, 1984, 2013). Par souci de concision et de clarté, la terminologie au sujet de ce sous-critère sera définie comme « pertinence aux buts » dans la suite du présent manuscrit.

## 2. L'implication

Pour que la pertinence aux buts des stimuli soit affinée, il est nécessaire d'en mesurer leurs implications potentielles. Comme son nom l'indique, le critère d'implication a pour objectif d'évaluer l'impact des stimuli par rapport aux buts et besoins courants des individus. Comme précisé par Grandjean et Scherer (2014), le concept de buts et besoins est un terme générique qui couvre l'ensemble des conduites, besoins, instincts, motifs, ou buts qui motivent les individus à agir pour atteindre leur bien-être et préserver ou assurer leur survie. L'implication serait établie via l'évaluation de cinq sous-critères, l'attribution causale, la probabilité des conséquences, la différence avec les attentes, la facilitation/obstruction aux buts et besoins et l'urgence :

- 1) L'évaluation de l'attribution causale aurait pour objectif d'établir si l'événement en cours d'évaluation est causé par un agent animé ou non. Par exemple, si un individu remarque que l'un des pneus de sa voiture est crevé, ce sous-critère viserait à déterminer si un autre individu est responsable de cela, ou bien s'il s'agit d'un défaut de fabrication ou du fait qu'il a roulé sur un clou. Dans le cas où un agent animé a été identifié, l'évaluation de ce sous-critère serait affinée pour caractériser les motivations de l'agent. En d'autres termes, est-ce que la crevaison du pneu a été réalisée de manière intentionnelle à la suite d'un conflit entre les protagonistes ou bien est-ce dû à un accident ?



- 2) L'évaluation de la probabilité des conséquences viserait à établir les répercussions possibles des événements et leur probabilité respective sur les états futurs des individus et de leur environnement. Dans la continuité de l'exemple précédent, imaginons que notre individu identifie un voisin comme responsable de la crevaison du pneu de sa voiture, dont les intentions seraient perçues comme une vengeance à la suite d'une dispute. L'évaluation du sous-critère de la probabilité des conséquences aurait pour objectif d'inférer si le voisin va probablement avouer ou nier son méfait une fois confronté et s'il risque d'accepter ou non de participer à la réparation du pneu.
- 3) L'évaluation de la différence avec les attentes établirait l'écart entre l'état de l'événement évalué et les croyances à son sujet. Ce sous-critère répond directement à celui de la probabilité des conséquences : si notre individu confronte son voisin concernant la crevaison du pneu de sa voiture, que le voisin reconnaît sa responsabilité, s'excuse et s'engage à réparer le pneu, un haut degré de décalage résulterait du critère de différence avec les attentes.
- 4) L'évaluation de la facilitation/obstruction aux buts et besoins déterminerait ensuite si les conséquences des événements vont plutôt amener à faciliter ou obstruer la réalisation des buts et besoins de l'individu. Dans la situation précédente, il semble évident que le pneu crevé sera évalué comme obstructif sur plusieurs plans pour notre individu. En effet, cet événement peut engendrer une situation désagréable comme une confrontation avec son voisin, l'engagement de frais dans la réparation du pneu, ou encore du retard dans les déplacements prévus.
- 5) Enfin, l'évaluation de l'urgence pourrait être définie comme l'approximation d'un poids attribué à l'importance des buts et besoins dans leur hiérarchie d'importance actuelle. En d'autres termes, ce sous-critère viserait à pondérer la nécessité d'accomplir un but ou un

besoin selon son importance actuelle et la pression temporelle qui lui est associée. Notre individu pourrait évaluer que réparer son pneu crevé n'est pas urgent s'il peut se rendre à son travail en transport en commun. Cependant, la réparation du pneu serait évaluée comme urgente si notre individu devait partir en vacances en voiture quelques heures plus tard.

### 3. Le potentiel de maîtrise

Une fois que les implications des événements ont été évaluées, les individus doivent ensuite évaluer jusqu'à quel point ils peuvent s'adapter aux conséquences produites. L'évaluation du potentiel de maîtrise vise à répondre à cette question. Dans le cas de conséquences évaluées comme facilitatrices pour les buts et besoins courants des individus, celles-ci seront à priori acceptées sans conflit. Aussi, l'évaluation du potentiel de maîtrise est majoritairement en jeu lorsque les conséquences sont évaluées comme étant obstructrices. Dans ce contexte, le problème à résoudre vise à évaluer l'ensemble des réponses possibles à ces conséquences et leur degré d'acceptabilité. Pour ce faire, trois sous-critères seraient évalués, le contrôle, la puissance et l'ajustement :

- 1) L'évaluation du contrôle, permettrait de déterminer la probabilité que les conséquences (plutôt obstructrices) d'un événement puissent être contrôlées par un agent animé. Dans la continuité de l'exemple concernant le pneu crevé, notre individu évaluerait la probabilité que les conséquences de cet événement (i.e., ne pas partir en vacances), puissent être contrôlées, en faisant par exemple réparer le pneu de sa voiture rapidement par un garagiste, ou l'un de ses amis.
- 2) En cas de contrôle possible, l'évaluation de la puissance aurait ensuite pour objectif d'évaluer la probabilité qu'un agent naturel puisse contrôler les conséquences de l'événement. Si les sous-critères de contrôle et de puissance semblent identiques, une

nuance réside en fait dans leur objet d'étude respectif. L'évaluation du sous-critère de contrôle concerne un calcul probabiliste orienté vers l'ensemble des possibles relatif au contrôle des conséquences. Au contraire, l'évaluation du sous-critère de puissance se focalise sur la probabilité d'un de ces possibles. En d'autres termes, si l'évaluation du critère de contrôle donne comme résultat la probabilité que les conséquences puissent être contrôlées, l'évaluation de la puissance donne comme résultat la probabilité qu'une des possibilités envisagées puissent être mise en œuvre. Par exemple, quelle est la probabilité que le garagiste soit en mesure de réparer le pneu de la voiture avant que notre individu parte en vacances ?

- 3) Une fois les probabilités de contrôle et de puissance établies, l'évaluation du critère d'ajustement aura pour objectif de déterminer dans quelle mesure l'individu pour s'adapter aux obstrucrices d'un événement, en tenant compte du contrôle et de la puissance envisagées. L'évaluation de l'ajustement pourra alors donner lieu à deux réponses possibles : la résignation ou la confrontation. Dans le cas de la résignation, notre individu pourrait accepter le fait de ne pas partir en vacances si aucun contrôle n'est envisageable pour réparer le pneu de sa voiture. Au contraire, il pourra réagir en mobilisant des ressources supplémentaires pour trouver un autre moyen de ne pas subir les conséquences de l'évènement, comme choisir un autre moyen de transport.

#### 4. La signification normative

Une fois que le processus d'évaluation cognitive touche à sa fin après avoir évalué la pertinence, l'implication et le potentiel de maîtrise d'un événement, l'individu aura probablement évalué une réponse comme étant celle qu'il choisira de mettre en place face à l'évènement. Cette réponse serait évaluée par un dernier critère, visant à la confronter par rapport aux standards, normes et valeurs de l'individu ou de la société. Cette évaluation

comprendrait deux sous-critères, l'évaluation des standards internes et l'évaluation des standards externes.

- 1) L'évaluation des standards internes viserait à approximer le degré de correspondance entre la réponse qu'un individu a sélectionné face à un événement et la représentation qu'il se fait de lui-même. Dans un cas extrême où notre individu se décide à voler une voiture pour partir en vacances à n'importe quel prix, cette réponse pourrait violer ses standards internes s'il considère – à juste titre – que le vol n'est pas une réponse acceptable.
- 2) L'évaluation des standards externes aurait pour objectif de déterminer si la réponse sélectionnée est en accord avec les standards que l'individu imagine pour son groupe social. Il s'agirait ici d'approximer le degré de correspondance entre la réponse sélectionnée face à événement et sa représentation de la réponse attendue par son groupe social. En poursuivant avec l'exemple de notre individu, il est possible que le fait de voler une voiture ne viole pas ses standards internes. Cependant, cette réponse pourrait violer ses standards externes. Par exemple, la représentation qu'il imagine que ses parents se font de la réponse envisagée pourrait le conduire à ne finalement pas voler de voiture.

#### **ii. L'aspect temporel du processus d'évaluation cognitive**

Si la proposition des critères d'évaluation cognitive au sein du MPC semble en accord avec les théoriciens de l'appraisal, la nature du processus et son organisation temporelle font débat. En se basant sur l'hypothèse d'un fonctionnement en terme de réseaux associationistes (Bower, 1981), certains auteurs suggèrent que l'évaluation de critères opérerait selon un processus d'activation d'associations (Forgas & Smith, 2007; C. Smith & Lazarus, 1990). Plus précisément, l'évaluation pourrait être décrite par un mécanisme de correspondance entre les caractéristiques extraites des stimuli et d'anciennes associations à propos de ces caractéristiques

concernant les différents critères. Selon un tel fonctionnement, la séquence temporelle du processus d'évaluation des critères serait flexible et pourrait même se dérouler en une seule étape. D'autres auteurs supposent que la séquence temporelle de l'évaluation est fixe, de sorte que l'évaluation de critères d'un traitement de niveau supérieur ne peut pas se produire avant l'évaluation d'un critère de niveau inférieur, bien qu'ils puissent se produire en parallèle (Grandjean & Scherer, 2008; Sander et al., 2005; Scherer, 1984). Cette position est défendue par les auteurs du MPC. Au sein du modèle, la notion de niveau de traitement entre les critères d'évaluation doit être comprise selon l'ordre séquentiel suivant (du plus bas au plus haut niveau) : la pertinence, l'implication, le potentiel de maîtrise, et la signification normative (voir Figure 1). Le raisonnement qui sous-tend cette hypothèse est que les résultats des critères d'évaluation des niveaux inférieurs sont une condition préalable au bon fonctionnement des critères d'évaluation des niveaux supérieurs. D'autres auteurs encore, se situent quelque part entre ces deux points de vue, supposant que la séquence temporelle des critères d'évaluation est flexible, à condition que le sous-critère de nouveauté ait été évalué préalablement (Ellsworth, 1991). Cette position met l'accent sur l'existence d'une micro-séquence dans le processus d'évaluation, dans laquelle l'évaluation de la nouveauté des stimuli est une condition nécessaire au déclenchement de l'évaluation des autres critères.

### **iii. La relation entre le processus d'évaluation cognitive et les sous-systèmes de l'organisme**

Durant le processus d'évaluation cognitive, le MPC suppose que les résultats d'évaluation de critères et sous-critères déclencheraient une activation synchronisée des sous-systèmes de l'organisme. Ces activations synchronisées auraient un but adaptatif, à savoir, faciliter les séquences comportementales que les individus auront sélectionnées comme étant les meilleures face à la signification évaluée des événements.

### 1. Les réactions physiologiques

Les réactions physiologiques auraient une fonction de support dans la conduite des séquences comportementales adaptatives (Levenson, 2006; Öhman & Wiens, 2009). En d'autres termes, une fois une séquence comportementale sélectionnée, l'état physiologique du corps serait défini sur un niveau facilitant cette séquence. Par exemple, dans un contexte où un individu se retrouve face à un animal menaçant, une accélération du rythme cardiaque pourrait être déclenchée afin d'assurer une oxygénation suffisante des muscles recrutés pour prendre la fuite. Comme mentionné précédemment, des réactions physiologiques spécifiques seraient donc déclenchées sur la base du résultat des différents critères d'évaluation cognitive (Pecchinenda, 2001; van Reekum et al., 2004).

### 2. Les tendances à l'action

Les tendances à l'action pourraient être définies comme une forme de préparation des séquences comportementales adaptatives (Frijda, 2009). La littérature fait état de deux grandes classes de tendance à l'action : l'approche et l'évitement (Maxwell & Davidson, 2007; Rutherford & Lindell, 2011), elles-mêmes divisées en sous-classes : se rapprocher, utiliser, posséder, pour l'approche ; éviter, fuir, se protéger, pour l'évitement. Au même titre que les réactions physiologiques, l'activation de ces différentes classes de tendance à l'action pourrait être différenciée sur la base des patterns de critères durant le processus d'évaluation cognitive (Frijda, 1987; Moors, 2014).

### 3. Les expressions motrices

Les expressions motrices auraient principalement une fonction de communication dans l'accompagnement des séquences comportementales adaptatives (Ekman et al., 1972). Cette fonction viserait à transmettre une intention aux autres individus présents lors de l'interaction. Les expressions motrices font référence à la posture, la prosodie de la voix, ou encore aux expressions faciales. Par exemple, si un individu subit une agression, le fait d'élever la voix ou

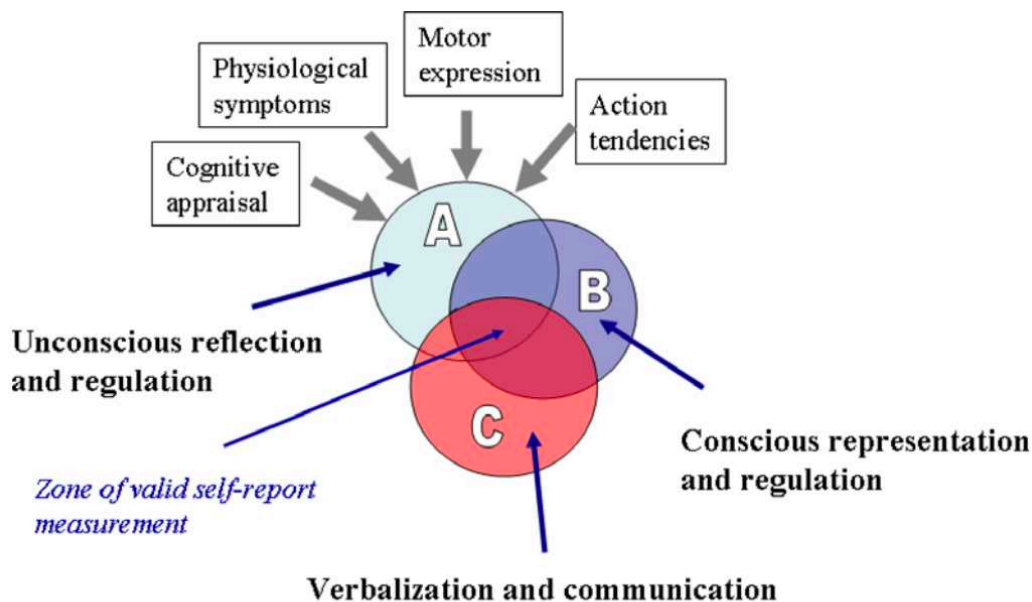
de froncer les sourcils pourrait être une manière de signaler à l'agresseur que l'individu ne compte pas se laisser faire. De la même manière que les sous-systèmes précédents, des patterns d'expressions motrices seraient déclenchés par des résultats spécifiques de critères pendant le processus d'évaluation. Une attention particulière a été portée lors de la dernière décennie à l'établissement de liens causaux entre les critères d'évaluation et les patterns d'expression faciale (Scherer et al., 2013, 2019; Scherer & Ellgring, 2007).

#### 4. Le sentiment subjectif

Bien que cela n'ait pas été mentionné précédemment, les théoriciens de l'appraisal postulent que les épisodes du processus d'évaluation cognitive et de l'activation synchronisée des sous-systèmes de l'organisme opéreraient en grande partie hors du champ de la conscience (Grandjean et al., 2008). Le processus d'évaluation cognitive serait effectué à plusieurs niveaux de traitement cognitif, depuis un niveau sensorimoteur, jusqu'à un niveau conceptuel (Scherer, 2005a). La partie consciente de ces épisodes (voir Figure 2), définie comme le sentiment subjectif, émergerait si les critères sont évalués à un niveau cognitif suffisamment élevé et si l'activation des sous-systèmes de l'organisme atteint des seuils importants de synchronisation (Dan Glauser & Scherer, 2008; Scherer, 2009).

**Figure 2**

*Diagramme de Venn représentant le sentiment subjectif du processus d'évaluation cognitive et de l'activation des sous-systèmes de l'organisme.*



*Note. L'ensemble A représente la partie non-consciente des régulations et représentations intégrées, ensemble dans lequel le processus d'évaluation cognitive et l'activation des sous-systèmes de l'organisme opère. L'ensemble B reflète la partie consciente des régulations et représentations intégrées. L'ensemble C fait référence à la partie verbalisable et communicable des expériences des individus. Le sentiment subjectif du processus d'évaluation cognitive et de l'activation des sous-systèmes de l'organisme correspond à la zone d'intersection des ensembles A, B et C. Figure adaptée de "Conscious emotional experience emerges as a function of multilevel, appraisal-driven response synchronization" par Grandjean et al., 2008.*

#### **iv. La rétroaction des sous-systèmes de l'organisme et le reappraisal**

Si l'activation synchronisée des sous-systèmes de l'organisme est déclenchée par les différentes évaluations des critères, le MPC postule que les états d'activation des sous-systèmes de l'organisme influenceraient en retour le résultat des critères d'évaluation (Grandjean & Scherer, 2014). Cette proposition s'inscrit dans le cadre de l'hypothèse de la rétroaction



corporelle selon laquelle la signification de l'environnement pourrait être modulée par les états d'activation du corps. Par exemple, comme il l'a été évoqué précédemment, les individus auraient tendance à évaluer un dessin animé comme plus humoristique s'ils tiennent un stylo dans leur bouche qui les force à activer les muscles liés au sourire (Strack et al., 1988). Cette proposition peut aussi être illustrée par l'hypothèse des marqueurs somatiques qui suppose que les états du corps influencent la prise de décision (Bechara et al., 2005; Harrison et al., 2016; Poppa & Bechara, 2018; Sugawara et al., 2020). Une étude de cas bien connue de cette tradition de recherche a montré que les réponses physiologiques déclenchées par l'évaluation des conséquences négatives des stimuli pouvaient influencer la prise de décision liée à ces stimuli dans les interactions ultérieures (Bechara et al., 2005).

Le résultat émergent de l'évaluation des stimuli selon les différents critères et sous-critères ne serait pas définitif à l'issue d'une seule séquence d'évaluation. En effet, une fois un premier processus d'évaluation cognitive effectué, ce résultat serait affiné au cours de répétitions multiples du processus, tenant compte des résultats des processus évaluatifs antérieurs et des états d'activation des sous-systèmes de l'organisme déclenchés (Ellsworth & Dougherty, 2016; Gross, 2015; Uusberg et al., 2023). On parle alors de réévaluation ou de reappraisal. Par analogie, ce phénomène d'affinage, pourrait être illustré par le travail d'un sculpteur qui ne cesse d'élaguer les courbes d'un bloc de granit jusqu'à obtenir une œuvre qu'il juge satisfaisante.

### **b. La nature des épisodes : non-affectifs ou émotionnels ?**

Bien que le MPC ait été initialement proposé pour rendre compte du concept d'émotion en tant qu'épisode, comme le propose les théories de l'appraisal (Folkman & Lazarus, 1985 ; Frijda, 1986 ; Moors et al., 2013 ; Scherer, 1993), la notion d'émotion a volontairement été évitée lors de la partie précédente détaillant le fonctionnement du modèle. Comme spécifié préalablement, les épisodes décrits par le MPC seraient impliqués dans la majorité des

interactions entre les individus et leur environnement, afin de donner un sens aux événements et de sélectionner les actions les plus adaptées. Cette idée inscrit sans équivoque les théories de l'appraisal et le MPC au sein du cadre de l'affectivisme, car elle suppose que la nature des épisodes non-affectifs et des épisodes affectifs, tels que les émotions, ne diffère fondamentalement pas (Dukes et al., 2021).

Pour illustrer cela, imaginons deux situations : l'une, représentant le déroulement d'un épisode non-affectif, où un individu ressent un état de soif et va réaliser une action lui permettant de satisfaire sa soif. L'autre, représentant le déroulement d'un épisode émotionnel, où ce même individu ressent un état de soif, mais n'est pas en mesure de satisfaire son besoin car l'environnement ne le permet pas (e.g., un désert aride), ce qui l'amène à ressentir de la peur. Ces deux épisodes sont déclenchés par l'évaluation d'un même événement : l'état interne de soif. Dans les deux cas, cet événement sera évalué selon les mêmes critères d'évaluation cognitive et le processus d'évaluation pourra déclencher une activation synchronisée des sous-systèmes de l'organisme afin de sélectionner l'action la plus adaptée. Cela amène donc à se poser la question suivante : comment différencier un épisode non-affectif d'un épisode émotionnel ? Selon les théories de l'appraisal, il s'agirait d'une question de remplissage de conditions (Sander, 2013; Scherer, 2005b). Un épisode doit être considéré comme émotionnel si : a) il est déclenché par un stimulus ; b) le stimulus a été évalué comme pertinent ; c) la synchronisation des sous-systèmes de l'organisme est importante ; d) l'impact comportemental est très élevé (par exemple, interruption des comportements en cours) ; et e) l'épisode est intense et bref. Dans notre exemple, si les deux épisodes sont tous deux déclenchés par un événement évalué comme pertinent (i.e., le ressenti de soif), le premier épisode ne remplira pas toutes les conditions et ne sera donc pas qualifié d'émotionnel. Dans le cas où une manière de satisfaire son besoin de soif est identifiée, la nécessité d'activer une importante synchronisation des sous-systèmes de l'organisme n'est pas nécessaire. Dans le second cas, où aucune manière de

satisfaire la soif n'a été identifiée, l'évaluation des implications et du potentiel de maîtrise vont probablement générer une importante synchronisation des sous-systèmes de l'organisme, car la survie de l'individu est engagée. Ce haut degré de synchronisation aura pour objectif de motiver l'individu à conserver son attention sur l'accomplissement de ce besoin et à lui fournir le soutien nécessaire pour l'effort important qu'il sera probablement amené à réaliser (e.g., marcher 10 km dans le désert). Dans la mesure où cet épisode remplit toutes les conditions énoncées préalablement, il pourra être considéré comme émotionnel.

### **c. La différenciation des émotions**

Selon les théories de l'appraisal, le sentiment subjectif, soit la partie consciente du processus d'évaluation cognitive et de l'activation synchronisée des sous-systèmes de l'organisme impliqués dans les épisodes d'interaction (voir Figure 2) représenterait la partie consciente de l'émotion (Grandjean et al., 2008). Si le processus d'évaluation cognitive est causal dans le déclenchement des activations synchronisées des sous-systèmes de l'organisme, cela signifie que les différents sentiments subjectifs pourraient être différenciés sur la base des patterns du processus d'évaluation. Aussi, le processus d'évaluation cognitive serait certes causal dans le déclenchement des émotions, mais il permettrait aussi de les différencier (Lazarus, 1991 ; Moors, 2013 ; Moors et al., 2013 ; Smith & Kirby, 2009). Par exemple, un événement évalué comme obstructif pour un but avec un faible potentiel de maîtrise de ses conséquences générerait une émotion de peur. Au contraire, si le potentiel de maîtrise lié à cet événement était évalué comme élevé, il générerait de la colère (Scherer et al., 2013, 2019 ; Scherer & Ellgring, 2007). En somme, il y aurait une relation nécessaire et suffisante entre les patterns de critères d'évaluation et la réponse émotionnelle (Siemer et al., 2007) : d'une part, un pattern spécifique d'évaluations de critères générerait une émotion donnée, d'autre part, une émotion donnée reflèterait un pattern spécifique d'évaluations de critères.

### 3. L'attention émotionnelle

Depuis les travaux de James (1890), l'attention peut être définie comme « la prise de possession par l'esprit, sous une forme claire et vive, d'un objet ou d'une suite de pensées parmi plusieurs présents simultanément [...]. Elle implique le retrait de certains objets afin d'en traiter plus efficacement d'autres ». Du point de vue de l'attention visuelle, trois composantes seraient engagées dans ce processus (Posner, 1980) : le déplacement, permettant à l'attention de s'orienter de façon appropriée dans le champ visuel ; l'engagement, permettant de fixer l'attention sur un emplacement précis du champ visuel afin d'en analyser le contenu ; et le désengagement, permettant de relâcher l'attention après analyse afin de pouvoir la déplacer pour la réengager ailleurs.

Dans la mesure où les ressources attentionnelles sont limitées (Kahneman, 1973), le cerveau mettrait en place des stratégies pour traiter en priorité et automatiquement certains stimuli. Ce phénomène d'orientation d'attention non-volontaire opérerait selon deux processus : 1) l'attention exogène, orientée vers les propriétés perceptives de bas niveau des stimuli suivant des processus *bottom-up* ou *stimulus-driven* (Brosch et al., 2011), il s'agirait par exemple de l'orientation de l'attention vers un stimulus saillant dans le champ visuel ; 2) l'attention émotionnelle (Vuilleumier, 2005), orientée vers des stimuli dits émotionnels, afin de répondre au bien-être et à la survie d'un individu en détectant par exemple la présence d'un stimulus menaçant dans son champ visuel (Yiend, 2010). Néanmoins, ces processus ne sont pas indépendants et peuvent opérer ensemble afin d'optimiser le traitement des différents stimuli potentiellement pertinents (Brosch et al., 2011; Fuller & Carrasco, 2006).

Si le concept d'attention émotionnelle ne soulève plus de débat dans la littérature, ni d'ailleurs sa base cérébrale principale, l'amygdale (Sander et al., 2003), plusieurs théories modernes de l'émotion ont été proposées pour en expliquer le mécanisme (Anderson, 2005; Brosch et al., 2008; Öhman, Flykt, et al., 2001). En effet, trois approches majeures ont été

proposées : a) la théorie du module de la peur, dérivée de la théorie des émotions de base (Ekman, 1992), dans laquelle l'amygdale serait définie comme « module de la peur » (Mineka & Öhman, 2002; Öhman & Mineka, 2001) et l'attention émotionnelle serait sensible aux stimuli menaçants (Öhman, Flykt, et al., 2001) ; b) la théorie de l'arousal, dérivée de la théorie bidimensionnelle (Russell, 1980), dans laquelle l'amygdale et l'attention émotionnelle seraient associées à l'éveil généré par les stimuli (Anderson, 2005; Anderson et al., 2003) ; et enfin, c) les théories de l'appraisal, selon lesquelles l'amygdale et l'attention émotionnelle seraient sensibles à la pertinence des stimuli (Brosch et al., 2008; Sander et al., 2003), comme décrit dans la section précédente.

### **a. La théorie du module de la peur**

La notion de module de la peur s'apparente aux théories des émotions de base (Ekman, 1992). Dans ce cadre théorique, les émotions découleraient d'un héritage phylogénétique visant à assurer le bien-être et la survie des individus (Öhman, 1999), dans lequel chacune d'elles serait associée à un réseau neuronal spécifique (LeDoux, 2003). En l'occurrence, le module régissant le réseau neuronal de la peur, implémenté au niveau de l'amygdale, comprendrait quatre caractéristiques (Mineka & Öhman, 2002; Öhman & Mineka, 2001) : 1) ce module serait préférentiellement activé en contexte aversif par des stimuli dont l'information de menace aurait été acquise de manière phylogénétique (e.g., des araignées ou des serpents), afin de déclencher les séquences comportementales adaptées ; 2) ce module serait gouverné par le principes d'automatisme ; 3) ce module serait encapsulé (Fodor, 1983), ce qui signifie que son activation serait automatique suite à la perception d'une menace et cette activation serait hermétique à tout contrôle cognitif ; et 4) l'amygdale, serait le chef d'orchestre de ce réseau neuronal spécialisé dans la peur, faisant le lien entre la perception des menaces et le déclenchement des séquences comportementales adaptées (LeDoux, 2003).

Selon cette théorie, la priorisation des stimuli par l'attention émotionnelle reposerait sur le fonctionnement du module de la peur. La capture attentionnelle automatique déclenchée par la nature émotionnelle de l'information serait réservée aux stimuli phylogénétiquement menaçants. Cette hypothèse a été étayée par des résultats montrant que des serpents et des araignées captaient plus fortement l'attention que des fleurs ou des champignons (Öhman, Flykt, et al., 2001). Il a également été montré que des visages menaçants (i.e., exprimant la colère) captaient plus fortement l'attention que des visages neutres ou exprimant d'autres émotions négatives (e.g., la tristesse ; Öhman, Lundqvist, et al., 2001).

Cependant, la théorie du module de la peur et les hypothèses qui en découlent ont souffert de plusieurs critiques importantes. Par exemple, il a été montré que la capture attentionnelle vers les stimuli menaçants n'était pas réservée aux menaces phylogénétiques, mais pouvait apparaître en direction de stimuli de menace ontogénétique (e.g., des armes à feu ; Brosch & Sharma, 2005). De plus, les stimuli positifs peuvent eux aussi capter l'attention de manière automatique (voir Pool et al., 2016), ce qui remet en cause les fondements de la théorie du modèle de la peur à l'origine de l'attention émotionnelle.

## **b. La théorie de l'arousal**

Contrairement à la théorie du module de la peur, la théorie de l'arousal est à même de rendre compte de la capture attentionnelle par les stimuli positifs et les stimuli ontogénétiquement menaçants. Cette théorie, directement issue de la théorie bidimensionnelle de Russel (1980), propose que l'amygdale, impliquée dans la capture attentionnelle est sensible à l'arousal des stimuli, à savoir leur potentiel d'éveil, quelle que soit leur valence (i.e., positive ou négative). Pour tester cette hypothèse, des auteurs ont mené une étude en IRMf proposant à des participants de sentir des odeurs qui différaient en termes de valence et d'arousal (Anderson et al., 2003). L'utilisation de l'olfaction pour cette expérience a été motivée par la relative facilité de contrôler les stimuli dans cette modalité sensorielle. Pour ce faire, les auteurs ont réalisé

deux préparations : l'une à base de citral (i.e., une odeur de citron), jugée comme positive, l'autre à base d'acide valérique (i.e., une odeur de transpiration), jugée comme négative. Ces préparations ont été séparées en deux versions d'intensité distinctes, l'une faible, l'autre forte. Aussi, quatre préparations ont résulté de ces manipulations : deux positives et deux négatives, l'une de faible arousal, l'autre de fort arousal. À la suite de la présentation de ces préparations aux participants dans l'IRM, les résultats ont montré que l'activation de l'amygdale était sensible au niveau d'arousal des odeurs, indépendamment de leur valence. En d'autres termes, une activité amygdalienne plus forte a été observée lorsque les préparations généraient un fort arousal par rapport à celles générant un faible arousal, et ce, qu'elles soient positives ou négatives. Le fait que l'attention émotionnelle est sensible au niveau d'arousal des stimuli, quel que soit leur niveau de valence, a été confirmé par Anderson (2005) dans une étude mesurant le niveau de capture attentionnelle allouée à des mots selon leur niveau de valence et d'arousal.

Cela étant, malgré l'existence d'une corrélation entre niveau d'arousal des stimuli et capture attentionnelle, la relation causale proposée par la théorie de l'arousal manque de précisions. De fait, l'arousal des stimuli fait référence au niveau d'activation corporelle que leur perception déclenche (Schimmack, 2005). Du point de vue de la temporalité, il a été montré que la capture attentionnelle par les stimuli précéderait l'activation corporelle qu'ils déclenchent (Hackley & Graham, 1983). Cela amène à penser qu'il est plus probable que l'arousal des stimuli soit un marqueur latent du déclenchement de la capture attentionnelle plutôt que sa cause (Bradley et al., 2001). Schimmack (2005) a proposé que l'arousal reflèterait le niveau d'intensité du traitement alloué aux stimuli qui serait fonction de leur pertinence.

### **c. Les théories de l'appraisal et pertinence**

Comme évoqué précédemment, les théories de l'appraisal, et plus particulièrement le MPC, postulent que l'attention serait capturée en fonction de l'évaluation de la pertinence des stimuli traités par les individus (i.e., selon la nouveauté, la pertinence intrinsèque et la pertinence aux

buts ; Grandjean & Scherer, 2014; Scherer, 2009). De plus, le rôle de l'amygdale serait défini comme un détecteur de pertinence (Sander et al., 2003). La force de cette hypothèse est qu'elle permet d'intégrer à la fois certains éléments de la théorie du modèle de la peur et de la théorie de l'arousal. D'une part, les stimuli phylogénétiquement menaçants captureraient l'attention (LoBue & DeLoache, 2008; Öhman, Flykt, et al., 2001; Öhman, Lundqvist, et al., 2001) car leur pertinence intrinsèque serait évaluée comme déplaisante, du fait de l'information de menace héritée au fil des générations. D'autre part, les stimuli associés à un fort arousal, qu'ils soient ontogénétiquement menaçants (Brosch & Sharma, 2005; Fox et al., 2007) ou plaisants (Pool et al., 2016), captureraient l'attention car leur pertinence intrinsèque serait évaluée comme déplaisante et plaisante, respectivement (informations acquises au cours des expériences). Le MPC propose également d'intégrer le concept d'attention exogène en son sein, selon l'évaluation du sous-critère de nouveauté. Ce sous-critère, permettrait de rendre compte de la capture attentionnelle selon les propriétés perceptives de bas niveau des stimuli (Brosch et al., 2011), et selon l'évaluation de leur soudaineté, familiarité, ou prédictibilité.

Cette dernière décennie, la littérature a consacré un intérêt tout particulier à étudier la capture attentionnelle qui pouvait être déclenchée par l'évaluation de la pertinence aux buts. L'effet de la pertinence due à l'obstruction d'un but a été majoritairement testé dans le cadre de punitions dans un contexte de jeu. Dans ce contexte, une capture attentionnelle a été mise en évidence vers des stimuli associés à une perte qui pouvait être évitée en pressant une touche (Müller et al., 2016; Sun et al., 2017). Dans ce cas de figure, on peut considérer que le potentiel de maîtrise face à la perte est probablement évalué comme maîtrisable, car une confrontation est possible pour éviter cette perte en pressant la touche. Dans une autre étude, des auteurs se sont intéressés à la capture attentionnelle générée par une obstruction du but dont le potentiel de maîtrise était évalué comme incontrôlable (Müller & Wentura, 2023). Ici, la perte était déclenchée automatiquement à l'apparition du stimulus, obligeant les participants à se résigner. Les



résultats n'ont pas mis en évidence de capture attentionnelle vers les stimuli associés à une pertinence obstruant un but sans potentiel de maîtrise possible.

L'impact de la pertinence aux buts sur la capture attentionnelle en cas de facilitation a été testé de manière plus exhaustive. Par exemple, il a été mis en évidence que des stimuli pouvant faciliter l'accomplissement d'un but ou d'un besoin capturaient l'attention dans le cas de bouteilles d'eau si les participants étaient induits à un état de soif (Mazzietti et al., 2014). Un même résultat a été observé avec des stimuli associés à la propreté (e.g., du savon) lorsque les participants avaient été induits au dégoût (Vogt et al., 2011), ou encore avec des stimuli associés à un gain dans un contexte de jeu (Vogt et al., 2013) ou des stimuli procurant un bien-être (i.e., évitant le déclenchement d'un bruit désagréable ; Vogt et al., 2017).

## **i. La relation entre la pertinence intrinsèque et la pertinence aux buts**

### **1. Leurs effets combinés**

Si l'impact de la pertinence intrinsèque et de la pertinence aux buts sur l'attention a majoritairement été questionné de manière indépendante, la réalité est souvent bien plus complexe. En effet, dans la plupart des cas, les stimuli sont probablement associés à la fois à une pertinence intrinsèque et à une (ou plusieurs) pertinence(s) aux buts. Comme mentionné précédemment, un gâteau au chocolat pourra par exemple être évalué à la fois comme plaisant et pouvant conduire à satisfaire la faim. Il est important de ne pas confondre le plaisir qu'un stimulus pourra procurer à un individu et sa pertinence intrinsèque. Dans le cas où le gâteau au chocolat est évalué comme pertinent pour une recherche de plaisir (e.g., une envie de sucre), cette pertinence doit être comprise comme une pertinence aux buts. Tandis que la pertinence intrinsèque se cantonne seulement à la préférence générale des individus concernant le gâteau (Scherer, 2005b), qu'elle soit héritée par l'évolution ou acquise par leurs expériences passées. Cette distinction est particulièrement importante pour préciser le caractère rigide de la pertinence intrinsèque dans le temps. Par exemple, en cas de faim, un aliment peu appétissant

pourra être considéré comme moins dégoûtant et pourra même déclencher un léger plaisir, mais cet aliment ne deviendra pas pour autant intrinsèquement plaisant (Hoefling et al., 2009).

La relation qui lie la pertinence intrinsèque et la pertinence aux buts pourrait être illustrée par la distinction « learning – liking – wanting » issue du modèle animal (Berridge, 2009; Berridge et al., 2009; Castro & Berridge, 2014). Dans ce modèle, le « learning » fait référence à l'acquisition d'informations associées aux stimuli à la suite des interactions de l'individu avec eux. L'idée est que si un stimulus facilite l'accomplissement d'un but ou d'un besoin, une libération de dopamine sera déclenchée, signifiant au cerveau que la manière dont le stimulus a conduit à cette libération dopaminergique mérite d'être mémorisée. Deux informations seront donc renforcées en mémoire : 1) le stimulus est probablement plaisant car il a permis de faciliter l'accomplissement d'un but (i.e., renforçant la préférence – ou la pertinence intrinsèque – du stimulus) et ; 2) ce stimulus peut permettre de faciliter l'accomplissement de ce type de but (i.e., information qui pourra être utilisée dans l'évaluation du sous-critère de facilitation aux buts et aux besoins de ce type de stimulus). Le « wanting », quant à lui, fait référence à la motivation de la recherche d'accomplissement de buts ou de besoins. Le « wanting » pourrait être guidé par la pertinence aux buts des stimuli, issue directement du processus de « learning », permettant d'informer les individus que des stimuli peuvent potentiellement permettre d'accomplir des buts ou besoins spécifiques. Enfin, le « liking », fait référence à l'état de plaisir ressenti à la suite de l'accomplissement d'un but, permettant de renforcer le « liking » des stimuli.

À notre connaissance, aucune étude ne s'est intéressée à la question de la manière dont les effets combinés de la pertinence intrinsèque et de la pertinence aux buts pourraient impacter la capture attentionnelle. Cependant, deux études se sont intéressées à leurs effets combinés sur différentes réponses somato-viscérales (Aue & Scherer, 2008, 2011). Dans ces études, les mesures ont porté sur l'activité musculaire acquise par électromyographie (i.e., zygomatique

majeur et mineur, corrugator supercilii et muscles extenseurs des doigts), le rythme cardiaque, la conductance électrodermale et la température cutanée. Ces réponses somato-viscérales ont été mesurées selon différents schémas d'évaluation de pertinence intrinsèque et de pertinence aux buts des stimuli. Pour ce faire, des images intrinsèquement plaisantes et déplaisantes ont été présentées sur un écran. Une seconde après l'apparition d'une image, un symbole venait se superposer au centre de l'image. Selon le symbole apparu, il était demandé aux participants de pousser un joystick vers l'avant ou de le tirer vers l'arrière. Si le joystick était poussé vers l'avant, l'image grandissait, tandis qu'en le tirant, l'image rétrécissait. Les auteurs ont proposé que si l'image plaisante rétrécissait ou que l'image déplaisante grandissait, ces effets seraient considérés comme générant une pertinence aux buts obstructrice. Au contraire, si l'image plaisante grandissait ou si l'image déplaisante rétrécissait, ces effets seraient considérés comme générant une pertinence aux buts facilitatrice. Les résultats ont montré que si les réponses somato-viscérales déclenchées par les différents types de pertinence intrinsèque et de pertinence aux buts étaient similaires, elles n'étaient pas identiques. Ces résultats illustrent l'importance de considérer la pertinence intrinsèque et la pertinence aux buts de manière séparée, ce qui est rarement effectué dans la littérature sur les émotions. Concernant les effets combinés de la pertinence intrinsèque et de la pertinence aux buts, la première étude (Aue & Scherer, 2008) n'a pas permis de conclure sur le type de relation que les caractérise. Cependant, dans la seconde étude (Aue & Scherer, 2011), des effets multiplicatifs de la pertinence intrinsèque et de la pertinence aux buts ont été observés sur toutes les réponses somato-viscérales à l'exception d'une seule, l'activité au niveau du site du corrugator supercilii.

## 2. Leur temporalité

Bien que cela n'ait pas été articulé dans les termes des théories de l'appraisal, plusieurs travaux laissent à penser que la pertinence intrinsèque et la pertinence aux buts auraient un impact différent sur la capture attentionnelle en fonction du temps. Plus précisément, la capture

attentionnelle précoce (i.e., à courte temporalité) serait davantage impactée par la pertinence intrinsèque, tandis que la capture attentionnelle tardive (i.e., à plus longue temporalité) serait davantage impactée par la pertinence aux buts. Par exemple, à l'aide d'un eye-tracker, des orientations initiales du regard plus fréquentes vers des aliments riches en graisses ont été observées pour un groupe de participants souffrant de troubles alimentaires (i.e., qui consomment habituellement ce type d'aliments) par rapport à un groupe contrôle (Castellanos et al., 2009; Nijs et al., 2010; Werthmann et al., 2011). Ces résultats ont été observés alors que le niveau d'envie et de satiété était contrôlé. De plus, il a été mis en évidence que le maintien du regard dans le temps vers les aliments était corrélé positivement avec l'état de manque. Des résultats similaires ont été obtenus pour les troubles addictifs tels que la dépendance à la nicotine et à l'alcool (Field & Cox, 2008). En outre, Forrest et al. (2022) ont montré que le niveau de capture attentionnelle vers des stimuli associés à une pertinence facilitatrice d'un but était plus important à une temporalité tardive (250 ms) qu'à une temporalité précoce (30 ms). Ces résultats sont en accord avec les théories de l'appraisal qui supposent que la pertinence intrinsèque est une évaluation de plus bas niveau que celle de la pertinence aux buts (Scherer, 2013). En effet, l'évaluation de la pertinence intrinsèque serait plus rapide, car elle ne nécessiterait de traiter que peu d'informations, celles correspondant au caractère d'agrément général des stimuli (Aue et al., 2007; Grandjean & Scherer, 2008; Lanctôt & Hess, 2007). Au contraire, l'évaluation de la pertinence aux buts serait établie sur la base des autres critères du processus d'évaluation (i.e., l'implication, le potentiel de maîtrise et la signification normative), générant une quantité d'informations plus importante et donnant lieu à un temps de traitement plus long.

## **4. Attention émotionnelle et trouble affectif : le cas de l'anxiété**

Dans la mesure où la problématique de l'attention émotionnelle dans le cas de l'anxiété est abordée dans l'une des études menée dans le cadre de cette thèse, il convient d'introduire ici cette thématique.

### **a. Capture attentionnelle vers la menace et troubles anxieux**

Si l'attention émotionnelle a initialement une fonction adaptative, à savoir maintenir les ressources attentionnelles des individus en direction des stimuli évalués comme pertinents pour leur bien-être et leur survie, elle peut devenir inadaptée. C'est le cas pour les individus souffrant de troubles affectifs tels que l'anxiété (Bar-Haim et al., 2007; Waters et al., 2010). L'anxiété peut émerger sous la forme d'un épisode émotionnel, si le processus d'évaluation d'un événement conduit à le qualifier de menaçant pour le bien-être et la survie des individus (Beck et al., 2005). L'anxiété peut également devenir pathologique, lorsque les épisodes anxieux sont expérimentés de manière extrêmement fréquente, ce qui conduit à biaiser le processus d'évaluation des événements par les individus (Milne et al., 2019). Plus précisément, les troubles anxieux seraient caractérisés par une surestimation de l'impact négatif ou de la sévérité des événements dans le processus évaluatif. Cette surestimation entraînerait une capture attentionnelle plus forte en direction de la menace pour les individus anxieux (Bar-Haim et al., 2007; Waters et al., 2010). Par effet boule de neige, le fait d'être focalisé sur les menaces dans l'environnement augmenterait la surestimation de l'impact négatif ou de la sévérité des événements dans le processus évaluatif, ce qui augmenterait en retour le niveau de capture attentionnelle en direction de la menace.

### **b. Une littérature en crise**

La plupart des études qui ont évalué la relation entre la capture attentionnelle vers la menace selon le niveau d'anxiété ont utilisé une tâche largement utilisée dans la littérature portant sur l'attention émotionnelle, à savoir, la tâche de dot-probe (MacLeod et al., 1986). Cette tâche se

défini par l'apparition consécutive de trois écrans : une croix de fixation, l'apparition simultanée de deux stimuli (un pertinent et un neutre) de part et d'autre de la croix de fixation, puis l'apparition d'une cible remplaçant l'un des deux stimuli. Les participants ont pour instruction de maintenir leur regard sur la croix de fixation pendant toute la durée de la tâche et d'indiquer le plus rapidement et le plus précisément possible le côté d'apparition de la cible en appuyant sur une touche. Les essais sont définis comme valides si la cible apparaît du côté du stimulus pertinent et comme non-valides si le point apparaît du côté du stimulus neutre. L'hypothèse qui sous-tend cette tâche est que lorsque les deux stimuli apparaissent, l'attention sera captée par le stimulus le plus pertinent. Si la cible apparaît du côté du stimulus pertinent (essai valide), le temps de réponse devrait être plus court car l'attention est déjà orientée de ce côté. En revanche, si la cible apparaît du côté du stimulus neutre (essai non-valide), le temps de réponse devrait être allongé, car l'attention doit être désengagée du côté du stimulus pertinent, puis déplacée et engagée du côté d'apparition du stimulus neutre, où la cible est apparue. Des Indices de Biais Attentionnel (IBA) sont généralement calculés pour interpréter les résultats de la tâche de dot-probe. Ils sont calculés en soustrayant les temps de réponse moyens des essais valides de la moyenne des essais non-valides (Macleod et al., 1986). Ainsi, un IBA positif signifie que les stimuli pertinents auront capté davantage l'attention que les stimuli neutres. Bien que de nombreuses études aient mis en évidence un biais attentionnel en faveur des stimuli liés à la menace chez les personnes souffrant de troubles anxieux (Bar-Haim et al., 2007; Waters et al., 2010), ces résultats n'ont pas été observés de manière systématique dans la littérature (Kruijt et al., 2019; McNally, 2019). En effet, une grande fluctuation dans l'IBA a été rapportée, à la fois en termes de mise en évidence de la capture attentionnelle vers la menace chez les individus anxieux, et dans la consistance des tests-retests chez les individus.

### c. Les pistes d'amélioration

En réponse à ce manque de consistance dans les résultats, des auteurs ont émis l'hypothèse que la variabilité des temps de réponse recueillis dans la tâche de dot-probe (donnant lieu à des IBA fluctuants) pouvait s'expliquer théoriquement. En effet, une partie de la littérature suggère que l'anxiété est, certes, caractérisée par un biais attentionnel *vers* la menace, mais aussi à *l'écart* de la menace (Zvielli et al., 2015). Selon ces auteurs, cet évitement de la menace chez les individus anxieux représenterait une forme de régulation de l'état d'anxiété en tentant de l'inhiber. Pour tenir compte de ce phénomène hypothétique, de nouvelles méthodes de calcul d'IBA de la tâche de dot-probe ont été proposées.

Premièrement, il a été proposé de calculer un indice de Variabilité du Biais Attentionnel (VBA), et ce, de deux manières différentes. Le calcul proposé par (Iacoviello et al., 2014) consiste à séparer les temps de réponse de la tâche de dot-probe en « bacs » de 20 essais successifs. Puis, l'ABI (i.e., la différence entre la moyenne des temps de réponse des essais valides soustrait à celle des essais non-valides) est calculé pour chaque bac. Enfin, l'écart-type entre les différents IBA est calculée, puis divisée par le temps de réponse moyen des participants. Le calcul proposé par Naim et al. (2015), est similaire, à la différence que les essais sont séparés par blocs comprenant 10 essais successifs valides et 10 essais successifs non-valides. L'idée d'utiliser un tel indice consiste à manipuler un score normalisé représentant la variabilité du biais attentionnel pour chaque participant plutôt que son ampleur.

Deuxièmement, le « Trial Level-Bias Score » (TL-BS) a été proposé par Zvielli et al. (2015). Le calcul de ce score consiste dans un premier temps à soustraire chaque temps de réponse de chaque essai non-valide par le temps de réponse de l'essai valide le plus proche temporellement. Ensuite, cinq indices TL-BS sont utilisés : 1) la moyenne et 2) le pic des TL-BS dont la valeur est négative (reflétant la moyenne et la valeur la plus élevée de la capture attentionnelle à l'écart de la menace) ; 3) La moyenne et 4) le pic des TL-BS dont la valeur est positive (reflétant la

moyenne et la valeur la plus élevée de la capture attentionnelle vers la menace) ; et 5) la variabilité du TL-BS. Ce dernier indice est calculé en additionnant toutes les distances entre tous les TL-BS successifs dans le temps, puis en divisant cette somme par le nombre total de TL-BS. Ces calculs permettraient de rendre compte à la fois du biais attentionnel en direction et à l'écart de la menace. Des études ont comparé ces différents indices au sein de tâches de dot-probe utilisées dans le cadre de la thématique de la capture attentionnelle vers la menace dans l'anxiété (Molloy & Anderson, 2020; Price et al., 2015). Selon les auteurs, ces indices permettraient d'augmenter – modestement – la fiabilité des résultats obtenus dans les tâches de dot-probe mesurant le lien entre capture attentionnelle vers la menace et niveau d'anxiété.

## **5. La théorie de l'inférence active**

Dans la présente thèse, un travail de réflexion articulant les théories de l'appraisal et la théorie de l'inférence active a été mené. Aussi, cette section est dédiée à une présentation succincte des postulats de la théorie de l'inférence active.

Le cœur de la théorie de l'inférence active consiste à définir le fonctionnement du cerveau en termes d'échanges de prédictions et d'erreurs de prédiction (Bottemanne et al., 2022; Hohwy, 2013). Dans ce contexte, les individus sont considérés comme des systèmes autoorganisés dont les interactions avec l'environnement sont décrites par un processus de prédiction hiérarchique répondant au but de maintenir leur homéostasie - ou de minimiser leur énergie libre (Kirchhoff et al., 2018; Seth, 2013). L'engouement récent pour ce domaine de recherche vise principalement à repenser les modèles existants du comportement, du fonctionnement du cerveau et d'esprit dans cette perspective. À terme, ce travail pourrait conduire à développer un cadre universel pour l'étude des êtres humains, unifiant les domaines de la psychologie, des neurosciences, de la philosophie et des mathématiques.



### **a. L'hypothèse du codage prédictif**

La théorie de l'inférence active est basée sur l'hypothèse du codage prédictif inscrite dans le paradigme du cerveau prédictif (Bottemanne et al., 2022; Hohwy, 2013). Selon cette hypothèse, le cerveau fonctionnerait comme une machine d'inférence construisant un modèle génératif du monde. Ce modèle viserait à prédire l'état actuel et futur des entrées sensorielles, ainsi que les différents niveaux de causes hiérarchiques qui pourraient expliquer ces états. À titre d'exemple, sur la base de l'état des entrées sensorielles d'une scène de perception visuelle, les causes hiérarchiques se référeraient aux caractéristiques locales de l'image (e.g., la luminance, la couleur), à la forme des objets (e.g., un cercle), et à la classe de l'objet (e.g., un ballon ; Kersten et al., 2004). À cette fin, le cerveau serait implémenté par deux modèles. D'une part, le modèle « forward », constitué de connexions top-down, coderait les prédictions concernant les différents niveaux de causes hiérarchiques des états des entrées sensorielles. Ces prédictions seraient réalisées par inférence Bayésienne (Colombo & Seriès, 2012; Durante & Dunson, 2018; Lin & Garrido, 2022), en codant les distributions de probabilité des différentes hypothèses représentant les croyances plausibles au sujet des causes hiérarchiques (Seth & Friston, 2016). D'autre part, le modèle « backward », constitué de connexions bottom-up, coderait les erreurs de prédiction des causes des états des entrées sensorielles. L'erreur de prédiction doit être comprise comme la partie de l'inférence d'un niveau de cause qui n'explique pas correctement un niveau d'état des entrées sensorielles. Par exemple, prédire la classe d'un objet comme étant un ballon est considéré comme une erreur de prédiction si sa forme est inférée comme étant un carré. Une telle conceptualisation du fonctionnement du cerveau implique de renverser la vision traditionnelle concernant la définition des messages top-down et bottom-up (Friston, 2018). En effet, plutôt que de définir les messages top-down comme le contrôle exécutif et les messages bottom-up comme la perception de l'environnement (Eysenck et al., 2007; Mogg & Bradley, 2016, 2018; Yantis, 1998), les messages top-down refléteraient

la transmission de prédictions, tandis que les messages bottom-up refléteraient la transmission d'erreurs de prédiction (Friston, 2008; Grossberg, 2009; Hesselmann et al., 2010; Rauss & Pourtois, 2013).

L'hypothèse de l'implémentation du cerveau selon l'hypothèse du codage prédictif implique deux postulats (Friston, 2005, 2018). Premièrement, les différents niveaux des causes hiérarchiques seraient attribués à des couches cérébrales spécifiques (e.g., la forme de l'objet, la classe d'objet). Au sein de chaque couche, un processus serait effectué, visant à inférer les hypothèses plausibles concernant les causes qui leur sont attribuées. Deuxièmement, les niveaux de causes hiérarchiquement supérieurs viseraient à prédire les causes des niveaux inférieurs (par le biais du modèle forward). En retour, les erreurs de prédiction concernant les niveaux de causes hiérarchiquement inférieurs seraient remontées vers les causes de niveaux supérieurs afin de corriger la prédiction (par le biais du modèle backward). Ainsi, la croyance concernant un niveau de cause donné – ou l'état de la couche qui lui est attribuée – résulterait d'un processus de sélection d'hypothèses basé sur la vraisemblance des hypothèses qui ont été déduites comme plausibles au sein de la couche. La vraisemblance des hypothèses serait inférée sur la base d'un calcul tenant compte des prédictions transmises par le modèle forward et des erreurs de prédictions transmises par le modèle backward au sein de la couche.

### **b. Le principe d'énergie libre**

Il a été proposé d'appliquer le principe de l'énergie libre pour rendre compte du processus de sélection de l'hypothèse la plus vraisemblable concernant les différents niveaux de causes hiérarchiques au sein des couches (Friston, 2005, 2010, 2012a).

Dans un premier temps, l'application de ce principe suppose que le cerveau manipulerait une quantité statistique d'erreurs de prédiction appelée « énergie libre » (Friston, 2012b; Smith et al., 2019). Dans un second temps, cela invite à considérer que deux calculs de quantité d'énergie libre seraient effectués au sujet de chaque niveau de cause au sein de chaque couche : 1) une

quantité d'énergie libre résultant de la différence entre chaque hypothèse plausible concernant la cause en question et la prédiction à son sujet envoyée par le modèle forward (cette quantité d'énergie libre refléterait la quantité d'erreurs de prédiction hypothétique que le modèle backward enverra à la couche attribuée au niveau de cause supérieur si cette hypothèse est sélectionnée); 2) et une quantité d'énergie libre résultant de la différence entre chaque hypothèse plausible concernant la cause donnée et l'erreur de prédiction envoyée à son sujet par le modèle backward (cette quantité d'énergie libre refléterait la quantité hypothétique d'erreurs de prédiction que le modèle backward sera amené à renvoyer à ce niveau de cause si la prédiction envoyée à la cause de niveau inférieure est réalisée sur la base d'une hypothèse plausible donnée).

Une autre hypothèse de l'application du principe de l'énergie libre au processus de sélection des hypothèses invite à considérer les couches du cerveau comme des systèmes autoorganisés (Kirchhoff et al., 2018). Plus précisément, cela suppose que ce processus au sein des couches serait régi par une règle interne, à savoir le maintien de leur homéostasie. Dans ce contexte, l'homéostasie des couches du cerveau fait référence au maintien de leur intégrité fonctionnelle visant à résister au désordre. En d'autres termes, maintenir l'homéostasie signifie minimiser l'erreur de prédiction ou plus précisément l'énergie libre au sein des couches (Friston, 2010). Ainsi, parmi les hypothèses plausibles concernant les différents niveaux de causes hiérarchiques, le processus de sélection s'effectuerait sur la base de celles qui minimisent le plus la quantité d'énergie libre. Il est important de noter que le calcul de la quantité d'énergie libre serait pondéré par le degré de confiance accordé aux prédictions (Smith et al., 2021). En d'autres termes, si une hypothèse plausible minimise l'énergie libre dans le calcul relatif à une prédiction associée à un haut degré de confiance, elle pourrait être sélectionnée quelle que soit la quantité d'énergie libre issue du calcul relatif à l'erreur de prédiction.

En résumé, l'hypothèse du codage prédictif régi par le principe de l'énergie libre invite à considérer le cerveau comme un système autoorganisé composé de sous-systèmes autoorganisés imbriqués (i.e., les couches du cerveau). Au niveau de l'ensemble du cerveau, le maintien de l'homéostasie – ou la minimisation de l'énergie libre – des différents niveaux de couches hiérarchiques concernant les entrées sensorielles conduit à maximiser la précision du modèle génératif du monde (K. Friston, 2012b; K. J. Friston et al., 2010; Kirchhoff et al., 2018).

### **c. L'inférence active**

La théorie de l'inférence active vise à appliquer l'hypothèse du codage prédictif et du principe d'énergie libre à un niveau d'abstraction plus large, à savoir, l'ensemble des interactions des individus avec leur environnement (Friston et al., 2017). Cette théorie s'inscrit dans une perspective incarnée dans laquelle les inférences – ou la cognition – sont imbriquées dans un corps qui est lui-même imbriqué dans un environnement (Friston et al., 2013). En d'autres termes, le modèle génératif du cerveau viserait à maintenir l'homéostasie des sous-systèmes du corps (i.e., leur intégrité structurelle et fonctionnelle) situés dans un environnement. À cette fin, le modèle mettrait également en œuvre un processus d'inférence intéroceptive visant à prédire les états intéroceptifs actuels et futurs ainsi que leurs différents niveaux de causes hiérarchiques (Barrett & Simmons, 2015; Seth, 2013, 2015; Seth & Friston, 2016). L'état intéroceptif fait référence à l'expérience subjective des états d'activation des sous-systèmes de l'organisme (Craig, 2002, 2003). Par exemple, il s'agirait de prédire qu'un effort robuste précédemment fourni (un niveau de cause hiérarchique) peut expliquer un état de faim surprenant (i.e., un état intéroceptif inféré). Dans ce cas, l'état de faim surprenant serait généré par la mise à jour de l'état intéroceptif inféré à partir du message afférent (erreur de prédiction) de la baisse du taux de sucre dans le sang (Seth, 2015).

Dans ce contexte, les actions seraient intégrées dans le modèle génératif cérébral en tant que réflexes autonomes, inférés pour corriger les erreurs de prédiction dans les différents niveaux

de causes hiérarchiques qui conduisent à des états indésirables (Adams et al., 2013; Buckley et al., 2017; Friston, 2011; Pezzulo et al., 2015). Le terme « action » est utilisé ici au sens large, incluant les actions externes (i.e., les séquences comportementales) et les actions internes (i.e., la régulation de l'état des sous-systèmes de l'organisme). Dans l'exemple précédent, la consommation de nourriture (action externe) et/ou le métabolisme des réserves de graisse (action interne) pourraient être inférés comme permettant de corriger la chute du taux de sucre dans le sang (erreur de prédiction) qui a conduit à l'état de faim (état intéroceptif indésirable). Cependant, le modèle génératif peut également mettre à jour ses prédictions en définissant l'état de faim comme viable pendant un certain temps, en abaissant les seuils de viabilité des niveaux de sucre dans le sang. La décision concernant le schéma résultant parmi la mise à jour des prédictions et/ou le déclenchement d'une action interne/externe serait prise par le processus de sélection d'hypothèses concernant les différents niveaux de causes hiérarchiques décrit précédemment (Friston, 2005, 2010, 2012b). En bref, la théorie de l'inférence active suppose que la perception et la signification de l'environnement, ainsi que les actions qui en résultent, sont le résultat d'un processus complexe d'inférence régi par le principe de minimisation de l'énergie libre. Bien que la notion de minimisation de l'énergie libre se réfère à la minimisation de l'erreur de prédiction au niveau neuro-computationnel, elle permet à terme d'atteindre l'homéostasie de l'individu (i.e., au service de son bien-être et de sa survie).

## **6. Conclusion de l'introduction générale**

Après cette introduction censée décrire les principales problématiques abordées dans cette thèse, nous présenterons successivement 5 articles publiés, en révision, ou soumis pour publication, qui ont fait l'objet de notre travail de thèse. Les trois premiers articles se concentrent sur les effets combinés de la pertinence intrinsèque et de la pertinence aux buts sur la capture attentionnelle.

Le premier article vise à qualifier la nature de ces effets combinés dans le cas où la pertinence intrinsèque et la pertinence aux buts des stimuli sont évaluées comme plaisante et facilitatrice, respectivement.

Le second article s'intéresse à comparer l'impact attentionnel de stimuli associés à une pertinence aux buts facilitatrice selon la nature plaisante ou déplaisante de leur pertinence intrinsèque.

Le troisième article a pour objectif de tester l'impact de stimuli intrinsèquement plaisants et déplaisants sur l'attention, selon que ceux-ci étaient associés soit à une induction de pertinence aux buts facilitatrice, à une induction de pertinence aux buts obstructrice sans contrôle possible, ou en l'absence d'induction.

Le quatrième article, quant à lui, est axé sur la thématique de l'attention émotionnelle dans le cadre des troubles anxieux. Il poursuit trois objectifs : 1) tester l'hypothèse de l'impact attentionnel des effets combinés de la pertinence intrinsèque et de la pertinence aux buts dans le contexte d'une capture attentionnelle vers la menace exacerbée par le niveau d'anxiété ; 2) tester la consistance des différents calculs d'indice de biais attentionnel proposés dans la littérature ; 3) mettre en place et valider une tâche novatrice permettant de mesurer le niveau de capture attentionnelle vers les stimuli, dérivée de la tâche de dot-probe classique.

Enfin, le cinquième article, consiste en un essai théorique visant à intégrer les différentes hypothèses des théories de l'appraisal dans le cadre de la théorie de l'inférence active.

## **Chapitre 2 - Article : “Combined effects of intrinsic and goal relevances on attention and action tendency during the emotional episode”**

Ce chapitre correspond au premier article que nous avons réalisé, paru dans la revue *Emotion*, dont la référence est la suivante :

Fournier, H., & Koenig, O. (2023). Combined effects of intrinsic and goal relevances on attention and action tendency during the emotional episode. *Emotion*, 23(2), 425-436.

<https://doi.org/10.1037/emo0001081>

## Résumé

Des travaux récents sur l'attention émotionnelle suggèrent que l'attention est capturée en priorité par les stimuli qui sont intrinsèquement pertinents et par les événements qui sont propices à la réalisation d'objectifs importants à court terme. L'évaluation de la pertinence intrinsèque et de la pertinence aux buts vise à alerter les individus de la présence d'événements qui pourraient nécessiter une action susceptible d'affecter leur bien-être et fait partie du processus qui conduit à un épisode émotionnel. Cependant, à notre connaissance, aucune donnée empirique n'a encore mis en évidence la manière dont la pertinence intrinsèque et la pertinence du but se combinent précisément pour déclencher des biais attentionnels. L'objectif de cette étude était de tester l'hypothèse selon laquelle la pertinence intrinsèque et la pertinence du but ont des effets multiplicatifs sur l'attention et déclenchent une tendance à l'action. À cette fin, les participants ont évalué le niveau de pertinence intrinsèque, de pertinence du but et de tendance à l'action associé à quatre stimuli (deux liés à la faim et deux liés à la soif), avant de réaliser une tâche de dot-probe dans laquelle ces stimuli étaient en compétition. Nous avons observé que les évaluations de la pertinence intrinsèque et de la pertinence du but avaient effectivement des effets multiplicatifs sur l'attention. De plus, la tendance à l'action rapportée par les participants était mieux prédite lorsque la pertinence intrinsèque et la pertinence au but étaient prises en compte simultanément. Ces résultats montrent l'importance de prendre en compte à la fois la pertinence intrinsèque et la pertinence au but dans la recherche sur l'émotion et l'attention.



**Combined effects of intrinsic and goal relevances on attention and action tendency  
during the emotional episode**

Hippolyte Fournier and Olivier Koenig

Laboratoire EMC, Université Lumière Lyon 2, Bron, France

EMC, Lyon 2 University, Bron, France

**Author note**

Hippolyte Fournier  <https://orcid.org/0000-0002-7861-2387>

Olivier Koenig  <https://orcid.org/0000-0002-5124-1889>

We have no conflicts of interest to disclose. We warrantee that participants in this experiment were treated in accordance with the ethical standards of APA. This research was made possible thanks to a doctoral grant to Hippolyte Fournier from the Banque Publique d'Investissement (BPI) under the grant agreement THERADIA. Thanks to Elise Lefèvre and Oussama Abdoun for our discussions and advice on data analyses. Thanks also to Victoria Chazal Piquand, Camille Lépingle, Angélique Veiga, Sarah Vieira Da Silva for their help in collecting the data.

All data, analysis code, and material from the study reported in this article are available through the Open Science Framework repository: <https://osf.io/27vfb/>.

Correspondence should be addressed to Hippolyte Fournier, Laboratoire EMC, Université Lumière Lyon 2, 5 Avenue Pierre Mendès France, Bron 69676, France. Email: [h.fournier@univ-lyon2.fr](mailto:h.fournier@univ-lyon2.fr)

### **Abstract**

Recent work on emotional attention suggests that attention is prioritized toward stimuli that are intrinsically relevant and toward events that are conducive to momentaneous important goals. The appraisal of intrinsic relevance and goal relevance is aimed to alert individuals to the presence of events that may require action that could affect their well-being and is part of the process that leads to an emotional episode. However, to our knowledge, no empirical data has yet highlighted how precisely intrinsic and goal relevances combine with each other to trigger attentional biases. The aim of this study was to test the hypothesis that intrinsic and goal relevances have multiplicative effects on attention and trigger action tendency. To this end, participants appraised the level of intrinsic relevance, goal relevance and action tendency associated with four stimuli (two related to hunger and two to thirst), before performing a dot-probe task in which these stimuli were in competition. We observed that intrinsic and goal relevances appraisals had indeed multiplicative effects on attention. In addition, the action tendency reported by participants was best predicted when intrinsic and goal relevances were considered together. These results show the importance of considering both intrinsic and goal relevances in research on emotion and attention.

*Keywords:* attention, action tendency, emotion, goal relevance, intrinsic relevance

### **Combined effects of intrinsic and goal relevances on attention and action tendency during the emotional episode**

Over the past few decades, research on emotional attention has evolved considerably. Indeed, it was first proposed that emotionally triggered attentional biases were restricted to phylogenetically threatening stimuli (Öhman et al., 2001), then to stimuli associated with a strong arousal (Anderson, 2005; Bradley et al., 2001; Lang, 1995). Henceforth, the literature seems to agree in defining the impact of emotion on attention as intended to indicate which stimuli are most important and will likely require action for the survival and well-being of individuals (Brosch et al., 2010). Specifically, the facilitative or obstructive nature of stimuli for the conduciveness toward goals that are momentarily important to the individuals would trigger attentional biases. For instance, studies have shown that attention is captured by a stimulus that may satisfy a goal that has been previously induced (Vogt et al., 2013). Other studies have shown that attention is captured by bottles of water, only when individuals are in a state of thirst (Mazzietti et al., 2014). Vogt et al. (2017) showed that in a threat context, attention is captured by both threatening and safe stimuli, that is, both stimuli that must be avoided and stimuli that may be conducive to safety to compensate threat. However, according to the Component Process Model (CPM; Scherer, 2009), the likelihood that a stimulus may require action is not only determined by the conduciveness nature of stimuli toward momentarily important goals (e.g., being safe), but also by its pleasantness, defined as the intrinsic relevance (Scherer, 2001). Thus, appraisal theorists suggests that attentional bias could also be impacted by the intrinsic relevance (Montagrín et al., 2019).

The CPM defines emotion as the period triggered by the detection of a relevant event appraised by several cognitive criteria (Grandjean & Scherer, 2014). The appraisal of these criteria triggers activations of organism' sub-systems (e.g., autonomic physiology, action tendency, motor expression, subjective feeling) which, in return, influences future appraisals.

In this theoretical framework, the *intrinsic relevance* could be defined as a continuum from unpleasant (generally encouraging avoidance) to pleasant (generally encouraging approach), independently of the organism state. The value of a stimulus on this continuum could be described as a rather stable synthesis of past emotional episodes including the involvement of this type of stimulus. According to the CPM, *goal relevance* depends on two pieces of information: the *conduciveness of stimuli toward goals*, that can be represented as a continuum from obstructive (generally encouraging avoidance) to facilitative (generally encouraging approach) to goals achievement in response to the stimulus, and the *goals urgency*, that can also be represented as a continuum, from unimportant to extremely important, within the hierarchy of goals from the individual point of view. Clearly defining intrinsic relevance and goal relevance and their relationships are major concerns for appraisal theorists. Stimuli that conduce the achievement of currently important goals generate states of pleasure (involving dopamine releases, Berridge, 2009; Berridge et al., 2009; Castro & Berridge, 2014; Ferreri, et al., 2019), which enhances the intrinsic relevance of stimuli in memory for future interactions (Smith et al., 2011). We emphasize that it is crucial to differentiate intrinsic relevance from state of pleasure. For instance, in condition of hunger, unpalatable food may be less disgusting and may even trigger a small amount of pleasure, yet the stimuli will not become pleasant (Hoefling et al., 2009). This example supports our view of the independence between intrinsic relevance and goal relevance.

The precise contribution of intrinsic and goal relevances on attention orientation is yet to be measured empirically. First, the specific contribution of the level of conduciveness toward goals and the level of goals urgency in the constitution of goal relevance remains unclear. Second, it is difficult to obtain separate measures of intrinsic relevance and goal relevance of stimuli. Aue and Scherer (2008; 2011) investigated the impact of intrinsic and goal relevances on somatovisceral responses (i.e., cardiac activity, skin conductance, skin temperature). They

highlighted similar but not identical response patterns triggered by these two appraisal processes on somatovisceral responses. Furthermore, their results suggested that the presence of intrinsic relevance and goal relevance triggers multiplicative effects on somatovisceral responses. If these two relevances have multiplicative effects on somatovisceral responses, this suggests that the same may be true for triggering the activation of all other organism subsystems during emotional episodes. Thus, we hypothesized that attentional biases are certainly triggered by goal relevance as proposed in the literature (Vogt et al., 2017), but also by intrinsic relevance. Furthermore, we hypothesized that these two types of relevance have multiplicative effects on attentional biases. On the one hand, we compared this hypothesis with the one proposed by Vogt et al. (2017) that suggests that only goal relevance has effects on attentional capture. On the other hand, to test whether intrinsic and goal relevances have multiplicative effects on attentional biases, we confronted the hypothesis with another type of relationship, namely addition, as done by Aue and Scherer (2011). Determining the type of relationship between intrinsic relevance and goal relevance tells us a lot about how relevance works. Indeed, if this relationship is multiplicative in nature, it suggests that the attentional bias is sensitive to the alignment of intrinsic relevance and goal relevance, which is not the true in the case of an additive relationship. Consider an example where three stimuli have the same level of total relevance, when computed by adding their intrinsic and goal relevances: the first is carried by intrinsic relevance only, the second by goal relevance only, and the third by intrinsic and goal relevances at an equal level. Thus, the three stimuli will capture attention equally. However, if we calculate the total relevance by multiplying their intrinsic and goal relevances, the total relevance of the third stimulus will be higher than that of the other two, and it will capture more attention. The literature provides arguments supporting the assumption of the multiplication relationship and the alignment of intrinsic and goal relevances. For instance, Correll et al.

(2014) showed that outgroup member faces no longer captured attention when these faces became goal-irrelevant in their experiment.

In the context of emotional attention, experimental designs have often opposed two categories of stimuli. Indeed, Öhman and Mineka (2001) opposed threatening vs. neutral stimuli; Vogt et al. (2008) high vs. low arousal stimuli; Mazzietti et al. (2014) conducive vs. not conducive to goal stimuli. Such experimental designs make it difficult to precisely study the effect of different relevance criteria on attention. Recently, several studies have turned to other methods of analysis in tasks investigating emotional attention. For example, using dot-probe tasks, Aday and Carlson (2019) and Carlson and Fang (2020) used linear correlation on attentional scores; Price et al. (2019) used drift-diffusion modeling; and Syrjänen et al. (2019) used a Bayesian linear model that highlighted the attentional preference between two competing stimuli. In line with these studies, we proposed an original paradigm to explore the impact of intrinsic and goal relevances on attentional biases.

First, we proposed a rating task to participants. They were asked to appraise the intrinsic relevance of two food and two drink stimuli. To construct a value of goal relevance of stimuli, we asked participants to appraise the two pieces of information that constitute the goal relevance: the level of conduciveness of stimuli toward goals to which they are linked (satisfying hunger for food stimuli and satisfying thirst for drink stimuli) and the current urgency of these goals. Second, participants performed a dot-probe task in which all possible couples among the four stimuli were presented. Thus, for each participant, we obtained an average value of the attentional bias generated by each stimulus relative to each others. Using the appraisals of the rating task and the attentional biases of the dot-probe task, we could confront our hypothesis that intrinsic and goal relevances have multiplicative effects on attention with the hypotheses that these relevances have additive effects on attention, and that only goal relevance has effects on attention. Furthermore, in the rating task, we also asked

participants to appraise the level of action tendency toward the four stimuli. Thus, we tested the hypothesis that intrinsic and goal relevances have multiplicative effects on action tendency. As done for attentional biases, we investigated whether the action tendency toward stimuli was best described by multiplying or adding their levels of intrinsic and goal relevances, or by considering their level of goal relevance only.

Although the appraisal mechanism seems to operate mainly outside the field of consciousness (Sander et al., 2005), we assumed that at least some part of the information of the appraisal is accessible to consciousness, as proposed by Scherer (2009). To check the validity of the responses of the rating task, participants were asked to perform it again after the dot-probe task. Variations on the goals urgency and the action tendency were expected between the two rating tasks, insofar as the place of goals varies over time in the goals hierarchy and thus has an impact on action tendency. In contrast, no variations were expected to intrinsic relevance and conduciveness of stimuli toward goals that are supposed to be fairly stable over time. We also wanted to check the reliability of the rating task responses using non-verbal evidence. Indeed, the opportunity of consuming the products seen during the experiment was offered to the participants, if they wished, when the experiment was over. The aim was to observe, in the second rating task, whether participants who decided to consume a product reported a higher level of urgency of satisfying the goal (i.e., by drinking or eating) and a higher level of action tendency than those who decided not to consume it.

## **Method**

### **Participants**

Forty-seven students (26 women, 20 men;  $M = 21.6$  years,  $SD = 4$ ) from Lyon University participated to the study. Participants had normal or corrected-to-normal vision.

We decided to base our sample size on Vogt et al. (2017) experiments to the extent that our experiment was quite similar to theirs. They calculated that the minimal sample size had to be

16 at an alpha of .05, a power of .85, and a size effect of  $d = .83$ . Since we fitted linear models, we judged the sample size of 16 to be insufficient. However, we did not have prior information on coefficient of determination generated by models on this type of data to calculate sample size. We therefore decided to set the sample size at a minimum of 40.

## **Materials**

The experiment was run on a 21" iMac in a soundproofed room using SuperLab 6 software (Haxby et al., 1991).

## **Stimuli**

Four product pictures were used as stimuli. Two drink pictures: a bottle of Vittel water noted "V" and a bottle of Coca-Cola noted "C", and two food product pictures: a Harry's sandwich bread noted "H" and a Mars chocolate bar "M". These pictures were cropped and pasted on a dark blue square (400\*400 pixels). The choice of these products pictures was made in part because their shape was quite similar. The size of each of them was adjusted so that their height ( $M = 336$  pixels,  $SD = 8$ ) and width ( $M = 93$  pixels,  $SD = 0$ ) were identical. Each product picture was edited to remove unnecessary information from the packaging. The objective was to standardize them by leaving only the product with its brand name in the center of the image. All treatments were performed on Affinity Photo.

A rating task was designed so that 12 questions were asked in relation with the product pictures. Each product picture appeared 3 times, each time with a different question. Questions about the intrinsic relevance were "Do you like this food in general terms?" or "Do you like this drink in general terms?". Questions about the conduciveness toward a goal were "Do you feel that this food could satisfy you hunger in general?" or "Do you feel that this drink could satisfy you thirst in general?". Questions about the action tendency were "Would you like to consume this food now?" or "Would you like to consume this drink now?". Finally, two additional questions were asked about the current urgency of satisfying physiological goals,



that is, hunger and thirst: “Do you feel that satisfying your hunger is urgent now? and “Do you feel that satisfying your thirst is urgent now?”.

### **Design and procedure**

The rating task and the dot-probe task were both presented on a black background on the computer screen.

#### ***Rating task design***

The rating task was proposed twice: before and after the dot-probe task in order to check the reliability of the participants' answers. Responses were given on a quantitative nine-point scale, from 1 (*not at all*) to 9 (*extremely*) using the keyboard keys 1 to 9.

#### ***Dot-probe task design***

Each trial of the dot-probe task corresponded to the successive appearance of three events, as described in Figure 1. First, a white fixation cross (16\*16 pixels) with two white squares (401\*401 pixels) whose center was at 250 pixels to the left and to the right of the cross, appeared during a random variable time (from 500 to 1000 ms). Second, a stimulus appeared in the center of each square for 250 ms. All the possible couples among the four pictures were presented (V vs. C, V vs. H, V vs. M, C vs. H, C vs. M, and H vs. M). Third, a white dot (26\*26 pixels) appeared during 100 ms in the center of one of the two squares, immediately after the pictures offset. The task was divided into four blocks. In each block, each couple was presented 20 times, 10 times with a given stimulus in the left square and 10 times with the same stimulus in the right square. In each case, the dot appeared 5 times in the middle of the left square and 5 times in the middle of the right square. Therefore, each participant performed 480 trials of the dot-probe task.

#### ***Procedure***

Before starting the experiment, participants had to complete a consent form. Then, participants were asked to sit in front of the computer to complete the rating task a first time.

Once the rating task was completed, participants were asked to maintain their head at a distance of sixty centimeters from the screen, and to place the left index on the "s" key, and the right index on the "k" key of the keyboard. Participants were also asked to stare at the fixation cross throughout the whole experiment and to respond as quickly and precisely as possible (Chica et al., 2014) with the "s" key if the dot appeared on the left square, and with the "k" key if it appeared on the right square. The first event of the next trial appeared immediately after the response. Once the instructions had been presented, participants carried out twelve practice training trials with the experimenter to ensure that the task was clearly understood. Participants could choose to repeat the training if they felt it was necessary; otherwise, the experimenter would leave the room so that they could start the dot-probe task. At the end of the dot-probe task, participants had to perform the rating task a second time and to fill out a debriefing questionnaire to ensure they understood and performed the task correctly. Finally, participants were proposed to consume one or two products encountered during the task: They could choose one drink, one piece of food, one drink plus one piece of food, or nothing. In case participants chose something, they were asked whether they intended to consume the product immediately or later.

### **Transparency and Openness**

This article met the level 2 requirement of the Transparency and Openness Guidelines (Nosek et al., 2015). We reported how we determined our sample size, all data exclusions, all manipulations, all measures in the study, and we followed JARS (Kazak, 2018). Raw data, pre-treatment and analysis codes, and material from the study reported in this article are available through the Open Science Framework repository: <https://osf.io/27vfb/>. This study was not preregistered. Data were analyzed using R (version 1.4) with the base package and seven add-ons: the package *nortest* version 1.0-4 (Gross et al., 2015), the package *lmttest* version 0.9-37 (Zeileis & Hothorn, 2002), the package *car* version 3.0-2 (Fox et al., 2007), the package *effsize*

version 0.7.4 (Torchiano & Torchiano, 2018), the package *lm.beta* version 1.5-1 (Behrendt, 2014), the package *heplots* version 1.3-5 (Fox et al., 2009), and the package *emmeans* version 1.6.0 (Lenth, 2021).

## **Data processing**

### ***Construction of variables from the rating task data***

Composite variables were calculated from the rating task data to obtain a value of goal relevance and values of multiplied and added intrinsic and goal relevances for each stimulus and for each participant. First, the composite variable of the goal relevance was calculated by taking a weighted mean of the responses to the two goal relevance questions (i.e., conduciveness of the stimuli toward goals and goals urgency). Second, the composite variables multiplied and added intrinsic and goal relevances (noted “multiplied relevances” and “added relevances”) were calculated by multiplying and adding, respectively, the responses to the intrinsic relevance question of the rating task with the goal relevance composite variable described previously. The choice of using a weighted mean to calculate goal relevance was motivated by the fact that the weighting factor of goals urgency to be applied on conduciveness of stimuli toward goals was not known. So, different weighting factors of goals urgency to be applied on conduciveness of stimuli toward goals were tested. Among these different weighting factors, those that allowed goal relevance, multiplied relevances, and added relevances to best describe attentional biases and action tendency on further analyses were selected. The details of these calculations are reported in the online supplemental material. Finally, the action tendency variable was simply defined as the responses to the action tendency question of the rating task.

### ***Calculation of Attentional Bias Indices***

Attentional Bias Indices (ABIs; Vogt et al., 2013) of couples of stimuli were constructed from the dot-probe task data to provide a measure of the attentional bias of the stimuli against

each others for each participant. The first and second terms of the couples were as follows: V vs. C, V vs. H, V vs. M, C vs. H, C vs. M, and H vs. M. ABIs were calculated by subtracting the average response times (RTs) of trials where the dot appeared at the location of the stimulus of the first term of the couples from the average RTs of trials where the dot appeared at the location of the stimulus of the second term of the couples. Positive ABIs mean that the stimulus of the first term of the couples captured attention stronger, whereas negative ABIs mean that the stimulus of the second term of the couples captured attention stronger.

### **Statistical analyses**

Seven participants were excluded from the analysis. One participant's file failed to save. The other six were not retained because they were identified as not sufficiently involved in the task. This was based on a threshold that we defined as more than 5% error in the dot-probe task. Trials that generated erroneous responses were discarded from the analyses on RT (this included 1.65% of the data). For each condition of each participant (i.e., a condition corresponded to all trials where the dot appeared at the location of a specific stimulus of each couple), RTs deviating more than 2.5 SDs from the mean RT were excluded (2.3% of the data). However, before performing this treatment, RTs shorter than 150 ms and longer than 1000 ms – which are generally defined as reflecting anticipatory and delayed responses, respectively (Vogt et al., 2013) – were discarded in order not to bias the computations of mean response times (2.4%). Thus, a total of 6.3% of the 19,200 data were discarded. Although the large number of trials that were carried out, no differences were observed on ABIs per participant and per stimuli couple between the first two and last two blocks  $t(39) = 1.4, p = .1$ .

### ***Model fitting with ABIs as dependent variable***

For each hypothesis, a corresponding linear regression model was built: the goal relevance driven ABIs model, the multiplied relevances driven ABIs model, and the added relevances driven ABIs model. Each model had the same shape, so that ABIs of couples were entered as

dependent variable with two predictors: the composite variables of the first term of the couples and the composite variables of the second term of couples. As mentioned in the data processing section, the best-fitting model of each composite variable on ABIs was selected according to the best weighting factor of goal urgency to be applied in the goal relevance calculation (see online supplemental material for details). Surprisingly, the best-fitting models were those that did not take into account the goal urgency on the goal relevance calculation. We suggested that for most participants, the difference between the urgency of the two goals (i.e., satisfying hunger and thirst) was not sufficiently important to impact ABIs. We observed that only 6 participants rated a substantial difference between the two urgencies. Thus, the step of weighting factor selection of goals urgency was carried out again by considering the level of goals urgency for these 6 participants who demonstrated the bigger difference (the level of goals urgency of the remaining participants being set to 0). The best weighting factor was 0.3 for the goal relevance driven ABIs model and 0.4 for the multiplied relevances driven ABIs model. Insofar the best weighting factor was 0 for the added relevances driven ABIs model, goals urgency was not taken into account in this case.

### ***Model fitting with action tendency as dependent variable***

For each hypothesis, a corresponding linear regression model was built: the goal relevance driven action tendency model, the multiplied relevances driven action tendency model, and the added relevances driven action tendency model. As done for ABIs, the best-fitting model of each composite variable on action tendency was selected according to the best weighting factor of goals urgency to be applied in the goal relevance calculation (see online supplemental material for details). The best weighting factor was 1 for the goal relevance driven action tendency model and 2 for the multiplied and added relevances driven action tendency models.

First, the model consistencies were tested by forward selection from null models. Two steps were carried out in the forward selection when ABIs were entered as dependent variable. First,

the goodness of fit of step 1, in which the only entered predictor was the composite variables of the term of the couples giving the best improvement, was compared to that of the null model. Second, the goodness of fit of step 2, in which the composite variables of the second term of the couples were also entered as predictor from the step 1, was compared to step 1. Only one step was carried out when action tendency was entered as dependent variable. Composite variables of stimuli were entered as predictor from null model, then its goodness of fit was compared to that of the null model. Second, the goodness of fit of consistent models were then compared with the Akaike Information Criterion (AIC) to highlight the most likely hypotheses. According to Burnham and Anderson (2004), the model with the lower AIC ( $AIC_{\min}$ ) is the best in terms of likelihood. The decision threshold was set on  $\Delta AIC (AIC_{\text{model}} - AIC_{\min}) > 2$  to decide whether the model associated with  $AIC_{\min}$  was more parsimonious than that associated with  $AIC_{\text{model}}$  (Burnham & Anderson, 2004; Hu, 2007). Finally, to verify whether the dependent variables correctly differed by the level of the composite variables of the most supported hypotheses, patterns were built based on the mean. On the one hand, four patterns were built concerning the ABIs to the extent that two variables were entered as predictors in the models: the “high-low” pattern was defined by above or equal to average rating of the first term and below-average rating of the second term of the couples. Conversely, the “low-high” pattern was defined by below-average rating of the first term and above or equal to average rating of the second term of the couples. The “high-high” pattern was defined by above or equal to average rating of the first term and above-average rating of the second term of the couples, and the “low-low” pattern was defined by below or equal to average rating of the first term and below-average rating of the second term of the couples. On the other hand, two patterns were built concerning the action tendency to the extent that only one variable was entered as predictor in the models: the “high” pattern was defined by above or equal to average rating of the stimuli and the “low” pattern was defined by below-average rating to the stimuli. One-way ANOVA was performed

on ABIs with the four patterns, and one-way ANOVA was performed on action tendency with the two patterns. Effect size was analyzed using  $\eta^2_p$  (small effect = .01, medium effect = .06, large effect = .14; Cohen, 1973). One-tailed planned comparisons with Bonferroni correction were then performed between the patterns on ABIs to the extent that the ANOVA was performed on more than two patterns. Effect size was measured by Cohen's  $d$  (small effect = .20, medium effect = .50, large effect = .80; Cohen, 1992) and 95% Confidence Interval (CI) was reported.

For all models, checks were used to assess for appropriate error normality (Kolmogorov-Smirnov test) and homoscedasticity (Goldfeld-Quandt test). Multicollinearity was checked for multiple models (variation inflation factor < 2.5 for each variable of each model) and autocorrelation was checked for models with continuous variables as predictors (Durbin-Watson test).

## Results

### Rating task reliability check

To check the reliability of the participants' answers in the rating task, we compared the answers in the rating task before and after the dot-probe task. As expected, participants reported higher rating scores in the second than the first rating task for the urgency to satisfying thirst (first rating:  $M = 5$ ,  $SD = 1.8$ ; second rating:  $M = 5.8$ ,  $SD = 1.9$ ),  $t(39) = 3.95$ ,  $p < .001$ , to satisfying hunger (first rating:  $M = 3.1$ ,  $SD = 1.8$ ; second rating:  $M = 3.7$ ,  $SD = 2.1$ ),  $t(39) = 3.01$ ,  $p < .005$ , and for action tendency toward products (first rating:  $M = 5.1$ ,  $SD = 2.8$ ; second rating:  $M = 4.6$ ,  $SD = 2.9$ ),  $t(39) = 4.16$ ,  $p < .001$ . In contrast, no difference appeared between the first and second rating tasks for intrinsic relevance and conduciveness, which were assumed to be relatively stable over time. Then, to obtain a non-verbal reliability check, we compared participants' responses to the second rating task with their decision to consume a product immediately after the task. This analysis could only be carried out on the Vittel water bottle

because a few participants only chose the other products ( $n = 18$  for the Vittel water bottle and  $n < 5$  for the sum of the other products). As expected, participants who directly consumed the product after the task had reported a higher score of thirst urgency (taken:  $M = 6.7$ ,  $SD = 1.4$ ; not taken:  $M = 5.1$ ,  $SD = 2$ ),  $t(39) = 2.96$ ,  $p < .005$ , and a higher score of action tendency towards product (taken:  $M = 8.5$ ,  $SD = .08$ ; not taken:  $M = 6.5$ ,  $SD = 2$ ),  $t(39) = 3.8$ ,  $p < .001$ , compared to those who did not take it. These results support the validity of using our questionnaires to represent participants' evaluation of their internal state and appraisal of stimuli values.

### **Relevances and attention**

The descriptive statistics and correlations between composite variables and ABIs of each model are reported in the Table 1 of the online supplemental material. Consistencies of the models analyzed for each hypothesis are reported in Table 1. For each model, the composite variables of the first term of the couples provided a better improvement than the second one. Thus, the composite variables of the first term of the couples were added in step 1 from the null model and the composite variables of the second term of the couples were added in step 2 from step 1. Regarding the model of goal relevance driven ABIs, the goodness of fit increased in step 1 compared to the null model,  $F(1, 238) = 25.23$ ,  $p < .001$ . However, no significant difference appeared between the goodness of fit of step 2 and step 1,  $F(1, 237) = 1.79$ ,  $p = .18$ . Regarding the model of multiplied relevances driven ABIs, the goodness of fit increased in step 1 compared to the null model,  $F(1, 238) = 26.25$ ,  $p < .001$ , and the goodness of fit increased in step 2 compared to step 1,  $F(1, 237) = 4.21$ ,  $p < .05$ . Regarding the model of added relevances driven ABIs, the goodness of fit increased in step 1 compared to the null model,  $F(1, 238) = 23.26$ ,  $p < .001$ . However, no significant difference appeared between the goodness of fit of step 2 and step 1,  $F(1, 237) = 3.27$ ,  $p < .10$ . Insofar as the model of multiplied relevances driven ABIs was consistent, we described it below in more details. The first term of the couples impacted ABIs positively,  $\beta = .31$ , 95% CI [.15, .34], and significantly,  $t(39) = 5.09$ ,  $p < .001$ .



In contrast, the second terms of the couples impacted ABIs negatively,  $\beta = -.13$ , 95% CI [-32, .01], and significantly,  $t(39) = -2.05$ ,  $p < .05$ . Thus, for the multiplied relevances driven ABIs model, the more the composite variables of the first term of the couples increased, the more the ABIs tended to be positive. In addition, the more the composite variables of the second term of the couples increased, the more the ABIs tended to be negative. Our results revealed that the effects of intrinsic and goal relevances seems to be rather multiplicative than additive on attentional biases. Thus, our results showed the importance to consider both intrinsic and goal relevances to explain attentional biases. In addition, it appeared that their effects on attentional biases are rather multiplicative than additive.

Finally, a one-way ANOVA was performed on ABIs with the four patterns of multiplied relevances driven ABIs to the extent that this model appeared to be better at predicting ABIs. Results are illustrated in Figure 2. This analysis revealed a significant main effect of the patterns on ABIs,  $F(3,236) = 8.85$ ,  $p < .001$ ,  $\eta^2_p = .10$ . Contrasts analyses showed a greater ABIs for “high-low” pattern than “low-high” pattern,  $t(39) = 4.82$ ,  $p < .001$ ,  $d = .84$ , 95% CI [.47, 1.21]. Hence, in the case of the multiplied relevances driven ABI model, when a stimulus was appraised as more relevant than another, the former captured attention stronger than the latter.

### **Relevances and action tendency**

The descriptive statistics and correlations between predictors and dependent variable are reported in the Table 2 of the online supplemental material. Model consistencies were analyzed for each hypothesis and are reported in Table 2.

Regarding the model of goal relevance driven action tendency, the inclusion of the goal relevance of stimuli in the null model increased the action tendency goodness of fit,  $F(1, 158) = 83.55$ ,  $p < .001$ . Regarding the model of multiplied relevances driven action tendency, the inclusion of the multiplied relevances of stimuli increased the action tendency goodness of fit,  $F(1, 238) = 166.61$ ,  $p < .001$ . Regarding the model of added relevances driven action tendency,

the inclusion of the added relevances of stimuli also increased the action tendency goodness of fit,  $F(1, 238) = 163.98, p < .001$ . Insofar the models of each hypothesis were consistent, these models were described in more details and their goodness of fit was compared.

The composite variables of goal relevance, multiplied relevances, and added relevances impacted action tendency positively and significantly, respectively,  $\beta = .59, 95\% \text{ CI } [.70, .88], t(39) = 9.14, p < .001$ ;  $\beta = .72, 95\% \text{ CI } [.09, .13], t(39) = 12.91, p < .001$ ; and  $\beta = .71, 95\% \text{ CI } [.48, .65], t(39) = 12.81, p < .001$ . Thus, for these three composite variables, the more the variable increased, the more the action tendency tended to be positive. The comparison of action tendency goodness of fit of goal relevance ( $\text{AIC} = 720.65$ ) with the one of added relevances ( $\text{AIC} = 674.66$ ) provided decisive support for the added relevances model  $\Delta\text{AIC} = 45.99$ . However, although the AIC of multiplied relevances driven action tendency ( $\text{AIC} = 673.36$ ) was lower than the one of added relevances driven action tendency, we could not conclude on a difference between the action tendency goodness of fit of the two models,  $\Delta\text{AIC} = -1.33$ . Results suggested that intrinsic and goal relevances seems to have effects on action tendency, but we cannot conclude on the quality of these effects.

Two one-way ANOVAs were performed on action tendency to the extent that substantial support was provided between multiplied relevances driven action tendency and added relevances driven action tendency models. Results are illustrated in Figure 3. Regarding the multiplied relevances model, the analysis revealed a significant main effect of the patterns on action tendency,  $F(1,158) = 95.66, p < .001, \eta^2_p = .38$ . Regarding the added relevances model, the analysis also revealed a significant main effect of the patterns on action tendency,  $F(1,158) = 114.95, p < .001, \eta^2_p = .42$ . Hence, for multiplied relevances and added relevances, the higher a stimulus is appraised, the stronger is the action tendency toward the stimulus.

## Discussion

Over the past decade, numerous studies have shown that attention is determined by events that are conducive to goals depending to the urgency for individuals (e.g., Vogt et al., 2013; Vogt et al., 2017). Our results suggest that attention is admittedly influenced by events that are conducive to momentarily important goals (i.e., goal relevance), but also by intrinsic relevance. In addition, it seems that intrinsic relevance and goal relevance have multiplicative effects on attention. Furthermore, our findings suggest that action tendency related to stimuli is also triggered by the appraisals of intrinsic and goal relevances.

However, recent literature evidenced that the impact of emotion on attention functioning is still not very well known. Indeed, numerous studies highlighted the poor reliability of attentional bias, even in the case of threat for anxious individuals that seemed indisputable (Kruijt et al., 2019; McNally, 2019). Recent works emphasize the importance to take into account both bottom-up and top-down processes when studying attentional biases to better understand their functioning (Abado et al., 2020; Macleod et al., 2019; Sussman et al., 2016).

The relevance appraisal aims to evaluate which stimuli are currently most important for the safety and well-being of individuals (Brosch et al., 2010; Grandjean & Scherer, 2014). A skeptical audience that does not distinguish intrinsic relevance from goal relevance (e.g., Feldman Barrett & Russell, 1998; Kron et al., 2015), might argue that goal relevance appraisal alone could address this concern. According to the predictive coding hypothesis and free-energy principle (Friston, 2005; 2010), organisms aim to respond as quickly as possible by minimizing free energy (e.g., predictive error). To this end, system architecture and data processing would tend to be as simple as possible. Specifically, some neural networks would have evolved to compress data that are perfectly predictable and that would not need to be re-appraised. As detailed in our introduction, stimuli that lead to the achievement of currently important goals generate states of pleasure (Berridge, 2009; Berridge et al., 2009; Castro & Berridge, 2014;

Ferreri, et al., 2019). All experiences of pleasure states stored in memory would then be concatenated into what we call intrinsic relevance in order to quickly and cheaply assign meaning to stimuli. Thus, intrinsic relevance would trigger a rapid action tendency on the basis of raw stimulus information and the action tendency would then be refined by assessing goal relevance. This last proposal raises a crucial point: Both intrinsic relevance and goal relevance trigger action tendency toward stimuli (Aue & Scherer, 2011; Scherer, 2010). Our results support this hypothesis by highlighting that action tendency is best predicted by the conjunction of intrinsic relevance and goal relevance.

Because valence appeals to the action tendency toward stimuli (see Lewin, 1951; Scherer, 2013), it is important to view it as multiple components, incorporating both intrinsic relevance and goal relevance. More specifically, we suggest that “positive valence”, which measures both intrinsic relevance and goal relevance and “pleasantness”, which measures intrinsic relevance and current pleasure state are not interchangeable. Let us digress for a sentence to clarify our point about the distinction between “current pleasantness” (i.e., measuring current pleasure state) and “general pleasantness” (i.e., measuring intrinsic relevance). Not considering the difference between valence and intrinsic relevance could create debates based on misinterpreted results. For instance, studies that manipulate goal relevance of stimuli and highlight change in positivity could be interpreted as providing evidence for the dependence of intrinsic relevance and goal relevance (Ferguson & Bargh, 2004; Moors & De Houwer, 2001; Moors et al., 2005; Spruyt et al., 2018). However, as we explained above, evidencing changes in positivity does not necessarily reflect changes in intrinsic relevance and does not constitute an argument for the previous proposal. Regarding the relationship between intrinsic relevance and goal relevance, we claim that our results support the assumption of independence of intrinsic relevance and goal relevance: Although variations of goal relevance were evidenced between the two reports of participants before and after the task, no difference was observed regarding

intrinsic relevance. In other words, intrinsic relevance did not appear to be affected by the variations of goal relevance.

We previously evoked the notion of temporality between intrinsic relevance and goal relevance. This idea is carried out by the assumption of a sequential temporal unfolding of the appraisal proposed by appraisal theorists (Aue et al., 2007; Lanctôt & Hess, 2007; Grandjean & Scherer, 2008). This assumption proposes that although they may overlap to some extent, intrinsic relevance appraisal occurs in a faster temporal window than goal relevance appraisal. Indeed, intrinsic relevance could be considered in terms of a fast, low-level, appraisal providing early but not precise information about the current necessity to approach or avoid a stimulus, while goal relevance could be considered as a relatively slower, high-level, appraisal providing late but more precise information about the current necessity to approach or avoid a stimulus (Scherer, 2013).

Thus, during the goal relevance appraisal processing of stimuli, the more time passes, the more refined the prediction of the current importance of the stimuli becomes. This imbalance between speed and precision between intrinsic relevance and goal relevance motivated us to propose the later assumption that two time windows of attentional biases could be proposed. In the first time window, attention would be impacted primarily by intrinsic relevance. In the second time window, the more time passes, the more attention would be impacted by goal relevance. We suggest that under the specific temporal conditions of our study, attention is modulated by both intrinsic and goal relevance. We develop further some empirical evidence that support our point.

Regarding the first time window: In several studies, among participants who reported no differences in craving and satiety levels, more frequent initial orientations toward high-fat food were observed in overweight individuals who consumed more high-fat food than healthy-weight individuals (Werthmann et al., 2011; Nijs et al., 2010; Castellanos et al., 2009). This early

attentional bias was also evidenced in the case of addiction disorders (e.g., nicotine and alcohol dependence; Field & Cox, 2008).

In addition, Chapman et al. (2019) tested the reliability of participants responses to the dot-probe task using angry and neutral faces as cues by manipulating SOA. Their results showed internal reliability – although poor – only when cues appeared for 100 ms (but not for 300, 500, and 900 ms).

Regarding the second time window: The design of our experiment, where the cue appeared for 250 ms, seems well suited to explore this time window where intrinsic relevance and goal relevance both contribute to affect attention. We believe that the credibility of the inference on goal relevance is not achieved yet. Therefore, the total relevance is calculated by combining intrinsic relevance and goal relevance to improve the credibility of the relevance inference. Furthermore, Field et al. (2004) evidenced that in social drinkers, alcohol craving was positively correlated with attention maintaining to alcohol pictures that remained on the screen for 2000 ms. Similarly, Werthmann et al. (2013) showed positive correlation between chocolate craving and maintaining attention on chocolate pictures during this second time window of the appraisal process (i.e., mean initial gaze duration of 189 ms with standard-deviation of 260 ms, and mean dwell time of 261 ms with standard-deviation of 361 ms). Finally, a recent study (Soleymani et al., 2020) used an attentional task (i.e., a free-viewing multi-stimuli eye tracking task) to predict addiction, specifically by trying to link attentional bias toward alcohol related stimuli with predictors related to alcohol disorder. Although in a first session, correlation was showed between maintaining attention (mean about 400 ms with a standard-deviation about 150 ms) and both craving and alcohol habits, results of a second session evidenced that maintaining attention was only correlated with craving, whereas mean first gaze fixation was only correlated with alcohol habits.

During these temporal windows, we suggest that the functioning of the relevance appraisal may operate in terms of Bayesian inference (Kolossa et al., 2015). Specifically, the probability that one stimulus is more important than another would be estimated based on their respective probability of having been important in the past. Based on prior knowledge, the brain would infer about intrinsic relevance (i.e., generally pleasant or unpleasant value) and goal relevance (i.e., goals that could be impacted by the stimuli, the importance of these goals according to the state of the organism, and their place in the current hierarchy of goals). An analogy of the functioning of the intrinsic relevance and goal relevance appraisal through the two temporal windows could be proposed using a machine learning example. Imagine a Bayesian model aims at predicting as accurately as possible the current importance of stimuli. The model receives input about the intrinsic relevance of stimuli and about the goal relevance of stimuli (i.e., two informations: Their conducive nature toward goals, and the current importance of these goals). The most accurate inference is the one that is computed from the goal relevance data, because it contains the most precise information about the current goals/needs to approach or avoid stimuli. However, the execution time of this computation is supposed to be longer than the computation that compiles the intrinsic relevance data because it requires a larger number of calculations. Thus, if we stop the model compilation at, say, 100 ms (corresponding to the first temporal window), only intrinsic relevance computation could return a credible prediction. By increasing the time constraint to nearly 250 ms (corresponding to the second temporal window), the computation of goal relevance could also return a credible prediction. Therefore, the best prediction could be provided by intrinsic relevance in some cases, and by goal relevance in other cases. This implies that in case of convergent predictions between the two computations, the final prediction accuracy would probably be very high. This is exactly the purpose of the multiplicative relationship. Indeed, in case of alignment between intrinsic relevance and goal relevance, the multiplicative relationship will generate a higher score of total relevance than the

additive relationship. In contrast, in the absence of alignment, that is, in the extreme case when the goal relevance is fixed to zero, for example, the multiplicative relationship will predict that the whole relevance will shift to zero (Correll et al., 2014).

In sum, the present experiment has great implication for the importance of considering both the intrinsic relevance and the goal relevance as triggers for the emotional episode, attention orientation, and action tendency. Specifically, our findings suggest that intrinsic and goal relevance have multiplicative effects on attention and trigger the action tendency toward stimuli.

### **Limitations**

The validity of the self-evaluation in our analyses could be discussed insofar as evaluations occur non-consciously (see Mazzietti et al., 2015). However, appraisal theorists suggest that some of these appraisals are accessible to individuals' consciousness (Scherer, 2009). So, the appraisal questionnaires seemed quite solid in view of the assumptions that were made about the controls we made. As expected, only the self-reported levels of thirst, hunger, and action tendency differed between the beginning and the end of the dot-probe task. Indeed, the physiological state of the organism can vary and more or less trigger the desire to eat or drink and generate more or less action tendency towards a stimulus depending on the level of importance of the goals that this stimulus may satisfy. In addition, participants who chose to take the bottle of water for direct consumption reported that they were more thirsty and more willing to consume it than others, which provided a non-verbal argument for the reliability of the measures we collected using questionnaires.



## References

- Abado, E., Aue, T., & Okon-Singer, H. (2020). The Missing Pieces of the Puzzle: A Review on the Interactive Nature of A-Priori Expectancies and Attention Bias toward Threat. *Brain Sciences*, *10*(10), 745. <https://doi.org/10.3390/brainsci10100745>
- Aday, J. S., & Carlson, J. M. (2019). Extended testing with the dot-probe task increases test–retest reliability and validity. *Cognitive Processing*, *20*(1), 65-72. <https://doi.org/10.1007/s10339-018-0886-1>
- Anderson, A. K. (2005). Affective influences on the attentional dynamics supporting awareness. *Journal of Experimental Psychology: General*, *134*(2), 258. <https://doi.org/10.1037/0096-3445.134.2.258>
- Aue, T., Flykt, A., & Scherer, K. R. (2007). First evidence for differential and sequential efferent effects of stimulus relevance and goal conduciveness appraisal. *Biological Psychology*, *74*(3), 347-357. <https://doi.org/10.1016/j.biopsycho.2006.09.001>
- Aue, T., & Scherer, K. R. (2008). Appraisal-driven somatovisceral response patterning: Effects of intrinsic pleasantness and goal conduciveness. *Biological Psychology*, *79*(2), 158-164. <https://doi.org/10.1016/j.biopsycho.2008.04.004>
- Aue, T., & Scherer, K. R. (2011). Effects of intrinsic pleasantness and goal conduciveness appraisals on somatovisceral responding: Somewhat similar, but not identical. *Biological Psychology*, *86*(1), 65-73. <https://doi.org/10.1016/j.biopsycho.2010.10.008>
- Behrendt, S. (2014). lm.beta: Add standardized regression coefficients to lm-Objects. *R package version 1.5, 1*. <https://CRAN.R-project.org/package=lm.beta>
- Berridge, K. C. (2009). ‘Liking’ and ‘wanting’ food rewards: brain substrates and roles in eating disorders. *Physiology & behavior*, *97*(5), 537-550. <https://doi.org/10.1016/j.physbeh.2009.02.044>

- Berridge, K. C., Robinson, T. E., & Aldridge, J. W. (2009). Dissecting components of reward: 'liking', 'wanting', and learning. *Current opinion in pharmacology*, 9(1), 65-73.  
<https://doi.org/10.1016/j.coph.2008.12.014>
- Bradley, M. M., Codispoti, M., Cuthbert, B. N., & Lang, P. J. (2001). Emotion and motivation I: Defensive and appetitive reactions in picture processing. *Emotion*, 1(3), 276.  
<https://doi.org/10.1037//1528-3542.1.3.276>
- Brosch, T., Pourtois, G., & Sander, D. (2010). The perception and categorisation of emotional stimuli: A review. *Cognition and Emotion*, 24(3), 377-400.  
<https://doi.org/10.1080/02699930902975754>
- Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods & Research*, 33(2), 261-304.  
<https://doi.org/10.1177/0049124104268644>
- Carlson, J. M., & Fang, L. (2020). The stability and reliability of attentional bias measures in the dot-probe task: Evidence from both traditional mean bias scores and trial-level bias scores. *Motivation and Emotion*, 44, 657-669. <https://doi.org/10.1007/s11031-020-09834-6>
- Castellanos, E. H., Charboneau, E., Dietrich, M. S., Park, S., Bradley, B. P., Mogg, K., & Cowan, R. L. (2009). Obese adults have visual attention bias for food cue images: evidence for altered reward system function. *International journal of obesity*, 33(9), 1063-1073.  
<https://doi.org/10.1038/ijo.2009.138>
- Castro, D. C., & Berridge, K. C. (2014). Advances in the neurobiological bases for food 'liking' versus 'wanting'. *Physiology & behavior*, 136, 22-30.  
<https://doi.org/10.1016/j.physbeh.2014.05.022>
- Chapman, A., Devue, C., & Grimshaw, G. M. (2019). Fleeting reliability in the dot-probe task. *Psychological Research*, 83(2), 308-320. <https://doi.org/10.1007/s00426-017-0947-6>

- Chica, A. B., Martín-Arévalo, E., Botta, F., & Lupiáñez, J. (2014). The Spatial Orienting paradigm: How to design and interpret spatial attention experiments. *Neuroscience & Biobehavioral Reviews*, 40, 35-51. <https://doi.org/10.1016/j.neubiorev.2014.01.002>
- Cohen, J. (1973). Eta-squared and partial eta-squared in fixed factor ANOVA designs. *Educational and Psychological Measurement*, 33(1), 107-112. <https://doi.org/10.1177/001316447303300111>
- Cohen, J. (1992). A power primer. *Psychological bulletin*, 112(1), 155. <https://doi.org/10.1037/0033-2909.112.1.155>
- Correll, J., Guillermo, S., & Vogt, J. (2014). On the flexibility of attention to race. *Journal of Experimental Social Psychology*, 55, 74-79. <https://doi.org/10.1016/j.jesp.2014.05.013>
- Feldman Barrett, L., & Russell, J. A. (1998). Independence and bipolarity in the structure of current affect. *Journal of Personality and Social Psychology*, 74(4), 967. <https://doi.org/10.1037/0022-3514.74.4.967>
- Ferguson, M. J., & Bargh, J. A. (2004). Liking is for doing: the effects of goal pursuit on automatic evaluation. *Journal of personality and social psychology*, 87(5), 557-572. <https://doi.org/10.1037/0022-3514.87.5.557>
- Ferreri, L., Mas-Herrero, E., Zatorre, R. J., Ripollés, P., Gomez-Andres, A., Alicart, H., ... & Rodriguez-Fornells, A. (2019). Dopamine modulates the reward experiences elicited by music. *Proceedings of the National Academy of Sciences*, 116(9), 3793-3798. <https://doi.org/10.1073/pnas.1811878116>
- Field, M., & Cox, W. M. (2008). Attentional bias in addictive behaviors: a review of its development, causes, and consequences. *Drug and alcohol dependence*, 97(1-2), 1-20. <https://doi.org/10.1016/j.drugalcdep.2008.03.030>

- Field, M., Mogg, K., Zetteler, J., & Bradley, B. P. (2004). Attentional biases for alcohol cues in heavy and light social drinkers: the roles of initial orienting and maintained attention. *Psychopharmacology*, *176*(1), 88-93. <https://doi.org/10.1007/s00213-004-1855-1>
- Fox, J., Friendly, G. G., Graves, S., Heiberger, R., Monette, G., Nilsson, H., Ripley B., Weisberg, S., & Zeleis, A. (2007). The car package. *R Foundation for Statistical Computing*, Vienna, Austria. <https://CRAN.R-project.org/package=car>
- Fox, J., Friendly, M., & Monette, G. (2009). Visualizing hypothesis tests in multivariate linear models: The heplots package for R. *Computational Statistics*, *24*(2), 233-246. <https://CRAN.R-project.org/package=heplots>
- Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological sciences*, *360*(1456), 815-836. <https://doi.org/10.1098/rstb.2005.1622>
- Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127-138. <https://doi.org/10.1038/nrn2787>
- Gauthier, J., & Bouchard, S. (1993). A French-Canadian adaptation of the revised version of Spielberger's State-Trait Anxiety Inventory. *Canadian Journal of Behavioural Science*, *25*(4), 559-578. <https://doi.org/10.1037/h0078881>
- Grandjean, D., & Scherer, K. R. (2008). Unpacking the cognitive architecture of emotion processes. *Emotion*, *8*(3), 341. <https://doi.org/10.1037/1528-3542.8.3.341>
- Grandjean, D., & Scherer, K. R. (2014). Chapitre 2. Théorie de l'évaluation cognitive et dynamique des processus émotionnels. In Sander, D. (Ed.). *Traité de Psychologie des Émotions*, 51-87. <https://doi.org/10.3917/dunod.sande.2014.01.0051>
- Gross, J., Ligges, U., & Ligges, M. U. (2015). Package 'nortest'. *Five omnibus tests for testing the composite hypothesis of normality*. <https://CRAN.R-project.org/package=nortest>

- Haxby, J. V., Parasuraman, R., Lalonde, F., & Abboud, H. (1993). SuperLab: General-purpose Macintosh software for human experimental psychology and psychological testing. *Behavior Research Methods, Instruments, & Computers*, 25(3), 400-405.  
<https://doi.org/10.3758/BF03204531>
- Hoefling, A., Likowski, K. U., Deutsch, R., Häfner, M., Seibt, B., Mühlberger, A., ... & Strack, F. (2009). When hunger finds no fault with moldy corn: Food deprivation reduces food-related disgust. *Emotion*, 9(1), 50-58. <https://doi.org/10.1037/a0014449>
- Hu, S. (2007). Akaike information criterion. *Center for Research in Scientific Computation*, 93.
- Kazak, A. E. (2019). Editorial: Journal article reporting standards. *American Psychologist*, 73, 1-2. <http://dx.doi.org/10.1037/amp0000263>
- Kolossa, A., Kopp, B., & Fingscheidt, T. (2015). A computational analysis of the neural bases of Bayesian inference. *Neuroimage*, 106, 222-237.  
<https://doi.org/10.1016/j.neuroimage.2014.11.007>
- Kron, A., Pilkiw, M., Banaei, J., Goldstein, A., & Anderson, A. K. (2015). Are valence and arousal separable in emotional experience? *Emotion*, 15(1), 35.  
<http://dx.doi.org/10.1037/a0038474>
- Kruijt, A. W., Parsons, S., & Fox, E. (2019). A meta-analysis of bias at baseline in RCTs of attention bias modification: no evidence for dot-probe bias towards threat in clinical anxiety and PTSD. *Journal of abnormal psychology*, 128(6), 563.  
<https://doi.org/10.1037/abn0000406>
- Lanctôt, N., & Hess, U. (2007). The timing of appraisals. *Emotion*, 7(1), 207.  
<https://doi.org/10.1037/1528-3542.7.1.207>
- Lang, P. J. (1995). The emotion probe: studies of motivation and attention. *American Psychologist*, 50(5), 372. <https://doi.org/10.1037/0003-066X.50.5.372>

- Lenth, R. (2021). Emmeans: Estimated marginal means, aka least-squares means. *R Package Version 1*, (2018). <https://CRAN.R-project.org/package=emmeans>
- Lewin, K. (1951). *Field theory in social science: Selected theoretical papers* (D. Cartwright, Ed.). Westport, CT: Greenwood Press. (Original work published 1938)
- MacLeod, C., Grafton, B., & Notebaert, L. (2019). Anxiety-linked attentional bias: is it reliable?. *Annual review of clinical psychology*, *15*, 529-554.  
<https://doi.org/10.1146/annurev-clinpsy-050718-095505>
- McNally, R. J. (2019). Attentional bias for threat: Crisis or opportunity?. *Clinical psychology review*, *69*, 4-13. <https://doi.org/10.1016/j.cpr.2018.05.005>
- Mazzietti, A., Sellem, V., & Koenig, O. (2014). From stimulus-driven to appraisal-driven attention: Towards differential effects of goal relevance and goal relatedness on attention? *Cognition and Emotion*, *28*(8), 1483-1492.  
<https://doi.org/10.1080/02699931.2014.884488>
- Mazzietti, A., Sellem, V., & Koenig, O. (2015). To See, or Not To See: Dynamic attentional interference by non-consciously perceived relevant stimuli.
- Moors, A., & De Houwer, J. (2001). Automatic appraisal of motivational valence: Motivational affective priming and Simon effects. *Cognition & Emotion*, *15*(6), 749-766.  
<https://doi.org/10.1080/02699930143000293>
- Moors, A., De Houwer, J., Hermans, D., & Eelen, P. (2005). Unintentional processing of motivational valence. *The Quarterly Journal of Experimental Psychology Section A*, *58*(6), 1043-1063. <https://doi.org/10.1080/02724980443000467>

Montagrin, A., Martins-Klein, B., Sander, D., & Mather, M. (2019). Effects of hunger on emotional arousal responses and attention/memory biases. *Emotion*.

<https://doi.org/10.1037/emo0000680>

Nijs, I. M., Muris, P., Euser, A. S., & Franken, I. H. (2010). Differences in attention to food and food intake between overweight/obese and normal-weight females under conditions of hunger and satiety. *Appetite*, *54*(2), 243-254. <https://doi.org/10.1016/j.appet.2009.11.004>

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., Buck, S., Chambers, C. D., Chin, G., Christensen, G., Contestabile, M., Dafoe, A., Eich, E., Freese, J., Glennerster, R., Goroff, D., Green, D. P., Hesse, B., Humphreys, M., Ishiyama, J., ... Yarkoni, T. (2015). SCIENTIFIC STANDARDS. Promoting an open research culture. *Science (New York, N.Y.)*, *348*(6242), 1422–1425.

<https://doi.org/10.1126/science.aab2374>

Öhman, A., Flykt, A., & Esteves, F. (2001). Emotion drives attention: detecting the snake in the grass. *Journal of Experimental Psychology: General*, *130*(3), 466.

<https://doi.org/10.1037/0096-3445.130.3.466>

Öhman, A., & Mineka, S. (2001). Fears, phobias, and preparedness: toward an evolved module of fear and fear learning. *Psychological Review*, *108*(3), 483.

<https://doi.org/10.1037/0033-295X.108.3.483>

Price, R. B., Brown, V., & Siegle, G. J. (2019). Computational modeling applied to the dot-probe task yields improved reliability and mechanistic insights. *Biological Psychiatry*, *85*(7), 606-612. <https://doi.org/10.1016/j.biopsych.2018.09.022>

Sander, D., Grandjean, D., & Scherer, K. R. (2005). A systems approach to appraisal mechanisms in emotion. *Neural Networks*, *18*(4), 317-352.

<https://doi.org/10.1016/j.neunet.2005.03.001>

- Scherer, K. R. (2001). Appraisal considered as a process of multilevel sequential checking. *Appraisal processes in emotion: Theory, methods, research*, 92(120), 57.
- Scherer, K. R. (2009). The dynamic architecture of emotion: Evidence for the component process model. *Cognition and Emotion*, 23(7), 1307-1351.  
<https://doi.org/10.1080/02699930902928969>
- Scherer, K. R. (2010). The component process model: A blueprint for a comprehensive computational model of emotion. In K. R. Scherer, T. Bänziger, & E. B. Roesch (Eds.), *Blueprint for affective computing: A sourcebook* (pp. 47–70). Oxford, UK: Oxford University Press.
- Scherer, K. R. (2013). The nature and dynamics of relevance and valence appraisals: Theoretical advances and recent evidence. *Emotion Review*, 5(2), 150-162.  
<https://doi.org/10.1177/1754073912468166>
- Smith, K. S., Berridge, K. C., & Aldridge, J. W. (2011). Disentangling pleasure from incentive salience and learning signals in brain reward circuitry. *Proceedings of the National Academy of Sciences*, 108(27), E255-E264.  
<https://doi.org/10.1073/pnas.1101920108>
- Soleymani, A., Ivanov, Y., Mathot, S., & de Jong, P. J. (2020). Free-viewing multi-stimulus eye tracking task to index attention bias for alcohol versus soda cues: Satisfactory reliability and criterion validity. *Addictive behaviors*, 100, 106117.  
<https://doi.org/10.1016/j.addbeh.2019.106117>
- Spruyt, A., Tibboel, H., De Schryver, M., & De Houwer, J. (2018). Automatic stimulus evaluation depends on goal relevance. *Emotion*, 18(3), 332-341.  
<https://doi.org/10.1037/emo0000361>



- Sussman, T. J., Jin, J., & Mohanty, A. (2016). Top-down and bottom-up factors in threat-related perception and attention in anxiety. *Biological psychology*, *121*, 160-172.  
<https://doi.org/10.1016/j.biopsycho.2016.08.006>
- Syrjänen, E., Fischer, H., & Olofsson, J. K. (2019). Background odors affect behavior in a dot-probe task with emotionally expressive faces. *Physiology & Behavior*, *210*, 112540.  
<https://doi.org/10.1016/j.physbeh.2019.05.001>
- Torchiano, M. & Torchiano, M.M. (2018) Package 'effsize'. <https://CRAN.R-project.org/package=effsize>
- Vogt, J., De Houwer, J., Koster, E. H., Van Damme, S., & Crombez, G. (2008). Allocation of spatial attention to emotional stimuli depends upon arousal and not valence. *Emotion*, *8*(6), 880. <https://doi.org/10.1037/a0013981>
- Vogt, J., De Houwer, J., Crombez, G., & Van Damme, S. (2013). Competing for attentional priority: temporary goals versus threats. *Emotion*, *13*(3), 587.  
<https://doi.org/10.1037/a0027204>
- Vogt, J., Koster, E. H., & De Houwer, J. (2017). Safety first: Instrumentality for reaching safety determines attention allocation under threat. *Emotion*, *17*(3), 528.  
<https://doi.org/10.1037/emo0000251>
- Werthmann, J., Roefs, A., Nederkoorn, C., & Jansen, A. (2013). Desire lies in the eyes: attention bias for chocolate is related to craving and self-endorsed eating permission. *Appetite*, *70*, 81-89. <https://doi.org/10.1016/j.appet.2013.06.087>
- Werthmann, J., Roefs, A., Nederkoorn, C., Mogg, K., Bradley, B. P., & Jansen, A. (2011). Can (not) take my eyes off it: attention bias for food in overweight participants. *Health Psychology*, *30*(5), 561. <https://doi.org/10.1037/a0024291>
- Zeileis, A., & Hothorn, T. (2002). Diagnostic checking in regression relationships. *R News*, *2*(3): 7-10. <https://CRAN.R-project.org/package=lmtest>

**Table 1**

*Forward Selection Results for the Composite Variables of the Terms of the Stimuli Couples in the Linear Regression with ABIs as Dependent Variable*

Variable	B (SE)	95% CI for B		$\beta$	$R^2$	$\Delta R^2$
		LL	UL			
<b>Goal relevance</b>						
Step 1					.09	.09 ***
Constant	-11.52 (2.51) ***	-16.45	-6.58			
First term	2.25 (.34) ***	1.36	3.13	.31		
Step 2					.10	.01
Constant	-9.18 (3.05) **	-15.19	-3.17			
First term	2.26 (.45) ***	1.38	3.14	.32		
Second term	-.73 (.54)	-1.80	.34	-.08		
<b>Multiplied relevances</b>						
Step 1					.10	.10 ***
Constant	-8.31 (1.9) ***	-12.07	-4.55			
First term	.25 (.05) ***	.15	.34	.32		
Step 2					.11	.01 *
Constant	-5.52 (2.34) *	-10.12	-.92			
First term	.24 (.05) ***	.15	.34	.31		
Second term	-.16 (.08) *	-.32	-.01	-.13		
<b>Added relevances</b>						
Step 1					.09	.09 ***
Constant	-12.28 (2.75) ***	-17.7	-6.87			
First term	0.97 (.20) ***	.57	1.36	.30		
Step 2					.10	.01 †
Constant	-7.51 (3.80) *	-14.99	-.02			
First term	.95 (.20) ***	.56	1.35	.30		
Second term	-.49 (.27) †	-1.02	.04	-.11		

**Table 2**

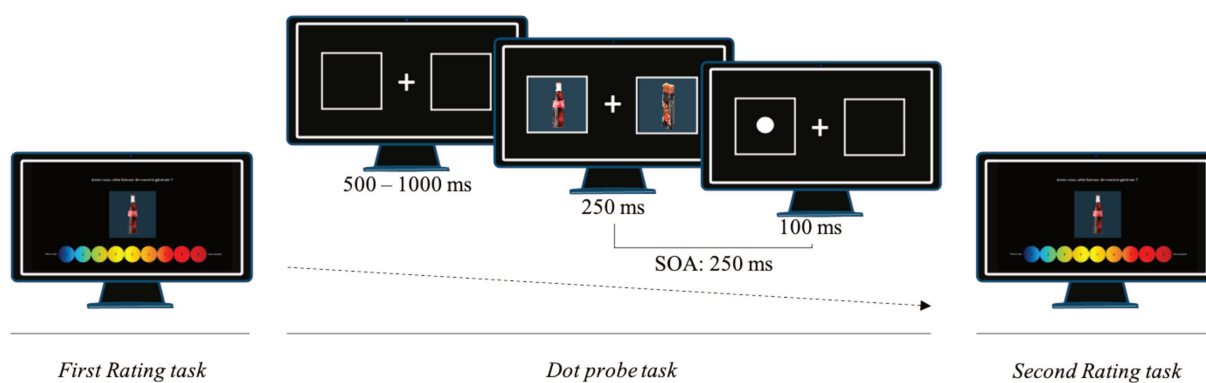
*Forward Selection Results for the Composite Variables of the Terms of the Stimuli Couples in the Linear Regression with Action Tendency as Dependent Variable*

Variable	<i>B</i> ( <i>SE</i> )	95% CI for <i>B</i>		$\beta$	$R^2$
		<i>LL</i>	<i>UL</i>		
Goal relevance					.35 ***
Constant	-.94 (.49)	-1.06	.88		
Coefficient	.89 (.09) ***	.69	1.08	.58	
Multiplied relevances					.51 ***
Constant	.99 (.28) ***	.43	1.56		
Coefficient	.11 (.01) ***	.10	.13	.71	
Added relevances					.51 ***
Constant	-1.67 (.48) ***	-2.61	-.73		
Coefficient	.57 (.04) ***	.48	.65	.71	

*Note.* CI = confidence interval; *LL* = lower limit; *UL* = upper limit; †  $p < .10$ . \*  $p < .05$ . \*\*  $p < .01$ . \*\*\*  $p < .001$ .

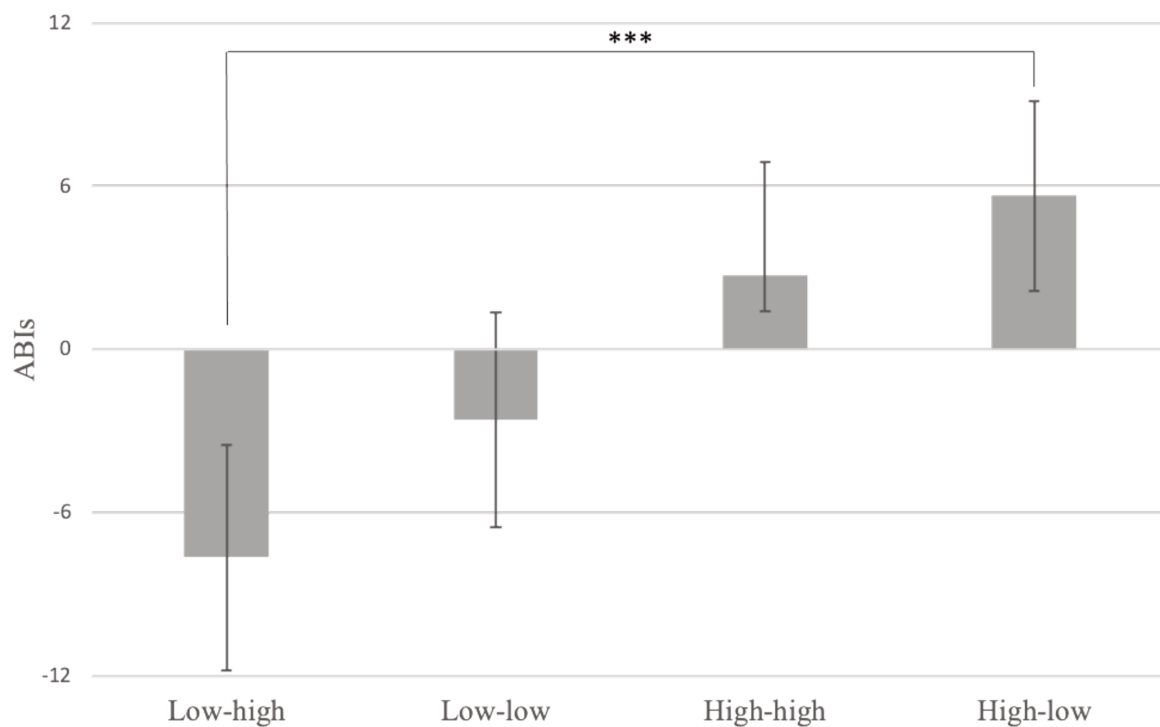
**Figure 1**

*Schematic Overview of the Experimental Design. The first and second rating tasks illustrate an example of how a question was presented and how participants had to respond. The three screens of the dot-probe task illustrate each critical event of each trial and its corresponding presentation time: The fixation cross with the two empty squares, the squares with the cues, and the empty squares with the dot.*



**Figure 2**

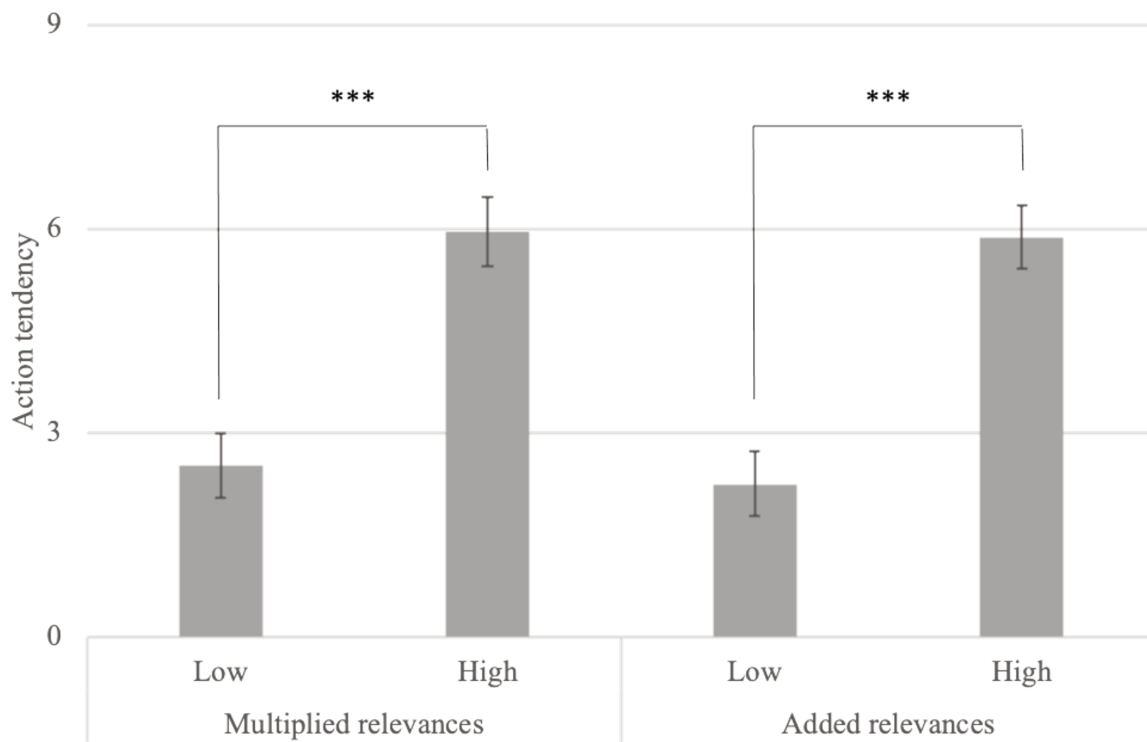
*ABIs as a function of the patterns of multiplied relevances reported by participants in the first rating task for stimuli of the two terms of couples in the dot probe task. Low-high patterns are defined by below-average rating of the first term and above or equal to average rating of the second term of the couples. High-low patterns are defined by above or equal to average rating of the first term and below-average rating of the second term of the couples. High-high patterns are defined by above or equal to average rating of the first term and above-average rating of the second term of the couples, and Low-low patterns are defined by below or equal to average rating of the first term and below-average rating of the second term of the couples.*



*Note.* †  $p < .10$ . \*  $p < .05$ . \*\*  $p < .01$ . \*\*\*  $p < .001$ ; Vertical lines depict 95% CI.

**Figure 3**

Action tendency values as a function of the patterns of multiplied relevances and added relevances reported by participants for stimuli in the first rating task. Low patterns are defined by below or equal to average rating of stimuli and High patterns are defined by above-average rating of stimuli.



Note. †  $p < .10$ . \*  $p < .05$ . \*\*  $p < .01$ . \*\*\*  $p < .001$ ; Vertical lines depict 95% CI.

## **Chapitre 3 - Article : “From emotion to action tendency: Time course of attentional effects of agonistic and antagonistic overlay of intrinsic and goal relevances”**

Ce chapitre correspond au deuxième article que nous avons réalisé, qui est actuellement en révision dans une revue scientifique dont la référence est la suivante :

Fournier, H. & Koenig, O. (Under review). From emotion to action tendency: Time course of attentional effects of agonistic and antagonistic overlay of intrinsic and goal relevances.

## Résumé

Au cours de la dernière décennie, la littérature sur l'attention émotionnelle a connu des avancées majeures. Il a été suggéré que la nature agréable (i.e., la pertinence intrinsèque) et la conductivité au but (i.e., la pertinence au but) des stimuli attirent l'attention. Récemment, il a été démontré que la superposition de la pertinence intrinsèque et de la pertinence au but augmentait le biais attentionnel (Fournier & Koenig, 2023). Cependant, ce résultat n'a été testé que lorsque la superposition était agonistique, c'est-à-dire lorsque les deux pertinences génèrent une tendance à l'approche parce que les stimuli sont agréables et propices à l'atteinte d'un objectif. Le problème se pose lorsque la superposition est antagoniste, par exemple lorsqu'un stimulus génère à la fois des tendances à l'approche et à l'évitement parce qu'il est à la fois désagréable et propice à l'atteinte du but. Nous avons émis l'hypothèse que la superposition antagoniste et agoniste de la pertinence intrinsèque et de la pertinence au but augmenterait de la même manière le biais attentionnel. En outre, les biais attentionnels précoces et tardifs sont supposés être davantage influencés par la pertinence intrinsèque et la pertinence au but, respectivement. Pour tester ces hypothèses, les participants ont effectué une tâche d'induction, visant à induire plusieurs types de pertinence pour les stimuli. Parallèlement, ils ont effectué une tâche de dot-probe, dans laquelle les stimuli étaient en concurrence en utilisant différents SOA, afin de mesurer les biais attentionnels précoces et tardifs. Nos résultats ont montré que la superposition agonistique et antagoniste de la pertinence intrinsèque et de la pertinence au but augmentait de la même manière le biais attentionnel. Cependant, nos résultats sont mitigés en ce qui concerne les effets de la pertinence intrinsèque et de la pertinence au but sur le biais attentionnel en fonction de la temporalité. Enfin, les tendances à l'approche et à l'évitement rapportées à propos des stimuli ont été mieux prédites lorsque leur caractère agréable, désagréable et la conductivité au but ont été pris en compte simultanément.




**From emotion to action tendency: Time course of attentional effects of agonistic and antagonistic overlay of intrinsic and goal relevances**

Hippolyte Fournier and Olivier Koenig

Laboratoire EMC, Université Lumière Lyon 2, Bron, France

EMC, Lyon 2 University, Bron, France

**Author note**

Hippolyte Fournier  <https://orcid.org/0000-0002-7861-2387>

Olivier Koenig  <https://orcid.org/0000-0002-5124-1889>

We have no conflicts of interest to disclose. We warrant that participants in this experiment were treated in accordance with the ethical standards of APA. This research was made possible thanks to a doctoral grant to Hippolyte Fournier from the Banque Publique d'Investissement (BPI) under the grant agreement THERADIA. Thanks to Mathilde Gaujard, Camille Roullet, Anaïs Quossi, Carole Allemandet, and Kenza Merland for their help in collecting the data.

All data, analysis code and material from the study reported in this article are available through the Open Science Framework repository: <https://osf.io/rnvpg/>.

Correspondence should be addressed to Hippolyte Fournier, Laboratoire EMC, Université Lumière Lyon 2, 5 Avenue Pierre Mendès France, Bron 69676, France. Email:

[h.fournier@univ-lyon2.fr](mailto:h.fournier@univ-lyon2.fr)

### Abstract

Over the past decade, there have been major advances in the literature on emotional attention. It has been suggested that both pleasantness (i.e., intrinsic relevance) and goal conduciveness (i.e., goal relevance) nature of stimuli capture attention. Recently, it has been shown that the overlay of intrinsic relevance and goal relevance increased attentional bias (Fournier & Koenig, 2023). However, this result was only tested when the overlay was agonistic, i.e., when both relevances generate approach tendency because stimuli are pleasant and goal conducive. The issue arises when the overlay is antagonistic, for instance when a stimulus generates both approach and avoidance tendencies by being both unpleasant and goal conducive. We hypothesized that antagonistic and agonistic overlay of intrinsic relevance and goal relevance would equally increase the attentional bias. Furthermore, early and late attentional biases were hypothesized to be driven more by intrinsic relevance and goal relevance, respectively. To test these hypotheses, participants performed an induction task, aimed at inducing several types of relevance to stimuli. Parallely, they performed a dot-probe task, in which stimuli competed using different SOA, to measure early and late attentional biases. Our results showed that agonistic and antagonistic overlay of intrinsic relevance and goal relevance equally increased the attentional bias. However, our results were mixed regarding the effects of intrinsic relevance and goal relevance on the attentional bias as a function of temporality. Finally, the approach and avoidance tendencies reported about stimuli were best predicted when their pleasantness, unpleasantness, and goal conduciveness were considered together.

*Keywords:* emotion, attention, appraisal theory, intrinsic relevance, goal relevance

**From emotion to action tendency: Time course of attentional effects, of agonistic and antagonistic overlay of intrinsic and goal relevances**

In a complex and dynamic world, attention allows individuals to interact with their environment appropriately, allocating cognitive resources to relevant information. Naturally, the question of the rules governing the hierarchy of information according to its relevance has given rise to much debate in the literature. Some researchers first proposed that the top of the hierarchy should be occupied by negative information (Öhman et al., 2001). For instance, negative stimuli such as snakes and spiders would capture attention because of their importance for survival (Fox et al., 2007; Brosch & Sharma, 2005; LoBue & DeLoache, 2008; Öhman, 1993; Öhman & Mineka, 2001). Later, it was shown that positive information also captures attention, such as erotic stimuli (Fromberger et al., 2012; Spiering & Everaerd, 2007; Steimke et al., 2017), baby faces (Brosch et al., 2007; Brosch et al., 2008) and happy faces (de Jong et al., 2007; Joormann & Gotlib, 2007). Another body of research proposed that both threatening and erotic information capture attention because such stimuli are associated with high arousal, and thus attention would prioritize information according to its arousal level (Anderson, 2005; Bradley et al., 2001; Lang, 1995). Knowing that arousal should likely be a consequence and not a cause of attentional capture, Schimmack (2005) proposed to consider stimulus-generated arousal as a marker of cognitive appraisal of stimulus relevance. In other words, attention would be captured based on the cognitive appraisal of stimulus relevance, and the arousal generated would reflect the depth of processing allocated to the stimulus according to its relevance. Currently, this statement is rather consensual in the literature (Brosch et al., 2010).

According to appraisal theories, the appraisal of relevance would be the trigger of an emotional episode, allowing to prepare individuals for the most appropriate behavior in response to likely important stimuli to their well-being and survival (Scherer, 2009, 2010).

Roughly speaking, stimuli appraised as negative will capture attention and will likely generate avoidance tendency, whereas stimuli appraised as positive will capture attention and will likely generate approach tendency. To this end, the appraisal of relevance would be divided into two sub-criteria that would determine whether the stimuli are rather negative or positive (Grandjean & Scherer, 2014). On the one hand, intrinsic relevance would appraise whether stimuli are rather unpleasant (likely generating avoidance tendency), or pleasant (likely generating approach tendency) in general, and this regardless of the state of the organism. On the other hand, goal relevance would appraise whether stimuli are currently goal obstructive (likely generating avoidance tendency) or goal conducive (likely generating approach tendency), which is important for the well-being or survival of the individual. Thus, the notion of intrinsic relevance would account for attentional capture toward stimuli associated with unpleasantness, such as snakes and spiders (LoBue & DeLoache, 2008; Öhman, 1993; Öhman & Mineka, 2001), or pleasantness, such as erotic stimuli (Fromberger et al., 2012; Spiering & Everaerd, 2007; Steimke et al., 2017), baby faces (Brosch et al., 2007; Brosch et al., 2008) and happy faces (de Jong et al., 2007; Joormann & Gotlib, 2007). In contrast, the notion of goal relevance would account for attentional capture toward stimuli associated with goal obstructiveness, such as loss-related stimuli in a game context (Müller et al., 2016; Sun et al., 2017), and goal conduciveness, such as water bottles for thirsty individuals (Mazzietti et al., 2014), clean-related stimuli after a disgust induction (Mazzietti et al., 2015; Vogt et al., 2011) and safety-related stimuli in a threat context (Vogt et al., 2017).

Although much attention has been given in the literature to the study of intrinsic relevance and goal relevance, most work has studied these two types of relevance separately. Indeed, few studies have examined the combined effects of intrinsic relevance and goal relevance. Yet, this question is crucial for designing the multilevel architecture of

appraisal criteria that form and differentiate emotional episodes. Among these studies, Aue & Scherer (2011) showed that intrinsic relevance and goal relevance have multiplied effects on somatovisceral responses. In a former study (Fournier & Koenig, 2023), we found the same evidence on attentional bias. More specifically, our findings evidenced that attentional capture was maximized when intrinsic relevance and goal relevance were aligned (i.e., if stimuli were appraised as both pleasant and goal conducive). This would reflect that attentional capture is maximized because both relevances generate approach tendencies. Indeed, in line with this interpretation, we showed that the approach tendency toward stimuli was best predicted by considering both pleasantness and goal conduciveness associated with them. That said, in this study, the combined effects of intrinsic relevance and goal relevance on attentional bias and action tendency were only tested in the case of an agonistic relevance overlay, when both relevances generated approach tendency by being pleasant and goal conducive. However, in many cases, relevance overlay can be antagonistic, for example, if stimuli are both unpleasant and goal conducive, generating both approach and avoidance tendencies. We can cite the so-called Miller experiment (1944) in which starving rats were confronted with food positioned at a location where electric shocks were delivered. One might ask whether an antagonistic stimulus, i.e., food associated with an electric shock, would capture attention in the same way as an agonistic stimulus. It would be conceivable that an antagonistic stimulus does not capture attention, since a stimulus that must be both approached because of its goal conduciveness and avoided because of its unpleasantness ultimately requires no action. However, the rats in the Miller experiment oscillated between avoidance and approach tendencies. These oscillations in action tendency would reflect a loop of reappraisals in which the stimulus was appraised as to be approached, then as to be avoided, etc... In any case, although a definitive appraisal struggled to emerge, the rats paid attention to the antagonistic stimulus

during this processing. Thus, it seems more conceivable to assume that antagonistic stimuli would capture attention in the same way as agonistic stimuli, to allow for further appraisal before deciding on the action to be taken.

Concerning the effects of intrinsic relevance and goal relevance on attention, another question arises regarding temporality. Although not articulated within the appraisal framework, several pieces of literature evidenced that early and late attentional biases would be driven more by intrinsic relevance (e.g., affective preferences) and goal relevance (e.g., craving), respectively. For instance, more frequent initial gaze orientations toward high-fat food were evidenced for overweight individuals (i.e., usually consumed this type of food) compared to healthy-weight individuals, while craving and satiety were controlled (Castellanos et al., 2009; Nijs et al., 2010; Werthmann et al. 2011). However, maintaining gaze over time was positively correlated with craving. Similar findings were found regarding addictive disorders such as nicotine and alcohol addictions (Field & Cox, 2008). Furthermore, Forrest et al., (2022) showed that goal-conducive stimuli captured more attention than unpleasant stimuli and that the attentional bias toward goal-conducive stimulus was greater when the Stimulus Onset Asynchrony (SOA) was longer (250ms vs. 30ms). According to appraisal theories, such observations may reflect a difference in appraisal processing time between the two types of relevance. Goal relevance appraisal would be time-consuming due to the compilation of two pieces of information, namely the current importance of a goal and the goal conduciveness of the stimulus. In contrast, intrinsic relevance appraisal would occur quickly (Aue et al., 2007; Grandjean & Scherer, 2008; Lanctôt & Hess, 2007), because it requires little information, namely the pleasantness of the stimulus only (Scherer 2013). In other words, while goal relevance is still being appraised, intrinsic relevance would be already available to approximate the relevance level of the stimulus. In terms of impact on the attentional bias, the intrinsic relevance impact

would be larger in the combined effects of relevance at an early temporality because goal relevance would still be under appraisal. The more time passes, the more complete the appraisal of the goal relevance is, and the larger its impact in the combined effects of relevances on the attentional bias (see Fournier & Koenig, 2023 for a discussion).

Let's take an example of the attentional competition between two stimuli among three: a neutral stimulus, a goal-relevant stimulus, and an overlay stimulus (with agonistic or antagonistic overlay), as a function of early and later temporality. First, we assume that in the pair contrasting the goal-relevant stimulus to the neutral stimulus, the attentional bias toward the goal-relevant stimulus would be greater in the later than the early temporality. This would reflect the larger effect of goal relevance on the attentional bias when the appraisal is completed. Second, we assume that in the pair contrasting the overlay stimulus to the neutral stimulus, the attentional bias toward the overlay stimulus would be greater in the later than the early temporality. This would reflect the larger effect of goal relevance in the combined effects of relevances on the attentional bias when the appraisal is completed. Third, we assume that in the pair contrasting the overlay stimulus to the goal-relevant stimulus, the attentional bias toward the overlay stimulus would be greater in the early temporality. This would reflect the larger effect of intrinsic relevance at the early temporality because goal relevance appraisal is not yet completed both for the overlay stimulus and the goal-relevant stimulus.

In the present experiment, we proposed to test whether the agonistic and antagonistic overlay of intrinsic relevance and goal relevance of stimuli increased the attentional bias compared to goal relevance alone. Second, we proposed to test whether the early attentional bias was driven more by intrinsic relevance and whether the late attentional bias was driven more by goal relevance.

To test these hypotheses, two groups of participants performed an induction task aiming to induce several relevances to three stimuli (i.e., three colored squares). The induction task functioned as a Go/No-Go task presented as a game (see, Vogt et al., 2013). The relevance inductions were a goal conduciveness induction, consisting of associating a stimulus with rewards, an unpleasantness induction, consisting of associating a stimulus with an unpleasant noise, and a pleasantness induction consisting of associating a stimulus with a pleasant noise. For the two groups of participants, one colored square was induced to the goal conduciveness, defined as the “goal-relevant stimulus”, and another was not induced to relevance, defined as the “neutral stimulus”. The difference between the two groups concerned the relevance induction applied to the third colored square, defined as the “overlay stimulus”. For the “agonistic group”, the overlay stimulus was induced to an overlay of agonistic relevances, namely, the goal conduciveness induction and the pleasantness induction. For the “antagonistic group”, the overlay stimulus was induced to an overlay of antagonistic relevances, namely, the goal conduciveness induction and the unpleasantness induction.

Parallel to the induction task, participants performed a dot-probe task in which the three stimuli were put in competition to measure attentional biases. The task was divided into two blocks, one with a short SOA, and one with a long SOA, to measure the impact of relevance on early and late attentional biases.

We hypothesized that the goal-relevant stimulus captured more attention than the neutral stimulus. Moreover, we hypothesized that the overlay stimulus captured more attention than other stimuli, and this, whether the overlay was agonistic or antagonistic. In addition, we expected that the attentional bias toward the goal-relevant stimulus and the overlay stimulus within the pairs contrasting them to the neutral stimulus was greater in the long SOA block than in the short SOA block. Eventually, we hypothesized that the attentional bias toward



the overlay stimulus within the pair contrasting it to the goal-relevant stimulus was greater in the short SOA block compared to the long SOA block.

A rating task was proposed to participants in which they were asked to appraise the level of intrinsic relevance (i.e., unpleasantness and pleasantness), the goal relevance (i.e., goal conduciveness), and the action tendencies (i.e., approach and avoidance) of the three stimuli. The rating task was first used as a material check. It was also used as an extension of our previous work (Fournier & Koenig, 2023) showing that the approach tendency level of stimuli was best predicted by taking into account both their pleasantness level and their goal conduciveness level. Here, we tested this hypothesis again taking into account the unpleasantness level, assuming that the approach tendency level of stimuli would be decreased by their unpleasantness level. Further, we tested the hypothesis that the avoidance tendency level of stimuli was best predicted by taking into account their level of pleasantness, goal conduciveness, and unpleasantness. It was assumed that the avoidance tendency level of stimuli would be decreased by their level of pleasantness and goal conduciveness and increased by their level of unpleasantness.

## **Method**

### **Participants**

One hundred and twenty-seven students (96 women;  $M = 20.22$  years,  $SD = 3.45$ ) from Lyon University participated in the study. Participants had normal or corrected-to-normal vision. The justification for the sample size was described in the Bayesian analysis section.

### **Materials**

The experiment was run on a 21" iMac in a soundproofed room using SuperLab 6 software (Haxby et al., 1991). "Sennheiser HD 219" headphones were used to deliver the sounds throughout the experiment. A French AZERTY keyboard was used to record participants' responses.

## Stimuli

Six colored squares (180\*180 pixels) were constructed using Affinity Photo. The luminance of each square was equalized to 34.5 cd/m<sup>2</sup>. Luminance measurement was carried out using the “Konica Minolta LS-110 luminance meter”. Three colored squares corresponded to the main stimuli: a blue, a pink, and an orange square, while the other three corresponded to the filler stimuli displayed in the induction task: a green, a yellow, and a red square. Two 650ms noises were used, an unpleasant and a pleasant one. The unpleasant sound (ID: 0360) was taken from the International Affective Digitized Sounds-Expanded (IADS-E; Yang et al., 2018), rated as highly negative ( $M = 1.92$ ,  $SD = 1.21$ ) according to the Self-Assessment Manikin (Bradley & Lang, 1994). This sound was delivered through the headphones at an intensity of 90 dBA so that the sound was unpleasant, but not harmful to physical integrity (Koster et al., 2004; Hobbs, 1990). The pleasant sound was taken from the database of sound effects of the video game "Zelda: Breath Of The Wild", snipped from the open-source database VG Resource. This sound was delivered through the headphones at an intensity of 80 dBA.

A rating task was designed so that 15 questions for each participant were asked about the three stimuli of the main set. Each stimulus appeared 5 times: accompanied by a question about its level of unpleasantness, pleasantness, goal conduciveness, avoidance tendency, or approach tendency. The questions regarding the level of unpleasantness and pleasantness were “How unpleasant do you find this square in general?” and “How pleasant do you find this square in general?”. The question regarding the level of goal conduciveness was “How much do you feel that this square could make you win?”. The questions regarding the level of avoidance tendency and approach tendency were “How much do you want this square not to appear on the screen?”, and “How much do you want this square to appear on the screen?”, respectively. In addition, one question was asked about the desire to win, “How much do you want to win?”, and two questions were asked about the level of unpleasantness and pleasantness of the noises “In

general, how unpleasant is this noise for you?” and “In general, how pleasant is this noise for you?”, respectively.

### **Design and procedure**

The rating task, the dot-probe task, and the induction task were presented on a black background on the computer screen.

#### ***Rating task design***

The rating task was proposed three times: once at the beginning, once in the middle, and once at the end of the experiment. The two questions about the level of unpleasantness and pleasantness of the noise were asked only once at the end of the experiment. Responses were given on a quantitative nine-point scale, from 1 (*not at all*) to 9 (*extremely*) using the keyboard keys 1 to 9.

#### ***Induction and dot-probe task design***

The dot-probe task corresponded to the successive appearance of four events. First, a white fixation cross (16\*16 pixels) with two empty squares with white outlines (400\*400 pixels) whose centers were at 250 pixels to the left and the right of the cross appeared during a random variable time (from 500 to 1000ms). Second, two stimuli appeared simultaneously in the center of each square during 100ms for the short SOA block or 250ms for the longer SOA block. Third, a white dot (26\*26 pixels) appeared during 100ms in the center of one of the two squares, immediately after the stimulus offset. Fourth, after the disappearance of the dot, responses required to indicate whether the dot appeared in the left square by pressing the “s” key or the right square by pressing the “ù” key.

The induction task consisted of the successive appearance of four events: First, one of the six stimuli from the main or the filler set appeared in a square in the center of the screen for 250ms. Second, the stimulus was replaced by a red question mark in the center of the screen that remained for 1,500ms or until the space bar was pressed. Third, a feedback screen appeared

for 650ms indicating whether the response was correct or not and whether points were gained or lost. Fourth, a final screen indicated the current score of the participant during 500ms.

### ***Procedure***

First, a consent form was completed. Then, participants were randomly assigned to the agonistic or the antagonistic group, to one of the two block orders (100ms SOA first or 250ms SOA first), and one of the six possible combinations of association between the three stimuli of the main set (blue, pink, and orange square) and the three stimulus types (overlay, goal-relevant, and neutral). All these variables were counterbalanced with an equivalent number of participants assigned to each possible combination.

Participants were asked to sit at sixty centimeters from the screen and to put the headphones over their ears. Then, they were asked to place their left index on the "s" key, their right index on the "ù" key, and one of their thumbs on the space bar of the computer keyboard. They were informed that the top five scores would win a FNAC gift certificate of 50€, 40€, 30€, 20€, and 10€, respectively. Then, the instructions were presented and divided into three phases.

First, the induction task was presented alone. The associations between the three stimuli of the main set and their assigned relevance induction type were explained to the participants. It was explained that when one of the stimuli of the main set appeared in the induction task, the space bar should be pressed when the red question mark appeared. If the space bar was pressed after the overlay stimulus appeared, the feedback screen informed a gain of 20 points, and a noise was produced (i.e., unpleasant for the antagonistic group and pleasant for the agonistic group). If the space bar was pressed after the goal-relevant stimulus appeared, the feedback screen informed a gain of 20 points. If the space bar was pressed after the neutral stimulus appeared, the feedback screen only informed that the response was correct. If the space bar was not pressed after a stimulus of the main set appeared, the feedback screen informed that the response was missed. If the space bar was pressed after a stimulus of the filler set appeared, the

feedback screen informed a loss of 20 points. Then, participants were asked to recall the roles assigned to the stimuli and to perform at least 20 trials without error.

Second, the dot-probe task was presented alone. Participants were asked to stare at the fixation cross throughout the whole experiment and to respond as quickly and precisely as possible (Chica et al., 2014) by pressing the “s” key if the dot appeared in the left square, and with the “ù” key if it appeared in the right square. The three possible pairs among the three stimuli of the main set could appear (the overlay stimulus vs. the goal-relevant stimulus, the overlay stimulus vs. the neutral stimulus, and the goal-relevant stimulus vs. the neutral stimulus). Once the instructions had been presented, participants performed twelve practice trials until no errors were made. It was explained to motivate participants that good performance in the dot-probe task increased the probability that winning stimuli would appear in the induction task.

Finally, participants were explained that a full trial of the experiment consisted of the succession of a trial of the dot-probe task and a trial of the induction task, as illustrated in Figure 1. After they were asked again to recall the relevance induction types assigned to the stimuli, participants performed twelve whole training trials of the experiment. If everything was understood correctly, the experimenter turned off the light and left the room.

The entire experiment consisted of the completion of the initial rating task, the first block of the dot-probe and induction tasks, the intermediate rating task, the second block of the dot-probe and induction tasks, and the final rating task. Each block contained 144 randomized trials. In the dot-probe task, each of the three possible pairs was presented 24 times with a given stimulus in the left square, and 24 times with the same stimulus in the right square. In half of these trials, the dot appeared on the left square, and in the other half on the right square. In the induction task, each of the six stimuli among the main set or the filler set appeared 24 times. Thus, the experiment consisted of a total of 288 trials.

At the end of the experiment, participants were asked if they had worn the headphones for the duration of the task and if they remembered the stimuli and their role. Finally, they were told that they would be contacted again if they won one of the FNAC gift certificates.

### **Bayesian Analysis**

Two points motivated the use of Bayesian analyses: The first point concerns the information richness compared to the Null Hypothesis Significance Test (NHST; Kruschke, 2013; Kruschke 2014). Indeed, obtaining a credibility level of evidence via a measure of the hypothesis likelihood, such as the Bayes Factor (BF), is more informative than the binary acceptance/rejection decision based on an arbitrary threshold such as p-value. Second, the Bayesian analysis allows us to support the null hypothesis rather than not being able to reject it (such as in the case of the NHST). Specifically, Bayesian models allow measuring the level of evidence in favor or against whether data are better predicted as a function of the inclusion of a given effect as a predictor. The interpretation of the BF is quite simple, even for a naive audience.  $BF_{10}$ , the BF providing the likelihood ratio in favor of the hypothesis that the data are better predicted by including a given effect as a predictor, can be interpreted as follows:  $BF_{10} \geq 3$  highlights moderate evidence,  $BF_{10} \geq 10$  highlights strong evidence, and  $BF_{10} \geq 100$  highlights decisive evidence (Kass & Raftery, 1995).  $BF_{01}$ , the BF providing the likelihood ratio in favor of the hypothesis that the data are better predicted by not including a given effect as a predictor, can be interpreted in the same way:  $BF_{01} \geq 3$  highlights moderate evidence,  $BF_{01} \geq 10$  highlights strong evidence, and  $BF_{01} \geq 100$  highlights decisive evidence.

The sample size was determined using the sequential Bayes Factor Design Analysis (Schönbrodt & Wagenmakers, 2018). The idea of such a design is to first determine a minimum number of participants ( $n_{\min}$ ) needed to correctly detect the expected effects, knowing a given BF threshold. Then, starting from the  $n_{\min}$ , the inclusion of participants is performed until all BFs of expected effects exceed the threshold, or until the defined maximum number of

participants ( $n_{\max}$ ) is reached. To guarantee the proper practice of the analysis, we followed the rules used in the step-by-step example proposed by Stefan et al. (2019) by defining the BF threshold as 6. The  $n_{\min}$  was set to 60 and the  $n_{\max}$  to 120. Since the BF of one expected effect did not exceed the threshold until the  $n_{\max}$ , the inclusion reached 120 participants and then stopped.

### **Transparency and Openness**

The level 2 requirement of the Transparency and Openness Guidelines (Nosek et al., 2015) was met for this article. The Journal Article Reporting Standards (Kazak, 2019) were followed. The sample size analysis, data exclusion, manipulation, and measurement were reported. Data, pre-treatment, code analyses, and material from the study are available through the Open Science Framework repository: <https://osf.io/rnvpg/>. This study was not preregistered. Data were analyzed using R (version 1.4): Models were fitted with 100,000 iterations using the package *BayesFactor* version 0.9.12-4.4 (Morey & Rouder, 2022) and the package *brms* (Bürkner, 2021). Default priors of the *brms* and *BayesFactor* packages were used (Rouder et al., 2012). The 95% CI of the standardized coefficients was computed using the package *bayestestR* version 0.9.0 (Makowski et al., 2019).

### **Data processing**

Seven participants were eliminated. Three reported removing the headphones from their ears, and the remaining four were identified as not being sufficiently engaged in the task. This identification was based on the error threshold of more than 5% error in the dot-probe task. For the remaining participants, the incorrect answers were discarded (223 of 34560 data). Then, RTs reflecting anticipatory and delayed responses (shorter than 150ms and longer than 750ms, respectively; Carlson & Fang, 2020) were discarded (925 of 34337 data).

In the dot-probe task, trials were divided into valid and invalid trials. Valid trials are generally defined by trials in which the dot replaced the supposedly most relevant stimulus of

a pair, while invalid trials are generally defined by trials in which the dot replaced the supposedly least relevant stimulus of the pair. Shorter RTs for valid than for invalid trials reflect the validation of the hypothesis that the supposedly most relevant stimulus of the pair captured more attention than the supposedly least relevant stimulus of the pair. Thus, in the present experiment, valid trials were defined by trials in which the dot replaced the overlay stimulus (in the overlay vs. goal-relevant stimulus pair and in the overlay vs. neutral stimulus pair), and the goal-relevant stimulus (in the goal-relevant vs. neutral stimulus pair).

In addition, Attentional Bias Indexes (ABI; Macleod et al., 1986) were calculated to measure the attentional bias of the pair's first and second stimulus. For each SOA condition of each pair of each participant, ABIs were calculated by subtracting the mean RTs of valid trials from the mean RTs of invalid trials. Positive ABI means that the first stimulus captured more attention than the second stimulus.

### **Data analysis plan**

A control was carried out from the three rating task data to check the reliability of the relevance inductions that we manipulated. First, it implied testing whether participants were sufficiently involved in the task. Second, it implied testing whether the noise assigned to the agonistic group was pleasant and that assigned to the antagonistic group was unpleasant. Then, it implied testing whether the overlay stimulus and the goal-relevant stimulus were more goal conducive than the neutral stimulus, to define them as goal-relevant. Finally, it implied to test whether the overlay stimulus was more pleasant for the agonistic group and more unpleasant for the antagonistic group than the goal-relevant stimulus and the neutral stimulus. This allowed us to define them as also intrinsically relevant (i.e., pleasant for the agonistic group, and unpleasant for the antagonistic group). To this end, Bayesian one-tailed t-tests were performed, to obtain the  $BF_{10}$  of the hypothesis that the effect size of one sample (or between two samples) was positive rather than null or negative.



To test our hypotheses on attentional biases, a Bayesian ANOVA (BANOVA) was performed on RTs with the main and interaction effects between the factors Validity (valid, invalid), Pair type (goal-relevant stimulus vs. neutral stimulus, overlay stimulus vs. goal-relevant stimulus, and overlay stimulus vs. neutral stimulus), SOA (100ms, 250ms), and Group (agonist, antagonist). Participant, Block order, and Square color were entered as random effects. Then, the data of the three pairs were entered separately in three BANOVAs containing the same fixed and random effects as the full BANOVA (except for the Pair type) to verify the consistency of effects within each pair.

To test the first hypothesis that the goal-relevant stimulus captured more attention than the neutral stimulus and that the overlay stimulus captured more attention than other stimuli, we tested for the presence of a Validity main effect in the full BANOVA. In case of a Validity  $\times$  Pair Type interaction, the consistency of the simple effect of Validity within each pair type would be checked in three separate BANOVAs. In each case, shorter RTs for valid than invalid trials would be expected. In addition, to test whether the overlay stimulus captured attention in the same way whether the overlay was agonistic or antagonistic, we expected no Validity  $\times$  Group interaction in the full BANOVA and the three separate BANOVAs related to each pair.

To test the second hypothesis that early and late attentional biases were driven more by intrinsic relevance and goal relevance, respectively, we tested for an interaction of the factors Validity  $\times$  Pair type  $\times$  SOA in the full BANOVA. Regarding the three separate BANOVAs related to each pair, we expected to observe a Validity  $\times$  SOA interaction modeled in different forms: Within the pair contrasting the goal-relevant stimulus vs. the neutral stimulus and the pair contrasting the overlay stimulus vs. the neutral stimulus, greater differences between valid and invalid RTs were expected in the 250ms than in the 100ms SOA condition. In addition, within the pair contrasting the overlay stimulus vs. the goal-relevant stimulus, greater

differences between valid and invalid RTs were expected in the 100ms than in the 250ms SOA condition.

To test our hypotheses on action tendency, two separate Bayesian linear models were fitted on the level of avoidance tendency and approach tendency of stimuli reported from the rating task data. For each model, the level of unpleasantness, pleasantness, and goal conduciveness of stimuli was entered as predictors. Participant and Time of rating (i.e., beginning, intermediate, or final rating task) were entered as random effects.

The hypothesis that the level of pleasantness and goal conduciveness of stimuli decreased their avoidance tendency level, while their level of unpleasantness increased it, involved observing that the best-fit model included all predictors. More precisely, it involved that the level of pleasantness and goal conduciveness of stimuli negatively impacted their avoidance tendency level, while the level of unpleasantness of stimuli positively impacted their avoidance tendency level.

Similarly, the hypothesis that the level of pleasantness and goal conduciveness of stimuli increased their approach tendency level, while their level of unpleasantness decreased it, involved observing that the best-fit model included all predictors. More precisely, it involved that the level of pleasantness and goal conduciveness of stimuli positively impacted their approach tendency level, while the level of the unpleasantness of stimuli negatively impacted their approach tendency level.

## **Results**

The control from the three rating task data to check the reliability of the relevance inductions we manipulated was reported in the supplemental material to lighten the body of the main manuscript. This control showed that participants were correctly involved in the task and that the noises assigned to the agonistic and antagonistic groups were rated as pleasant and unpleasant, respectively. Regarding the relevance inductions performed on stimuli, both groups

rated the overlay stimulus and the goal-relevant stimulus as more goal conducive than the neutral stimulus. In addition, the overlay stimulus was rated as more pleasant for the agonistic group and more unpleasant for the antagonistic group than the goal-relevant stimulus and the neutral stimulus. In sum, these controls ensured the reliability of the relevance inductions that we manipulated.

### **Attentional bias analysis**

Regarding the hypothesis that the goal-relevant stimulus and the overlay stimulus captured more attention than the neutral stimulus, as expected, the full BANOVA revealed a Validity main effect,  $BF_{10} > 1000$ , as well as a Validity  $\times$  Pair type interaction,  $BF_{10} > 1000$ . This Validity main effect was consistently observed in the three separate BANOVAs performed on the data of each pair ( $BF_{10} = 30.54$  for the goal-relevant vs. neutral stimulus pair,  $BF_{10} > 1000$  for the overlay vs. goal-relevant stimulus pair, and  $BF_{10} > 1000$ , for the overlay vs. neutral stimulus pair).

Within each pair, RTs of valid trials were shorter than RTs of invalid trials. The mean RTs and standard errors as a function of pair type and validity are reported in Figure 2. These results suggest that the goal-relevant stimulus captured more attention than the neutral stimulus, and the overlay stimulus captured more attention than the goal-relevant stimulus and the neutral stimulus.

Regarding the comparison of agonistic and antagonistic overlay relevances, the full BANOVA revealed no interaction between Validity and Group,  $BF_{01} = 41.67$ . This absence of interaction was consistently observed in the three BANOVAs performed on each pair's data ( $BF_{01} = 17.54$  for the goal-relevant vs. neutral stimulus pair,  $BF_{01} = 20$  for the overlay vs. goal-relevant stimulus pair, and  $BF_{01} = 30.30$  for the overlay vs. neutral stimulus pair). The corresponding mean RTs and standard errors are reported in Figure 3. These results suggest that

the attentional bias previously evidenced within each pair is equivalent whether the relevance overlay was agonistic or antagonistic.

Regarding the hypothesis that early and late attentional biases were driven more by intrinsic relevance and goal relevance, respectively, the full BANOVA revealed no interaction between Validity, Pair type, and SOA,  $BF_{01} = 45.45$ . The absence of Validity  $\times$  SOA interaction was also observed in two of the three BANOVAs performed on the data of each pair ( $BF_{01} = 26.32$ , for the goal-relevant vs. neutral stimulus pair, and  $BF_{01} = 11.76$ , for the overlay vs. goal-relevant stimulus pair). Regarding the BANOVA performed on the data of the pair contrasting the overlay stimulus vs. the neutral stimulus, although moderate evidence of no Validity  $\times$  SOA interaction was observed, the BF did not exceed the threshold of 6,  $BF_{01} = 5.41$ . The corresponding mean RTs and standard errors are reported in Figure 4. These results suggest that the attentional bias previously evidenced within pairs contrasting the goal-relevant stimulus vs. the neutral stimulus, and the overlay stimulus vs. the goal-relevant stimulus is equivalent to whether attentional bias was early or late. In addition, although moderate evidence of attentional bias equivalence was found within the pair contrasting the overlay stimulus vs. the neutral stimulus as a function of early or late attentional bias, we could not conclude on this effect because the BF threshold was not reached.

To better understand why this effect was not conclusive, we decided to perform an exploratory analysis on the attentional bias of the pair contrasting the overlay stimulus vs. the neutral stimulus in the 100ms and 250ms SOA conditions. The BANOVA allowed answering whether the expected effect likely produced the data knowing the impact level of the other fixed and random effects entered in the model. To go further with a less restrictive approach, we decided to test whether the attentional bias of the pair contrasting the overlay stimulus vs. the neutral stimulus differed between 100ms and 250ms SOA conditions. To this end, a Bayesian t-test was performed on the ABI of this pair as a function of 100ms and 250ms SOA conditions.

Decisive evidence of greater ABI was observed for the pair contrasting the overlay stimulus vs. the neutral stimulus at 250ms SOA compared to 100ms SOA condition,  $BF_{10} > 1000$ . In sum, this exploratory analysis suggests that the attentional bias toward the overlay stimulus compared to the neutral stimulus was greater in the case of a 250ms than a 100ms SOA, although this difference was not strong enough to predict data.

### **Action tendency analysis**

Descriptive statistics and correlations are reported in Table S3.

Regarding the Bayesian linear model fitted on the level of avoidance tendency of stimuli, the best-fit model included all predictors,  $R^2 = .37$ . More precisely, the level of avoidance tendency of stimuli was negatively impacted by their level of pleasantness,  $BF_{10} > 1000$ ,  $\beta = 0.43$ , 95% CI [0.39, 0.48], and their level of goal conduciveness,  $BF_{10} = 91.45$ ,  $\beta = -0.12$ , 95% CI [-0.18, -0.05], whereas the level of avoidance tendency of stimuli was positively impacted by their unpleasantness level,  $BF_{10} > 1000$ ,  $\beta = 0.32$ , 95% CI [0.26, 0.38]. These results suggested that both pleasantness and goal conduciveness of stimuli increased approach tendency while unpleasantness of stimuli decreased it.

Regarding the Bayesian linear model fitted on the level of approach tendency of stimuli, the best-fit model included all predictors,  $R^2 = .65$ . More precisely, the level of approach tendency of stimuli was positively impacted by their level of pleasantness,  $BF_{10} > 1000$ ,  $\beta = -0.27$ , 95% CI [-0.34, -0.19], and their level of goal conduciveness,  $BF_{10} > 1000$ ,  $\beta = 0.37$ , 95% CI [0.33, 0.41], whereas the level of approach tendency of stimuli was negatively impacted by their unpleasantness level,  $BF_{10} > 1000$ ,  $\beta = -0.12$ , 95% CI [-0.16, -0.08]. These results suggested that both pleasantness and goal conduciveness of stimuli increased approach tendency while unpleasantness of stimuli decreased it.

## Discussion

Recently, it was shown that intrinsic relevance and goal relevance have combined effects on the attentional bias (Fournier & Koenig, 2023). Broadly speaking, it means that the attentional bias toward stimuli is increased if they are both intrinsically relevant and goal-relevant. However, this result was evidenced in the case the overlay of intrinsic relevance and goal relevance on stimuli was agonistic (i.e., both pleasant and goal-conducive stimuli). Our results revealed that the attentional bias is similarly enhanced if the overlay of intrinsic relevance and goal relevance is antagonistic (i.e., overlay of unpleasantness and goal conduciveness). Furthermore, a growing body of evidence suggests that intrinsic relevance induces an early attentional bias, whereas goal relevance induces a later attentional bias. Although our main results did not support this hypothesis, the results of our exploratory analysis were consistent with this prediction. Finally, our findings showed that both approach and avoidance tendencies related to stimuli are triggered by the appraisal of their intrinsic relevance (i.e., unpleasantness and pleasantness), and their goal relevance (i.e., goal conduciveness).

The operationalization of the hypothesis that the attentional bias is similarly increased by the agonistic and antagonistic overlay of intrinsic and goal relevance was related to several findings. First, we showed that the goal-relevant stimulus captured more attention than the neutral stimulus. This suggests that attention is captured by goal-relevant stimuli to indicate to individuals that the stimuli are likely to be approached to achieve well-being and survival (Scherer, 2009, 2010). This result was widely reported in the literature where we observed an attentional bias toward water bottles for thirsty individuals (Mazzietti et al., 2014), toward clean-related stimuli after a disgust induction (Mazzietti et al., 2015; Vogt et al., 2011), and toward safety-related stimuli in a threat context (Vogt et al., 2017). Second, we showed that the overlay stimulus captured more attention than the goal-relevant and

neutral stimuli, both for the agonistic and antagonistic groups. This means that overlaying intrinsic relevance on goal relevance increases the attentional bias toward stimuli, whether the overlay is agonistic or antagonistic. To discuss these results, it is necessary to further qualify what intrinsic relevance and goal relevance are.

Appraising the relevance of a stimulus would be like making a prediction (Barrett, 2017). More precisely, a probability distribution of the stimulus relevance would be simulated by Bayesian inference based on available prior beliefs about the stimulus characteristics (Seth & Friston, 2016). Concerning intrinsic relevance, it implies inferring the likelihood that the stimulus is more generally pleasant or unpleasant based on past experiences with the stimulus in general. Concerning goal relevance, it implies inferring the likelihood that the stimulus can conduce or obstruct the achievement of a currently important goal based on past experiences involving the stimulus in the context of this goal. The notion of prediction means that the relevance appraisal of a stimulus can be erroneous, which can be dramatic for the individual's well-being and survival (e.g., underestimating the goal conduciveness nature of a stimulus, such as the thirst-quenching nature of a cactus for a thirsty individual in the desert). To deal with this problem, the brain would follow an adaptative principle dictating to constantly seek to minimize the prediction error (Friston, 2005; 2010).

Regarding the increase in attentional bias by the agonistic overlay of intrinsic relevance and goal relevance, we suggest that this reflects a way for the brain to minimize prediction error. If a stimulus is appraised as both pleasant and goal conducive, the probability that the stimulus must be approached for well-being and survival is maximized because both relevances generate approach tendency (Fournier & Koenig, 2023).

Regarding the increase in attentional bias by the antagonistic overlay of intrinsic relevance and goal relevance, we suggest that the explanation differs somewhat. As explained above, the brain struggles to deal with prediction errors and constantly seeks to

minimize them (Friston, 2005; 2010). If a stimulus is appraised as both goal-conducive and unpleasant, it follows that the stimulus must probably be approached and avoided for well-being and survival, because the relevances generate both approach and avoidance tendencies. In this case, the prediction error is highly maximized because intrinsic relevance and goal relevance are in conflict. Thus, we suggest that an antagonistic overlay of intrinsic relevance and goal relevance captures attention to resolve this conflict through more in-depth appraisal, and ultimately minimizing prediction error.

Our results about relevance and action tendency shed light on their hypothesized relationship in appraisal theories (Grandjean & Scherer, 2014). More specifically, the approach tendency and the avoidance tendency of stimuli were best predicted by the conjunction of their pleasantness, goal conduciveness, and unpleasantness. On the one hand, the approach tendency of stimuli was increased by their pleasantness and their goal conduciveness and decreased by their unpleasantness. On the other hand, the avoidance tendency of stimuli was increased by their unpleasantness and decreased by their pleasantness and their goal conduciveness. These findings are in line with the hypothesis that approach tendency is generated by the pleasant nature of intrinsic relevance appraisal and the goal-conducive nature of goal relevance appraisal, while avoidance tendency is generated by the unpleasant nature of intrinsic relevance appraisal. According to appraisal theories (Grandjean & Scherer, 2014), avoidance tendency is also generated by the goal obstructive nature of goal relevance. However, the design of the present experiment did not include a goal obstructiveness induction, which prevented us from assessing the relationship between goal obstructiveness and action tendency. Moreover, our findings suggest that relevance may have inhibitory effects on action tendency. Indeed, approach tendency would be reduced by the unpleasant nature of intrinsic relevance appraisal, and avoidance tendency



would be reduced by the pleasant nature of intrinsic relevance appraisal and the goal conducive nature of goal relevance appraisal.

The above observations raise a critical point about the notion of valence, widely used in the literature on emotion (Arioli et al., 2021; Lewis et al., 2019; Ossenfort & Isaacowitz, 2021; Yeung & Fernandes, 2021). This notion was defined as the degree of repulsion or attraction of stimuli for individuals (Lewin, 1938/1951), and was later extended to the differentiation of emotions (Feldman Barrett & Russel, 1998; Kron et al., 2015; Posner et al., 2005). So, broadly speaking, valence could be defined by a continuum from avoidance tendency to approach tendency (Scherer, 2013). However, our results showed that stimuli can be associated with both approach and avoidance tendencies. Thus, considering valence as unidimensional is somewhat problematic for characterizing the emotional nature of stimuli. We emphasize the importance of considering valence as multidimensional, depending on the stimuli's intrinsic relevance and goal relevance (Scherer, 2013; Scherer & Moors, 2019; Shuman et al., 2013).

Concerning the hypothesis that early and late attentional biases would be driven more by intrinsic relevance and goal relevance, respectively, our results are mixed. Specifically, equivalence in the magnitude of attentional bias was found between 100ms and 250ms SOA conditions within the pairs contrasting the goal-relevant stimulus vs. the neutral stimulus and the overlay stimulus vs. the goal-relevant stimulus. On the one hand, this means that goal relevance alone did not have a greater impact on late attentional bias than on early attentional bias. On the other hand, this means that in the attentional bias between the overlay of intrinsic relevance and goal relevance vs. goal relevance alone, the relevance overlay did not have a greater impact on attention than goal relevance alone at early or later temporality. However, within the pair contrasting the overlay stimulus vs. the neutral stimulus, results did not allow to conclude in favor or against a difference between early

and late attentional bias. An exploratory analysis was conducted using a Bayesian t-test to further investigate these results. The Bayesian t-test answers whether the data likely differ across two modalities of an effect, while the BANOVA answers whether an effect likely produces data accounting for the impact level of the other effects entered in the model. Although the Bayesian t-test is not a robust method that can be used to firmly answer a hypothesis, it can provide clues about the potential impact of an effect on the data. The Bayesian t-test performed on the pair contrasting the overlay stimulus vs. the neutral stimulus showed decisive evidence of greater attentional bias toward the overlay stimulus at later than early temporality. We emphasize that this result is not inconsistent with that of the main analysis because the tests answer different questions and taken together, indicate that the effect is likely to exist, even though we failed to highlight it.

One reason we failed to highlight that early and late attentional bias would be driven more by intrinsic relevance and goal relevance, respectively, may be related to the fact that the SOA defined for the early temporality was too long. In a prior experiment (Fournier & Koenig, 2023), we assumed that in the combined effects of relevance on attention, the time window in which attentional bias would be driven more by intrinsic than goal relevance would be approximately 100ms. The short SOA of the present study was defined based on this statement. However, we did not account for the dot presentation time of 100ms. Thus, the early attentional bias that we measured probably already accounted for a significant weight of goal relevance because the processing time of the relevance appraisal was 200ms. Thus, we suggest that choosing a shorter SOA to assess early attentional bias, such as 30ms, as done by Forrest et al. (2022), would be better suited to compare the different impacts of intrinsic and goal relevance.

### Limitations

The current study presents two major limitations. First, the experimental design did not include a goal obstructiveness induction. This would have been interesting to assess the impact of an agonistic overlay of unpleasantness and goal obstructiveness and an antagonistic overlay of pleasantness and goal obstructiveness on attentional biases. Further, it would have been interesting to assess if approach and avoidance tendencies would be best predicted by accounting for goal obstructiveness in addition to pleasantness, unpleasantness, and goal conduciveness. The goal obstructiveness induction was not included because of the complexity of the experimental design instructions in its current state. An initial experimental design also included a goal obstructiveness induction. But participants struggled to retain the inductions assigned to the four colored squares, in addition to integrating the instructions of the dot-probe task and the induction task to be performed in parallel. Thus, we decided to remove the goal obstructiveness induction.

Second, a choice of 100ms SOA was defined to measure early attentional bias. As explained above, the early attentional bias we measured corresponded instead to the impact of relevance across a 200ms processing time window given the dot time presentation. We emphasize the importance of defining a shorter SOA to observe the impact of relevance on attentional biases.

### References

- Arioli, M., Basso, G., Poggi, P., & Canessa, N. (2021). Fronto-temporal brain activity and connectivity track implicit attention to positive and negative social words in a novel socio-emotional Stroop task. *NeuroImage*, 226, 117580.  
<https://doi.org/10.1016/j.neuroimage.2020.117580>
- Aue, T., Flykt, A., & Scherer, K. R. (2007). First evidence for differential and sequential efferent effects of stimulus relevance and goal conduciveness appraisal. *Biological Psychology*, 74(3), 347-357. <https://doi.org/10.1016/j.biopsycho.2006.09.001>

- Aue, T., & Scherer, K. R. (2011). Effects of intrinsic pleasantness and goal conduciveness appraisals on somatovisceral responding: Somewhat similar, but not identical. *Biological Psychology*, 86(1), 65-73. <https://doi.org/10.1016/j.biopsycho.2010.10.008>
- Barrett, L. F. (2017). The theory of constructed emotion: an active inference account of interoception and categorization. *Social cognitive and affective neuroscience*, 12(1), 1-23. <https://doi.org/10.1093/scan/nsx060>
- Bradley, M. M., Codispoti, M., Cuthbert, B. N., & Lang, P. J. (2001). Emotion and motivation I: defensive and appetitive reactions in picture processing. *Emotion*, 1(3), 276-298. <https://doi.org/10.1037//1528-3542.1.3.276>
- 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. [https://doi.org/10.1016/0005-7916\(94\)90063-9](https://doi.org/10.1016/0005-7916(94)90063-9)
- Brosch, T., Sander, D., & Scherer, K. R. (2007). That baby caught my eye... Attention capture by infant faces. *Emotion*, 7(3), 685–689. <https://doi.org/10.1037/1528-3542.7.3.685>
- Brosch, T., Sander, D., Pourtois, G., & Scherer, K. R. (2008). Beyond fear: Rapid spatial orienting toward positive emotional stimuli. *Psychological Science*, 19(4), 362-370. <https://doi.org/10.1111/j.1467-9280.2008.02094.x>
- Brosch, T., & Sharma, D. (2005). The Role of Fear-Relevant Stimuli in Visual Search: A Comparison of Phylogenetic and Ontogenetic Stimuli. *Emotion*, 5(3), 360–364. <https://doi.org/10.1037/1528-3542.5.3.360>
- Bürkner, P.C. (2021). Bayesian Item Response Modeling in R with brms and Stan. *Journal of Statistical Software*, 100(5), 1-54. <https://doi.org/10.18637/jss.v100.i05>
- Castellanos, E. H., Charboneau, E., Dietrich, M. S., Park, S., Bradley, B. P., Mogg, K., & Cowan, R. L. (2009). Obese adults have visual attention bias for food cue images: evidence

- for altered reward system function. *International Journal of Obesity*, 33(9), 1063-1073.  
<https://doi.org/10.1038/ijo.2009.138>
- Carlson, J. M., & Fang, L. (2020). The stability and reliability of attentional bias measures in the dot-probe task: Evidence from both traditional mean bias scores and trial-level bias scores. *Motivation and Emotion*, 44(5), 657-669. <https://doi.org/10.1007/s11031-020-09834-6>
- Chica, A. B., Martín-Arévalo, E., Botta, F., & Lupiáñez, J. (2014). The Spatial Orienting paradigm: How to design and interpret spatial attention experiments. *Neuroscience and Biobehavioral Reviews*, 40, 35–51. <https://doi.org/10.1016/j.neubiorev.2014.01.002>
- de Jong, P. J., Koster, E. H., van Wees, R., & Martens, S. (2009). Emotional facial expressions and the attentional blink: Attenuated blink for angry and happy faces irrespective of social anxiety. *Cognition and Emotion*, 23(8), 1640-1652.  
<https://doi.org/10.1080/02699930802490227>
- Feldman Barrett, L., & Russell, J. A. (1998). Independence and bipolarity in the structure of current affect. *Journal of Personality and Social Psychology*, 74(4), 967-984.  
<https://doi.org/10.1037/0022-3514.74.4.967>
- Fournier, H., & Koenig, O. (2023). Combined effects of intrinsic and goal relevances on attention and action tendency during the emotional episode. *Emotion*, 23(2), 425-436.  
<https://doi.org/10.1037/emo0001081>
- Fox, E., Griggs, L., & Mouchlianitis, E. (2007). The detection of fear-relevant stimuli: Are guns noticed as quickly as snakes? *Emotion*, 7(4), 691-696. <https://doi.org/10.1037/1528-3542.7.4.691>
- Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological sciences*, 360(1456), 815-836. <https://doi.org/10.1098/rstb.2005.1622>

- Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature reviews neuroscience*, *11*(2), 127-138. <https://doi.org/10.1038/nrn2787>
- Fromberger, P., Jordan, K., von Herder, J., Steinkrauss, H., Nemetschek, R., Stolpmann, G., & Müller, J. L. (2012). Initial orienting towards sexually relevant stimuli: Preliminary evidence from eye movement measures. *Archives of Sexual Behavior*, *41*(4), 919-928. <https://doi.org/10.1007/s10508-011-9816-3>
- Grandjean, D., Sander, D., & Scherer, K. R. (2008). Conscious emotional experience emerges as a function of multilevel, appraisal-driven response synchronization. *Consciousness and Cognition*, *17*(2), 484-495. <https://doi.org/10.1016/j.concog.2008.03.019>
- Grandjean, D., & Scherer, K. R. (2014). Chapter 2. Theory of cognitive appraisal and dynamics of emotional processes. In Sander, D. (Ed.). *Treatise on the Psychology of Emotions*, 51-87. <https://doi.org/10.3917/dunod.sande.2014.01.0051>
- Haxby, J. V., Parasuraman, R., Lalonde, F., & Abboud, H. (1993). SuperLab: General-purpose Macintosh software for human experimental psychology and psychological testing. *Behavior Research Methods, Instruments, & Computers*, *25*(3), 400-405. <https://doi.org/10.3758/BF03204531>
- Hobbs, R. J. (1990). Noise and vibration. In *Safety at work* (pp. 418-440). Butterworth-Heinemann. <https://doi.org/10.1016/B978-0-7506-1018-6.50031-1>
- Joormann, J., & Gotlib, I. H. (2007). Selective attention to emotional faces following recovery from depression. *Journal of Abnormal Psychology*, *116*(1), 80. <https://doi.org/10.1037/0021-843X.116.1.80>
- Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American statistical association*, *90*(430), 773-795. <https://doi.org/10.1080/01621459.1995.10476572>
- Kazak, A. E. (2019). Editorial: Journal article reporting standards. *American Psychologist*, *73*, 1-2. <http://dx.doi.org/10.1037/amp0000263>

- Koster, E. H. W., Crombez, G., Van Damme, S., Verschuere, B., & De Houwer, J. (2004). Does Imminent Threat Capture and Hold Attention? *Emotion, 4*(3), 312–317. <https://doi.org/10.1037/1528-3542.4.3.312>
- Kron, A., Pilkiw, M., Banaei, J., Goldstein, A., & Anderson, A. K. (2015). Are valence and arousal separable in emotional experience? *Emotion, 15*(1), 35-44. <http://dx.doi.org/10.1037/a0038474>
- Kruschke, J. K. (2013). Bayesian estimation supersedes the t-test. *Journal of Experimental Psychology: General, 142*(2), 573. <https://doi.org/10.1037/a0029146>
- Kruschke, J. (2014). Tools in the Trunk. In Kruschke, J. (Ed.). *Doing Bayesian Data Analysis* (2<sup>nd</sup> ed., pp. 721-736). Academic Press. <https://doi.org/10.1016/B978-0-12-405888-0.00025-8>
- Lanctôt, N., & Hess, U. (2007). The timing of appraisals. *Emotion, 7*(1), 207–212. <https://doi.org/10.1037/1528-3542.7.1.207>
- Lewin, K. (1951). *Field theory in social science: Selected theoretical papers* (D. Cartwright, Ed.). Greenwood Press. (Original work published 1938)
- Miller, N. E. (1944). Experimental studies of conflict. In Hunt, J. (Ed), *Personality and the behavior disorders*, 431-465. New York: Ronald Press.
- Lewis, E. J., Blanco, I., Raila, H., & Joormann, J. (2019). Does repetitive negative thinking affect attention? Differential effects of worry and rumination on attention to emotional stimuli. *Emotion, 19*(8), 1450-1462. <https://doi.org/10.1037/emo0000535>
- LoBue, V., & DeLoache, J. S. (2008). Detecting the snake in the grass: Attention to fear-relevant stimuli by adults and young children. *Psychological Science, 19*(3), 284-289. <https://doi.org/10.1111/j.1467-9280.2008.02081.x>

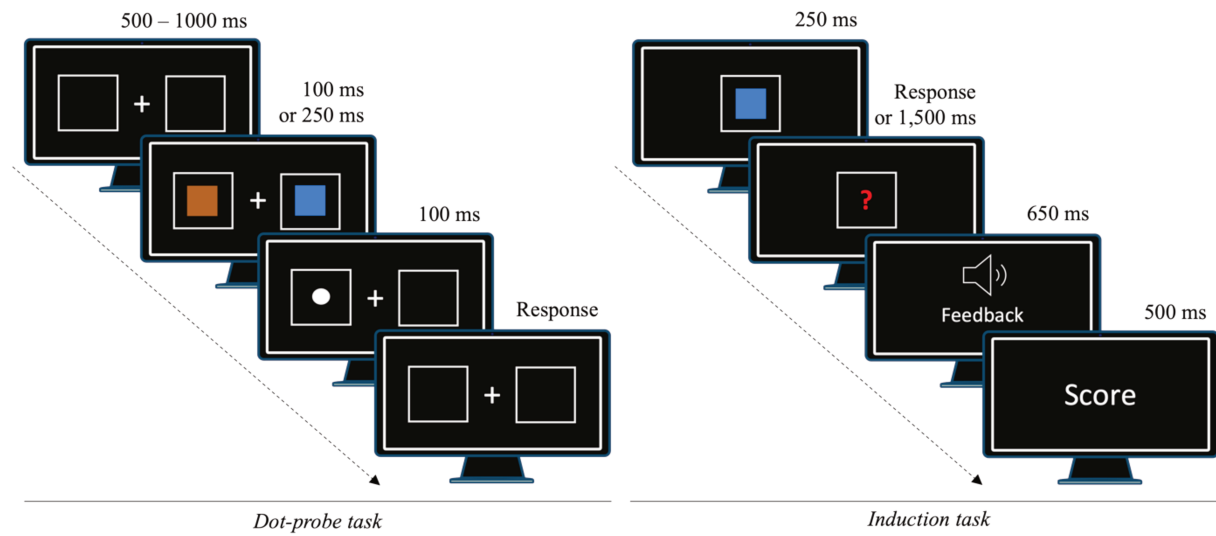
- Makowski, D., Ben-Shachar, M., & Lüdtke, D. (2019). bayestestR: Describing Effects and their Uncertainty, Existence, and Significance within the Bayesian Framework. *Journal of Open Source Software*, 4(40), 1541. <https://doi.org/10.21105/joss.01541>
- Mazzietti, A., Sellem, V., & Koenig, O. (2014). From stimulus-driven to appraisal-driven attention: Towards differential effects of goal relevance and goal relatedness on attention? *Cognition and Emotion*, 28(8), 1483-1492. <https://doi.org/10.1080/02699931.2014.884488>
- Mazzietti, A., Sellem, V., & Koenig, O. (2015, July 9). To see, or not to see: Dynamic attentional interference by non-consciously perceived relevant stimuli [Conference session]. Bi-Annual Conference of the International Society for Research On Emotion, Geneva, Switzerland. <https://www.unige.ch/cisa/isre2015/sites/default/files/Mazzietti.pdf>
- Miller, N. E. (1944). Experimental studies of conflict. In Hunt, J. (Ed), *Personality and the behavior disorders*, 431-465. New York: Ronald Press.
- Morey, R. D., & Rouder, J. N. (2022). BayesFactor: Computation of Bayes Factors for Common Designs. R package version 0.9. 12-4.4. <https://CRAN.R-project.org/package=BayesFactor>
- Nosek, B.A., Alter, G., Banks, G.C., Borsboom, D., Bowman, S.D., Breckler, S.J., Buck, S., Chambers, C.D., Chin, G., Christensen, G., Contestabile, M., Dafoe, A., Eich, E., Freese, J., Glennerster, R., Goroff, D., Green, D.P., Hesse, B., Humphreys, M., Ishiyama, J., Karlan, D., Kraut, A., Lupia, A., Mary, P., Madon, T., Malhotra, N., Mayo-Wilson, E., McNutt, M., Miguel, E., Levy Paluck, E., Simonsohn, U., Soderberg, C., Spellman, B.A., Turitto, J., Vandenbos G., Vazire, S., Wagenmakers, E.J., Wilson., & Yarkoni, T. (2015). Scientific Standards. Promoting an open research culture. *Science (New York, N.Y.)*, 348(6242), 1422–1425. <https://doi.org/10.1126/science.aab2374>



- Öhman, A. (1993). Fear and anxiety as emotional phenomena: Clinical phenomenology, evolutionary perspectives, and information-processing mechanisms. In M. Lewis & J. M. Haviland (Eds.), *Handbook of Emotions* (pp. 511–536). The Guilford Press.
- Öhman, A., & Mineka, S. (2001). Fears, phobias, and preparedness: toward an evolved module of fear and fear learning. *Psychological Review*, *108*(3), 483-522.  
<https://doi.org/10.1037/0033-295X.108.3.483>
- Ossenfort, K. L., & Isaacowitz, D. M. (2021). Spatial attention to arousing emotional stimuli in younger and older adults. *Motivation and Emotion*, *45*(6), 790-797.  
<https://doi.org/10.1007/s11031-021-09899-x>
- Posner, J., Russel, J., & Peterson, B. (2005). The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and Psychopathology*, *17*(3), 715-734.  
<https://doi.org/10.1017/S0954579405050340>
- Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of mathematical psychology*, *56*(5), 356-374.  
<https://doi.org/10.1016/j.jmp.2012.08.001>
- Scherer, K. R. (2009). The dynamic architecture of emotion: Evidence for the component process model. *Cognition and Emotion*, *23*(7), 1307-1351.  
<https://doi.org/10.1080/02699930902928969>
- Scherer, K. R. (2010). The component process model: A blueprint for a comprehensive computational model of emotion. In K. R. Scherer, T. Bänziger & E. B. Roesch (Eds.), *Blueprint for affective computing: A sourcebook* (pp. 47–70). Oxford University Press.
- Scherer, K. R. (2013). The nature and dynamics of relevance and valence appraisals: Theoretical advances and recent evidence. *Emotion Review*, *5*(2), 150-162.  
<https://doi.org/10.1177/1754073912468166>

- Scherer, K. R., & Moors, A. (2019). The emotion process: Event appraisal and component differentiation. *Annual Review of Psychology*, *70*, 719-745.  
<https://doi.org/10.1146/annurev-psych-122216-011854>
- Schimmack, U. (2005). Attentional Interference Effects of Emotional Pictures: Threat, Negativity, or Arousal? *Emotion*, *5*(1), 55–66. <https://doi.org/10.1037/1528-3542.5.1.55>
- Schönbrodt, F. D., & Wagenmakers, E. J. (2018). Bayes factor design analysis: Planning for compelling evidence. *Psychonomic bulletin & review*, *25*(1), 128-142.  
<https://doi.org/10.3758/s13423-017-1230-y>
- Shuman, V., Sander, D., & Scherer, K. R. (2013). Levels of valence. *Frontiers in Psychology*, *4*, 261. <https://doi.org/10.3389/fpsyg.2013.00261>
- Spiering, M., & Everaerd, W. (2007). The sexual unconscious. In E. Janssen (Ed.), *The Psychophysiology of Sex* (pp. 166–184). Indiana University Press.
- Stefan, A. M., Gronau, Q. F., Schönbrodt, F. D., & Wagenmakers, E. J. (2019). A tutorial on Bayes Factor Design Analysis using an informed prior. *Behavior Research Methods*, *51*(3), 1042-1058. <https://doi.org/10.3758/s13428-018-01189-8>
- Steimke, R., Nomi, J. S., Calhoun, V. D., Stelzel, C., Paschke, L. M., Gaschler, R., Goschke, T., Walter, H., & Uddin, L. Q. (2017). Salience network dynamics underlying successful resistance of temptation. *Social Cognitive and Affective Neuroscience*, *12*(12), 1928-1939.  
<https://doi.org/10.1093/scan/nsx123>
- Sun, L., Ding, C., Xu, M., Diao, L., & Yang, D. (2017). Engagement attentional bias toward value-associated stimuli. *Current Psychology*, *36*(4), 747-754.  
<https://doi.org/10.1007/s12144-016-9462-y>
- Vogt, J., De Houwer, J., Crombez, G., & Van Damme, S. (2013). Competing for attentional priority: temporary goals versus threats. *Emotion*, *13*(3), 587.  
<https://doi.org/10.1037/a0027204>

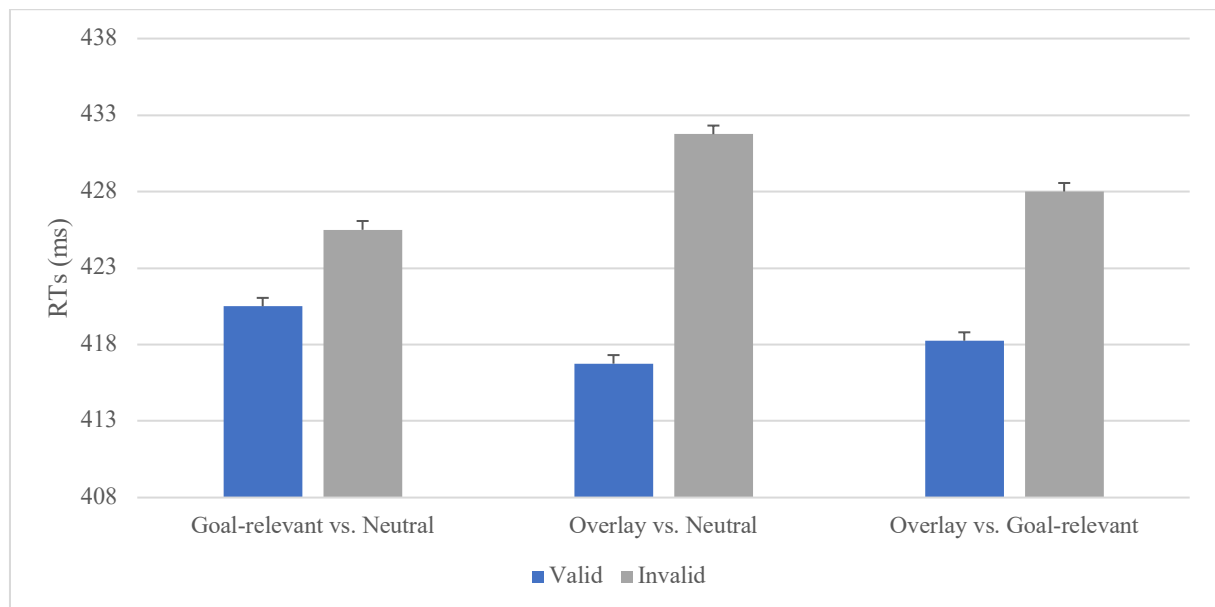
- Vogt, J., Koster, E. H., & De Houwer, J. (2017). Safety first: Instrumentality for reaching safety determines attention allocation under threat. *Emotion, 17*(3), 528.  
<https://doi.org/10.1037/emo0000251>
- Vogt, J., Lozo, L., Koster, E. H., & De Houwer, J. (2011). On the role of goal relevance in emotional attention: Disgust evokes early attention to cleanliness. *Cognition and Emotion, 25*(3), 466-477. <https://doi.org/10.1080/02699931.2010.532613>
- Werthmann, J., Roefs, A., Nederkoorn, C., Mogg, K., Bradley, B. P., & Jansen, A. (2011). Can(not) take my eyes off it: Attention bias for food in overweight participants. *Health Psychology, 30*(5), 561–569. <https://doi.org/10.1037/a0024291>
- Yeung, R. C., & Fernandes, M. A. (2021). Divided attention at encoding or retrieval interferes with emotionally enhanced memory for words. *Memory, 29*(3), 284-297.  
<https://doi.org/10.1080/09658211.2021.1887896>

**Figure 1.***Schematic Overview of the Experimental Design*

*Note.* The dot-probe task consisted of the appearance of four screens: a fixation cross with two empty squares, the cues in the two squares, the dot in one of the squares, and the response screen. The induction task immediately started after the participant's response to the dot-probe task and consisted of the appearance of four screens: a cue in the center of a square, a red question mark in the square, a feedback that could be accompanied by a noise depending on the cue that appeared, and the updated score.

**Figure 2**

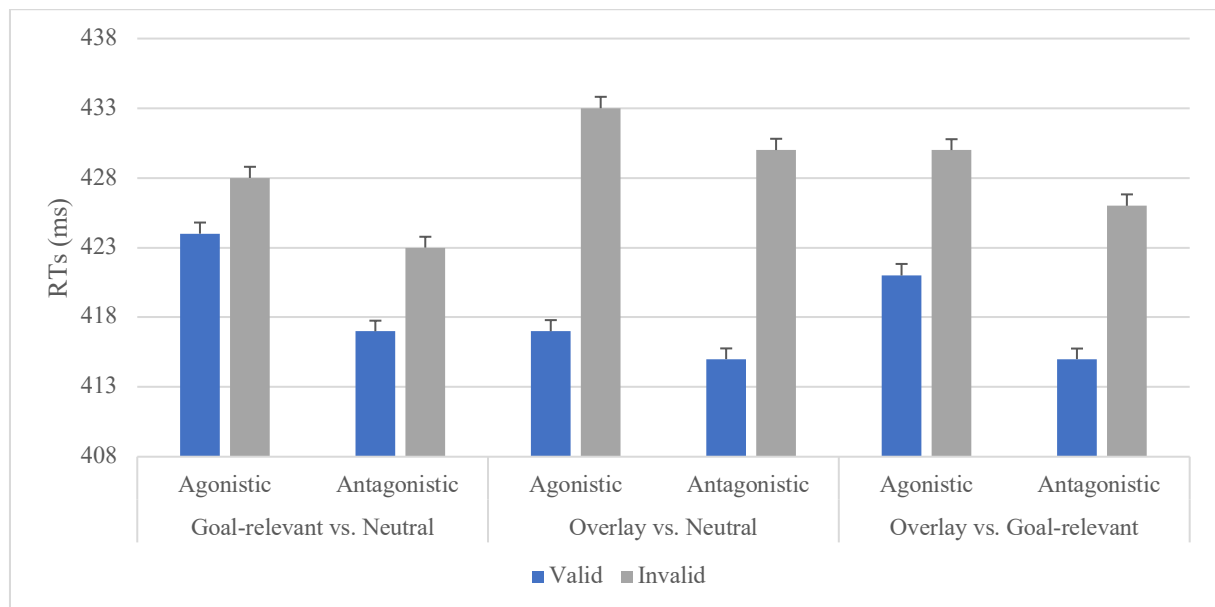
*Mean Reaction Times and Standard Errors as a Function of Pair type and Validity*



*Note. Goal-relevant vs. Neutral refers to the pair contrasting the goal-relevant stimulus with the neutral stimulus. Overlay vs. Neutral refers to the pair contrasting the overlay stimulus with the neutral stimulus. Overlay vs. Goal-relevant refers to the pair contrasting the overlay stimulus (i.e., both goal conducive and intrinsically relevant, regardless of whether it was pleasant or unpleasant) with the goal-relevant stimulus. Valid refers to the trials in which the dot replaced the first stimulus of the pair, while Invalid refers to the trials in which the dot replaced the second stimulus of the pair.*

**Figure 3**

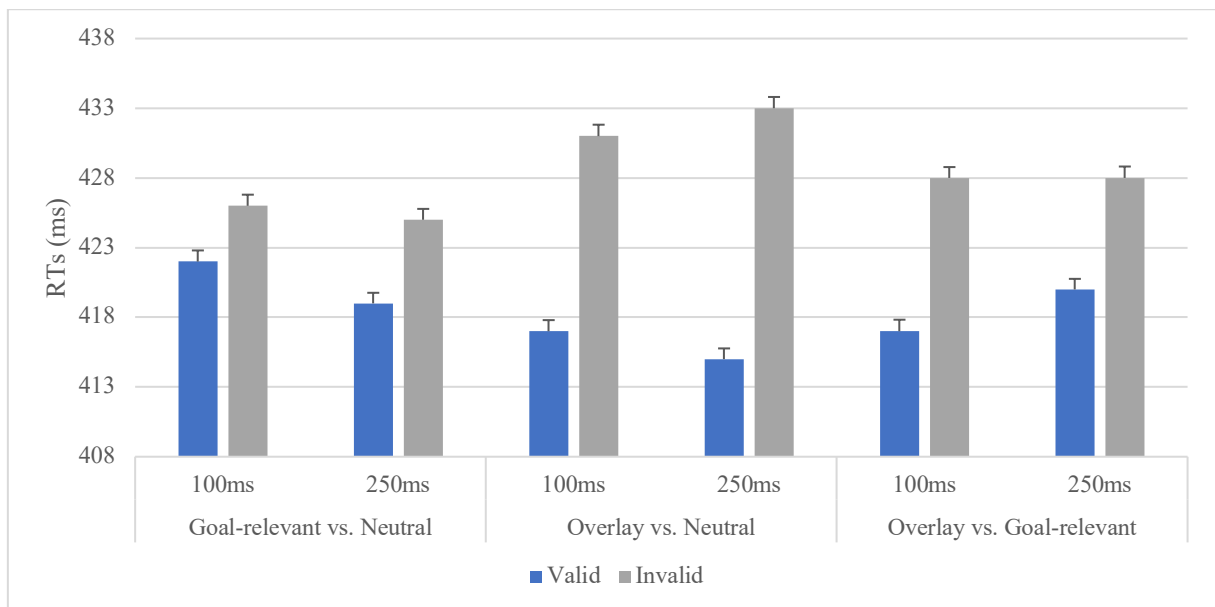
*Mean Reaction Times and Standard Errors as a Function Pair type, Group, and Validity*



*Note. Goal-relevant vs. Neutral refers to the pair contrasting the goal-relevant stimulus with the neutral stimulus. Overlay vs. Neutral refers to the pair contrasting the overlay stimulus with the neutral stimulus. Overlay vs. Goal-relevant refers to the pair contrasting the overlay stimulus (i.e., both goal conducive and intrinsically relevant), with the goal-relevant stimulus. Agonistic refers to the agonistic group in which the intrinsic relevance assigned to the overlay stimulus was pleasant, whereas antagonistic refers to the antagonistic group in which the intrinsic relevance assigned to the overlay stimulus was unpleasant. Valid refers to the trials in which the dot replaced the first stimulus of the pair, while Invalid refers to the trials in which the dot replaced the second stimulus of the pair.*

**Figure 4**

*Mean Reaction Times and Standard Errors as a Function of Pair type, SOA, and Validity*



*Note. Goal-relevant vs. Neutral refers to the pair contrasting the goal-relevant stimulus with the neutral stimulus. Overlay vs. Neutral refers to the pair contrasting the overlay stimulus with the neutral stimulus. Overlay vs. Goal-relevant refers to the pair contrasting the overlay stimulus (i.e., both goal conducive and intrinsically relevant), with the goal-relevant stimulus. 100ms and 250ms refers to the two SOA values. Valid refers to the trials in which the dot replaced the first stimulus of the pairs, while Invalid refers to the trials in which the dot replaced the second stimulus of the pair.*

***Chapitre 4 - Article : “Modulation of attentional capture by combining intrinsic pleasantness and goal relevance or unavoidable goal obstruction”***

Ce chapitre décrit le troisième article que nous avons écrit qui est actuellement en soumission dans une revue, dont la référence est la suivante :

Fournier, H. & Koenig, O. (Submitted a). Attentional impact of relevance overlay: Effects of goal conduction and unavoidably goal obstruction inductions on pleasant and unpleasant stimuli.



## Résumé

Dans un environnement complexe et dynamique, les stimuli agréables ou désagréables favorisent ou entravent souvent la réalisation d'un objectif. Récemment, il a été démontré que la propension à atteindre un but augmentait la capture attentionnelle des stimuli agréables et désagréables (Fournier & Koenig, Under review). Cependant, à notre connaissance, l'impact attentionnel de l'obstruction d'un but sur des stimuli agréables et désagréables n'a jamais été testé. Alors que des stimuli neutres obstruant inévitablement un but n'ont pas révélé d'effet attentionnel (Müller & Wentura, 2022), cet effet attentionnel n'a jamais été testé dans le cas où les stimuli sont agréables ou désagréables. Dans la présente étude, nous avons testé l'impact attentionnel de la superposition d'une facilitation ou d'une obstruction inévitable de l'atteinte d'un but par des stimuli agréables et désagréables. À cette fin, les participants ont effectué une tâche de dot-probe dans laquelle des stimuli agréables et désagréables étaient en concurrence avec des stimuli neutres. Parallèlement, les participants ont effectué une tâche d'induction dans laquelle les stimuli étaient induits à la conductivité au but (gain de points) ou à l'obstruction inévitable du but (perte automatique de points). Les résultats ont montré que la capture attentionnelle vers les stimuli agréables et désagréables augmentait s'ils étaient également facilitateurs dans l'atteinte du but et diminuait s'ils étaient également obstructifs dans l'atteinte du but. Ces résultats ont été discutés dans le contexte des théories de l'appraisal des émotions.

**Modulation of attentional capture by combining intrinsic pleasantness and goal  
relevance or unavoidable goal obstruction**

Hippolyte Fournier and Olivier Koenig  
Laboratoire EMC, Université de Lyon

**Author details**

Hippolyte Fournier, affiliated with Laboratoire EMC, Université de Lyon.

 <https://orcid.org/0000-0002-7861-2387>

Olivier Koenig, affiliated with Laboratoire EMC, Université de Lyon.

 <https://orcid.org/0000-0002-5124-1889>

Correspondence should be addressed to Hippolyte Fournier, Laboratoire EMC, Université

Lumière Lyon 2, 5 Avenue Pierre Mendès France, Bron 69676, France. Email:

[h.fournier@univ-lyon2.fr](mailto:h.fournier@univ-lyon2.fr)

### **Abstract**

In a complex and dynamic environment, pleasant or unpleasant stimuli are often conducive or obstructive to a goal achievement. Recently, goal conduciveness has been shown to increase the attentional capture of both pleasant and unpleasant stimuli (Fournier & Koenig, Under review). However, to our knowledge, the attentional impact of goal obstruction on pleasant and unpleasant stimuli has never been tested. While neutral stimuli inevitably obstructing a goal did not reveal any attentional effect (Müller & Wentura, 2022), the presence of such an effect has never been investigated in the case of pleasant or unpleasant stimuli. In the present study, we proposed to test the attentional impact of superimposing goal conduction or unavoidable goal obstruction to pleasant and unpleasant stimuli. To this end, participants performed a dot-probe task in which pleasant and unpleasant stimuli competed with neutral stimuli. In parallel, participants performed an induction task in which stimuli were induced to goal conduction (point gain) or unavoidable goal obstruction (automatic point loss). Results showed that the attentional capture toward pleasant and unpleasant stimuli increased if they were also goal-conducive and decreased if they were also goal-obstructive. These results were discussed in the context of appraisal theories of emotion.

*Keywords:* emotion, attention, appraisal, intrinsic relevance, goal relevance

## **Modulation of attentional capture by combining intrinsic pleasantness and goal relevance or unavoidable goal obstruction**

The literature on the impact of emotion on attention is extensive. Although many different assumptions have been proposed, they are all based on the common idea that the purpose of attention is intended to indicate which stimulus should be processed first to ensure the well-being and survival of individuals (Brosch et al., 2010). First, researchers proposed that attentional capture be reserved for negative stimuli such as snakes, spiders, or firearms to indicate that they likely require to be avoided (Brosch & Sharma, 2005; Fox, et al., 2007; LoBue & DeLoache, 2008; Öhman, 1993; Öhman & Mineka, 2001). Later, it was shown that positive stimuli such as baby faces, happy faces, or erotic content also capture attention to indicate that they likely require to be approached (Brosch et al., 2007; Brosch et al., 2008; de Jong et al., 2007; Fromberger et al., 2012; Joormann & Gotlib, 2007; Spiering & Everaerd, 2007; Steimke et al., 2017). To take into account all of these observations, a consensual view of the prioritization of emotional attentional was proposed under the notion of cognitive appraisal of stimuli relevance (Grandjean & Scherer, 2014; Schimmack, 2005).

According to appraisal theories, relevance appraisal would be the trigger of both attentional capture and action tendency toward stimuli. More precisely, the relevance appraisal of stimuli would be carried out by two sub-criteria. On the one hand, intrinsic relevance would appraise whether stimuli are generally pleasant (likely generating approach tendency) or rather generally unpleasant (likely generating avoidance tendency). The attentional bias toward pleasant stimuli such as baby faces, happy faces, and erotic content, or unpleasant stimuli such as firearms, snakes, and spiders would reflect an attentional bias toward intrinsically relevant stimuli. On the other hand, goal relevance would appraise whether stimuli are currently goal conducive (likely generating approach tendency) or goal-obstructive (likely generating avoidance tendency). Regarding goal conduciveness, attentional biases have been shown toward water

bottles for thirsty individuals, clean-related stimuli after a disgust induction, gain-related stimuli in a game context, and safety-related stimuli in a threat context (Mazzietti et al., 2014; Mazzietti et al., 2015; Müller & Wentura, 2022; Vogt et al., 2011; Vogt et al., 2017). In the case of a stimulus being appraised as goal obstructive, the appraisal of an additional criterion is triggered, namely, the control potential over the goal obstruction (Grandjean & Scherer, 2014). In some cases, the goal obstruction is controllable and could be avoidable by taking an action. In other cases, the goal obstruction is uncontrollable and could not be avoided. The evidence of the impact of goal obstruction on attention was shown in the case of avoidable goal obstruction when the loss triggered by the obstructive stimuli could be avoided by pressing a key (Müller et al., 2016; Müller & Wentura, 2022; Sun et al., 2017). However, in the case of unavoidable goal obstruction (i.e., when the goal obstruction cannot be avoided by an action), no attentional effect could be measured toward goal-obstructive stimuli (Müller & Wentura, 2022). It is conceivable that the effect would emerge with a larger sample size, or that no attentional effect would occur in this case because the unavoidable goal obstruction would not generate any avoidance tendency since obstruction cannot be avoided.

The reality is probably more complex because stimuli are often overlaid by both an intrinsic relevance and a goal relevance, such as illustrated in the approach-avoidance conflict literature (Aupperle et al., 2015; Bach et al., 2014; Ito & Lee, 2016). Let us take the following example about food. Food can be appraised both as pleasant or unpleasant and as goal conducive (to satiety) or goal obstructive (to diet). Thus, the relevance overlay can be agonistic if the two relevances both generate either an approach tendency (if food is pleasant and goal conducive) or an avoidance tendency (if food is unpleasant and goal obstructive). In contrast, the relevance overlay can be antagonistic if the two relevances generate both approach and avoidance tendencies (i.e., if food is pleasant and goal obstructive or unpleasant and goal conducive). Fournier & Koenig (2023) theorized the impact of relevance overlay on attention based on a

predictive view of brain function (Howhy, 2013). Indeed, relevance appraisal is nothing more than inference based on prior beliefs emerging from past interactions with stimuli (Barrett, 2017). The purpose of confronting the predictions of intrinsic relevance and goal relevance about a stimulus would provide information about the quality of the predictions, namely if the stimulus should likely be approached or avoided.

Fournier and Koenig (2022) predicted that regarding the impact of relevance overlay on attention, the overlay of agonistic and antagonistic relevances would increase attentional bias toward the stimulus in the same way. In the case of agonistic relevance overlay, the increase in attentional bias toward the stimulus would reflect the likelihood that it should be approached or avoided, and this likelihood would be maximized because the two overlaid relevances generate either approach or avoidance tendency. In the case of antagonistic relevance overlay, the increase in attentional bias toward the stimulus would reflect the application of one fundamental principle of the brain function according to which the uncertainty must constantly be minimized. (Friston, 2005; 2010). Thus, uncertainty is high since the two relevances indicate that the stimulus should likely be both approached and avoided. Given that, the attentional bias toward the stimulus would reflect the need for further appraisal to decide which action tendency is more credible and ultimately may reduce uncertainty. Evidence in favor of an increased attentional bias toward the stimulus in the case of an overlay of pleasantness and goal conduction or unpleasantness and goal conduction was reported in the literature (Fournier & Koenig, Under review). However, the effect of the impact of relevance overlay on the attentional bias when goal relevance is goal obstructive has never been tested to our knowledge.

One can imagine that the outcome may be different depending on whether the goal obstruction can be avoided or not. In the case the goal obstruction can be avoided, the attentional bias toward the stimulus would increase since a tendency would be generated to act against the obstruction of the goal. Thus, if the intrinsic relevance is unpleasant, the likelihood that the

stimulus should be avoided would be maximized. In addition, if the intrinsic relevance is pleasant, the uncertainty of the required action would be maximized. In the case the goal obstruction cannot be avoided, one can assume that only the intrinsic relevance of the stimulus increases the attentional bias, since no action against the obstruction of the goal can be generated. Thus, the likelihood that the stimulus should be approached or avoided is only carried by the fact that its intrinsic relevance is pleasant or unpleasant, respectively.

In the present study, we proposed to test two hypotheses regarding the impact of the relevance overlay on attentional bias. On the one hand, considering that pleasant and unpleasant stimuli capture more attention than neutral stimuli, we hypothesized that their attentional capture would increase if they are also goal conducive following an induction procedure. On the other hand, we hypothesized that the attentional capture toward pleasant and unpleasant stimuli would not differ following an unavoidable goal obstruction induction.

To test these hypotheses, a paradigm alternating trials of a dot-probe task with an induction task was proposed to the participants (Vogt et al., 2013; Vogt et al., 2017; Forrest et al., 2022). In this task, two relevant stimuli, one pleasant (a naked woman's breast) and one unpleasant (a spider) were pitted against two neutral stimuli (a plant and a mushroom) to measure differences in attentional capture. The induction task was presented as a game in which the aim was to reach the maximum number of points, the real purpose being to induce goal relevance to the stimuli. Two types of induction were used: A goal conduction (a gain of points if a key was pressed) and an unavoidable goal obstruction (an automatic loss of points). The participants performed the tasks while these inductions were attributed, respectively: to the pleasant and the unpleasant stimuli, to the unpleasant and the pleasant stimuli, to the plant and the mushroom, to the mushroom and the plant, or to none of them (by presenting a set of filler stimuli in the induction task). This allowed us to measure the attentional capture within and between pairs contrasting the relevant (i.e., pleasant or unpleasant) stimuli with the neutral stimuli depending

on the following induction patterns: The gain was overlaid to the relevant stimuli (O+), the automatic loss was overlaid to the relevant stimuli (O-), the gain competed with the relevant stimuli (C+), the automatic loss competed with the relevant stimuli (C-), no induction was used (NI). Thus, the relevant stimuli were induced in the overlay conditions, whereas the neutral stimuli were induced in the competition conditions. First, we expected that the relevant stimuli would capture more attention than the neutral stimuli in the NI induction pattern. Second, we expected that the attentional capture toward the relevant stimuli would increase in the O+ compared to the NI induction pattern. Third, we expected attentional capture toward the relevant stimuli to be equivalent between NI and O- induction patterns, assuming that the unavoidable goal obstruction did not increase attentional capture. Eventually, the competition conditions were used to test the reliability of the inductions. We expected that the neutral stimuli would capture more attention than the relevant stimuli in the C+ induction pattern, whereas relevant stimuli would still capture more attention than neutral stimuli in the C- induction pattern. This was done to emphasize the attentional capture-enhancing power of goal conduction and the lack of attentional effect of unavoidable goal obstruction. These hypotheses were made for the two relevant stimuli, namely pleasant and unpleasant.

Finally, attentional capture within the pairs was unpacked according to whether it was better explained by facilitated engagement or delayed disengagement. To this end, we used the vigilance and disengagement indices proposed by Koster et al. (2004). The calculations of these indices required a baseline measurement of the participant's response times in the dot-probe task. To this end, an additional pair contrasting the same stimulus with itself (i.e., a mixture of the shapes and colors of the other stimuli) was included in the dot-probe task. The reliability of the assumption that the mean RTs of this pair could be used as a baseline is a critical point for the use of these indices. To check this, the pairs contrasting the pleasant with the unpleasant stimuli and the plant with the mushroom were presented in the dot-probe task. It was assumed



that when the stimuli were not subjected to induction, the mean RTs of these pairs had to be equivalent to the mean RTs of the pair contrasting the mixed stimulus with itself. This was checked before the analysis of the results about the vigilance and disengagement indices.

At the end of the task, participants were asked to report the pleasantness level of stimuli.

## Method

### *Participants*

One hundred and seventeen students (117 men,  $M_{age} = 21.6$ ;  $SD_{age} = 3.12$ ) from Lyon University participated in the study. Since the pleasant stimulus depicted the breast of a naked woman, only men were recruited as participants (Jiang et al., 2006; Lykins et al., 2008). Participants had normal or corrected-to-normal vision. The determination of the sample size is reported in the Bayesian analysis plan section.

### *Material*

The experiment was run on a 21" iMac in a soundproofed room using SuperLab 5 software (Haxby et al., 1991).

The main set of stimuli consisted of four pictures cropped and pasted on a dark blue square (400\*400px): two relevant stimuli, one pleasant (a naked woman's breast), one unpleasant (a spider), and two neutral stimuli (a plant named "neutral 1" and a mushroom named "neutral 2"). In the center of each image of the main set, a black circle (diameter 20px) was drawn where the dot in the dot-probe task was to appear, so as not to bias the dot detection task by perceptual differences depending on the images. The luminance was corrected and equalized across the stimuli (i.e.,  $M = 10.51$  cd/m<sup>2</sup>;  $SD = 0.05$ ). Measurements were made with a "Konica Minolta LS-110 luminance meter" (Konica Minolta, Mahwah, NJ). A filler set of stimuli was also constructed consisting of four pictures cropped and pasted on the same dark blue square (400\*400px): a bike, a mug, a car, and a lamp. An additional stimulus, the mixed stimulus, was constructed by mixing the shapes and colors of the four pictures of the main set. This stimulus

also contained a hole in its center. Stimuli were created using images from the Pixabay open-source database. All treatments were performed on Gimp. (<https://www.gimp.org>).

### ***Design and Procedure***

[Figure 1 near here]

#### *Dot-probe and induction tasks design*

A successive trial of the dot-probe task and the induction task is illustrated in Figure 1. Each trial of the dot-probe task started with the appearance of a white fixation cross (16\*16px) in the center of the computer screen with two empty squares with white outlines (401\*401px) at 250px to the left and right of the fixation cross. After a random delay from 500 to 1000ms, a stimuli pair from the main set of stimuli appeared in the center of each square for 250ms. The stimulus pair contrasted either one of the relevant stimuli with one of the neutral stimuli, the mixed stimulus with itself, both relevant stimuli, or both neutral stimuli. Immediately after the stimuli offset, a white dot (16\*16px) appeared in the center of one of the two squares during 100ms. The trial ended when the “s” key or the “k” key was pressed. Six hundred milliseconds after the end of the dot-probe task event, a trial of the induction task started with the appearance of a stimulus in the center of the screen for 800ms or until the space bar key was pressed. This stimulus could be one of the main stimuli set or the mixed stimulus, except for some participants, for whom the stimulus set was the filler set. When the stimulus was turned off, feedback appeared for 600ms indicating whether points had been gained or lost and whether the response key had been correctly pressed. Finally, the updated total score appeared for 600ms.

#### *Procedure*

First, participants were divided equally into each induction pattern subgroup. Gain inductions and automatic loss induction could be attributed, respectively: to the pleasant and

unpleasant stimuli, to the unpleasant and pleasant stimuli, to the neutral 1 and neutral 2 stimuli, to the neutral 2 and neutral 1 stimuli, or to two stimuli in the filler set.

Before starting the experiment, participants had to complete a consent form. Then, they sat in front of the computer, sixty centimeters from the middle of the screen, and placed the left index on the “s” key, the right index on the “k” key, and the left or right thumb on the space bar of the computer keyboard. The instructions were then explained to them. Participants were asked to stare at the fixation cross throughout the whole experiment. Regarding the dot-probe task, the instructions were to respond as quickly and precisely as possible (Chica et al., 2014) with the “s” or “k” keys, depending on whether the dot appeared on the left or the right, respectively. The induction task was presented as a game where the goal was to obtain the maximum number of points. Participants were told that the speed of response had no impact on the game. To maintain their motivation throughout the tasks it was explained to participants that if they obtained one of the best scores, they could choose a nickname that would appear in the top 3 players. Then, it was explained that if the stimulus attributed to the gain induction appeared in the induction task, pressing a key would result in a gain of 20 points. Pressing a key after the appearance of a stimulus not assigned to this induction was considered an error. If the stimulus attributed to the automatic loss induction appeared, it resulted in an unavoidable loss of 10 points. The two remaining stimuli could not impact the score. Then, participants performed 10 training trials in the presence of the experimenter. The training restarted until all ten responses were correct, to ensure that the induction worked from the beginning of the task. Then, the experimenter left the room, and the experiment began.

Participants completed 260 trials so that, in the dot-probe task, each of the seven pairs was randomly presented 10 times in each presentation condition. Specifically, the six pair combinations among the stimuli of the main set appeared 40 times, with a given stimulus appearing 20 times in the left square and 20 times in the right square, followed by a dot

appearing 10 times on the left and 10 times on the right in each instance. Since the pair contrasted the mixed stimulus with itself, no counterbalancing was necessary, so this pair appeared only 20 times followed by a dot appearing 10 times on the left and 10 times on the right. In the induction task, each of the four stimuli of one set and the mixed stimulus randomly appeared 52 times. In the end, participants were asked to rate the intrinsic relevance level of the stimuli of the main set on a continuous seven-point scale, ranging from 1 (*very unpleasant*) to 7 (*very pleasant*), with 4 being considered neutral. All stimuli randomly appeared twice to increase the reliability of the measure.

### ***Data Processing***

Seventeen participants were eliminated because they were identified as not being sufficiently engaged in the task. Elimination was based on a threshold that we defined as more than 5% error in the dot-probe task. For the remaining participants, erroneous responses were discarded (0.02% of the data). Then, RTs reflecting anticipatory (i.e., shorter than 150ms) and delayed (i.e., longer than 750) responses were discarded in order not to bias the computations of mean response times (Carlson & Fang, 2020). This concerned 0.01% of the data remaining data.

### ***Measures***

The measurement focused solely on the data of the dot-probe task contrasting the pleasant or the unpleasant stimulus with one of the two neutral stimuli as a function of different induction patterns. Within the dot-probe task, trials in which the dot replaced the stimulus expected to be the most relevant of a pair are commonly defined as valid trials, while trials in which the dot replaced the stimulus expected to be the less relevant of a pair are defined as invalid trials. Shorter RTs for valid than invalid trials means that the stimulus expected to be the most relevant captured more attention than the other, whereas shorter RTs for invalid than valid trials means that the stimulus expected to be the less relevant captured more attention than the other. Within the pairs, valid trials were defined by trials in which the dot replaced the pleasant or the

unpleasant stimuli, and invalid trials were defined by trials in which the dot replaced the neutral stimuli. This was irrespective of whether the stimuli were attributed to gain or automatic loss induction.

Regarding the unpack of attentional capture within the pairs according to facilitated engagement or delayed disengagement, the analysis could not be conducted. The pair contrasting the mixed stimulus with itself was unreliable as the baseline of participants' mean RTs. Indeed, the mean RTs of this pair was longer than the mean RTs of the pairs contrasting the relevant stimuli with the neutral stimuli.

### ***Bayesian Analysis***

#### *The rationale for the use of Bayesian analysis*

Bayesian analyses were performed because they are richer and more informative than the Null Hypothesis Significance Test (NHST; Kruschke, 2013; Kruschke 2014). First, the Bayesian estimators such as Bayes Factor (BF) allow obtaining a measure of a hypothesis likelihood rather than a binary decision based on an arbitrary threshold such as in the case with the p-value. Second, Bayesian analysis allows supporting the null hypothesis rather than not being able to reject it. The  $BF_{10}$  provides the likelihood ratio in favor of the hypothesis that the effect generated the data versus the hypothesis that the effect did not generate the data. In contrast, the  $BF_{01}$  provides the likelihood ratio in favor of the hypothesis that the effect did not generate the data versus the hypothesis that the effect generated the data. They can be interpreted as follows:  $BF \geq 3$  highlights moderate evidence,  $BF \geq 10$  highlights strong evidence, and  $BF \geq 100$  highlights decisive evidence (Kass & Raftery, 1995). In this framework, the Bayesian ANOVA (BANOVA) procedure works as a stepwise regression method. In a complex design containing main and interaction effects, the BF of the effects of interests are first observed from the full BANOVA (i.e., containing all main and interaction effects). Then, the data are entered separately into multiple BANOVAs following the same structure as the full

BANOVA. This is done to observe the consistency of simple effects through the BF when data of no interest to the effects are dropped from the BANOVA.

### *Determination of the Sample Size*

The sample size was determined using the sequential Bayes Factor Design Analysis (BFDA; Schönbrodt & Wagenmakers, 2018). The idea of such a design is to first determine a minimum number of participants ( $n_{\min}$ ) needed to correctly detect the expected effects, knowing a given BF threshold. Then, starting from the  $n_{\min}$ , the inclusion of participants is performed until all BFs of expected effects exceed the threshold in favor or against the expected hypothesis, or until the defined maximum number of participants ( $n_{\max}$ ) is reached. To guarantee the proper practice of the analysis, we followed the rules used in the step-by-step example proposed by Stefan et al. (2019) by selecting the most restrictive effect size,  $d = 0.2$ , setting the  $BF_{10}$  and the  $BF_{01}$  thresholds at 6, and choosing default prior. Using these parameters, a power between 80% and 95% was reached for a number of observations between 235 and 498 per sample regarding the  $BF_{10}$  threshold. While a power between 80% and 95% was reached for a number of observations between 120 and 330 per sample regarding  $BF_{01}$  (The BFDA report is available through the osf repository). Thus, we decided to set the  $n_{\min}$  at 100, so that each sample within modalities among the interaction effects and random effects entered in the BANOVAs contained at least 400 observations per sample. These 400 observations consisted of the replication of 20 trials of each modality performed by 20 participants. Trial modalities referred to validity conditions (i.e., valid or invalid), induction patterns (i.e., NI, O+, O-, C+, and C-), and stimuli assigned to induction patterns (i.e., pleasant or unpleasant stimuli for O+ and O-, and neutral 1 or neutral 2 stimuli for C+ and C-). The  $n_{\max}$  was set to 400. Because all expected effects exceed the BF threshold from the  $n_{\min}$ , the inclusion stopped at 100 participants. This  $n_{\min}$  may seem small for a behavioral study, so we emphasize the importance of considering this sample size consistent with the robustness of the effects found (most BFs exceeded the

decisive evidence threshold). Thus, we assume that the sample size of this study was sufficient to generalize the results.

## Result

### *Ratings of the intrinsic relevance level of stimuli*

We first verified that the pleasant and unpleasant stimulus were indeed perceived as such, that gain induction made the stimuli more pleasant and that automatic loss induction made the stimuli more unpleasant. Descriptive statistics are presented in Table 1.

To this end, a  $4 \times 3$  BANOVA was performed on the ratings of the intrinsic relevance of stimuli that included the main and interaction effects of Stimulus (pleasant, unpleasant, neutral 1, and neutral 2) and Induction pattern (no induction, gain induction, and automatic loss) with Participant as random effect. As expected, a Stimulus main effect and an Induction pattern main effect were observed,  $BF_{10} > 1000$  in both cases. These results confirmed that the ratings of intrinsic relevance differed as a function of the stimulus and the induction pattern.

[Table 1 near here]

In order to better understand the nature of the differences regarding the main effect of the stimulus, five separate BANOVAs were fitted in which the ratings of either the pleasant stimulus or the unpleasant stimulus were entered in pairs with the ratings of each other stimuli. A Stimulus main effect was consistently observed in each of these BANOVAs. On the one hand, the pleasant stimulus ( $M = 5.32$ ,  $SD = 1.53$ ) was rated more pleasant than the unpleasant stimulus ( $M = 2.78$ ,  $SD = 1.44$ ),  $BF_{10} > 1000$ , than the neutral 1 stimulus ( $M = 4.44$ ,  $SD = 1.16$ ),  $BF_{10} > 1000$ , and than the neutral 2 stimulus ( $M = 3.93$ ,  $SD = 1.22$ ),  $BF_{10} > 1000$ . On the other hand, the unpleasant stimulus was rated more unpleasant than the neutral 1 stimulus,  $BF_{10} > 1000$ , and than the neutral 2 stimulus,  $BF_{10} > 1000$ . This check ensured that the pleasant stimulus was perceived as more pleasant than other stimuli and that the unpleasant stimulus was perceived as more unpleasant than other stimuli.

In order to better understand the nature of the differences regarding the main effect of the Induction pattern, three separate BANOVAs were fitted in which the ratings of the stimuli attributed to each modality of the Induction were entered in pairs with those of the other modalities. These analyses revealed that the stimuli induced to gain ( $M = 4.88$ ,  $SD = 1.74$ ) were rated more pleasant than the stimuli not induced ( $M = 4.06$ ,  $SD = 1.48$ ),  $BF_{10} > 1000$ , and than the stimuli induced to automatic loss ( $M = 3.51$ ,  $SD = 1.63$ ),  $BF_{10} > 1000$ . Moreover, the stimuli induced to automatic loss were rated more unpleasant than the stimuli not induced,  $BF_{10} > 1000$ . This check revealed that gain induction made the stimuli more pleasant, while automatic loss induction made them more unpleasant.

### ***Attentional capture and relevance***

The analysis of the dot-probe task data focused on the RTs of the four stimulus pairs contrasting the pleasant or the unpleasant stimuli with one of the two neutral stimuli as a function of the induction patterns. The purpose of this analysis was to verify that: 1) relevant stimuli captured more attention than the neutral stimuli in the NI induction pattern; 2) attentional capture by the relevant stimuli increased in the O+ compared to the NI induction pattern; 3) attentional capture by the relevant stimuli was equivalent between the NI and O- induction patterns; 4) neutral stimuli captured more attention than relevant stimuli in the C+ induction pattern, and that; 5) relevant stimuli always captured more attention than neutral stimuli in the C- induction pattern.

First, a  $2 \times 5 \times 2$  BANOVA was performed on RTs including the main and interaction effects of Validity (valid or invalid), Induction pattern (NI, O+, O-, C+, and C-), and Pleasantness (pleasant or unpleasant). Participant and Neutral stimulus type (neutral 1 and neutral 2) were entered as random effects. As expected, a Validity  $\times$  Induction pattern interaction was observed,  $BF_{10} > 1000$ , showing that attentional capture differed depending on whether the relevant or the neutral stimuli were induced to gain or to automatic loss or were not induced.



*Attentional capture by relevance overlay*

To address our hypotheses 1), 2), and 3), a series of separate BANOVAs was fitted from the full BANOVA. First, three BANOVAs including RTs of the NI, O+, and O- induction patterns were fitted. This was done to test the consistency of the Validity main effect within the Induction patterns. Then, the three BANOVAs were split in half to isolate RTs of the pleasant and the unpleasant stimuli. This was done to test the consistency of the Validity mean effect for the two levels of Pleasantness of stimuli within the Induction patterns.

Regarding the BANOVAs fitted separately on RTs of the NI induction pattern and of RTs of the O+ induction pattern, the Validity main effect was systematically found with shorter RTs for valid than invalid trials. This was observed in the BANOVAs in which data were not separated according to Pleasantness for RTs of NI and the O+ induction patterns,  $BF_{10} > 1000$  in both cases. The same result was observed in the BANOVAs in which data were separated according to Pleasantness,  $BF_{10} > 1000$ , for the BANOVA centered on RTs of the pleasant stimulus of the NI induction pattern,  $BF_{10} > 1000$ , for the BANOVA centered on RTs of the pleasant stimulus of the O+ induction patterns,  $BF_{10} > 1000$  for the BANOVAs centered on RTs of the unpleasant stimulus of the NI induction pattern, and  $BF_{10} > 1000$ , for the BANOVA centered on RTs of the unpleasant stimulus of the O+ induction pattern. The corresponding mean RTs and standard errors are reported in Figure 2.

These results suggest that when stimuli were not induced or when relevant stimuli were induced to the gain, they captured more attention than neutral stimuli, whether pleasant and unpleasant.

Surprisingly, regarding the BANOVAs fitted on the RTs of the O- induction pattern, an absence of Validity main effect was systematically observed. This was the case for the BANOVA in which data were not separated according to Pleasantness,  $BF_{01} = 23.26$ , the BANOVA centered on RTs of the pleasant stimulus,  $BF_{01} = 14.29$ , and the BANOVA centered

on RTs of the unpleasant stimulus,  $BF_{01} = 8.47$ . The corresponding mean RTs and standard errors are reported in Figure 2. These results suggest that when stimuli were induced to automatic loss, pleasant and unpleasant stimuli captured attention in the same way as neutral stimuli.

Then, two BANOVAs were fitted by including RTs from pairwise induction patterns. This was done for RTs of NI and O+ induction patterns and RTs of NI and O- induction patterns. These BANOVAs were split in half to isolate RTs of the pleasant and the unpleasant stimuli. The aim of this analysis was to test the consistency of the interaction effect of Validity  $\times$  Induction pattern.

With respect to the BANOVAs fitted on RTs of the NI and O+ induction patterns, the interaction effect of Validity  $\times$  Induction pattern was consistently observed with a greater magnitude between valid and invalid trials for the RTs of the O+ induction pattern,  $BF_{10} > 1000$  for the BANOVA in which data were not separated according to Pleasantness,  $BF_{10} > 100$  for the BANOVA centered on RTs of the pleasant stimulus, and  $BF_{10} > 100$  for the BANOVA centered on RTs of the unpleasant stimulus. The corresponding mean RTs and standard errors are reported in Figure 2. These results suggest that attentional capture by relevant stimuli was greater in the O+ compared to the NI induction pattern, whether relevant stimulus was pleasant or unpleasant.

With respect to the BANOVAs fitted on RTs of NI and O- induction patterns, an interaction effect of Validity  $\times$  Induction pattern was observed with a greater magnitude between valid and invalid trials for RTs of the NI induction pattern,  $BF_{10} > 1000$  for the BANOVA in which data were not separated according to Pleasantness,  $BF_{10} > 1000$  for the BANOVA centered on RTs of the pleasant stimulus, and  $BF_{10} > 1000$  for the BANOVA centered on RTs of the unpleasant stimulus,  $BF_{10} > 1000$ . The corresponding mean RTs and standard errors are reported in Figure 2. Contrary to our expectations, attentional capture toward the relevant stimuli was lower in the

O- compared to the NI induction pattern, whether the relevant stimulus was pleasant or unpleasant.

[Figure 2 near here]

#### *Attentional capture by relevance competition*

To address our hypotheses 4) and 5), four separate BANOVAs were fitted from the full BANOVA. These BANOVAs aimed to test the consistency of the Validity main effect as a function of the C+ and C- induction patterns, whether the relevant stimulus included in the pairs was pleasant or unpleasant. This implied fitting two BANOVAs including separately RTs of the C+ and C- induction patterns, then dividing them into two BANOVAs to isolate RTs corresponding to the pleasant and the unpleasant stimulus.

Regarding the BANOVA fitted on RTs of the C+ induction pattern, the Validity main effect was consistently found with shorter RTs for valid than invalid trials,  $BF_{10} > 1000$  for the BANOVA in which the data were not separated according to Pleasantness,  $BF_{10} > 1000$  for the BANOVA centered on RTs of the pleasant stimulus, and  $BF_{10} > 1000$ , for the BANOVA centered on RTs of the unpleasant stimulus. The corresponding mean RTs and standard errors are reported in Figure 3. These results suggest that when neutral stimuli were induced to gain, they captured more attention than the relevant stimulus, whether it was pleasant or unpleasant.

Surprisingly, regarding the BANOVAs fitted on RTs of the C- induction pattern, an absence of Validity main effect was observed. This was the case for the BANOVA in which data were not separated according to Pleasantness,  $BF_{01} = 7.19$ , for the BANOVA centered on RTs of the pleasant stimulus,  $BF_{01} = 10.64$ , and for the BANOVA centered on RTs of the unpleasant stimulus,  $BF_{01} = 7.58$ . The corresponding mean RTs and standard errors are reported in Figure 3. These results suggest that when neutral stimuli were induced to automatic loss, pleasant and unpleasant stimuli captured attention in the same way as neutral stimuli.

[Figure 3 near here]

## Discussion

### *Ratings of the intrinsic relevance level of stimuli*

In the first place, the check conducted on the ratings of the level of intrinsic relevance of stimuli allowed us to ensure that the pleasant stimulus and the unpleasant stimulus were rated, respectively, more pleasant and unpleasant than the other stimuli. Furthermore, this check evidenced that the stimuli submitted to the gain induction were rated more pleasant than when they were not induced or submitted to an automatic loss induction. In contrast, stimuli submitted to an automatic loss induction were rated more unpleasant than when they were not induced or submitted to a gain induction. This provides information on the relationship between intrinsic relevance and goal relevance by showing that experiencing the goal relevance of stimuli impacts their intrinsic relevance. This is consistent with the hypothesis that intrinsic relevance is the result of concatenated information about the goal relevance of stimuli in past experiences (Fournier & Koenig, 2023). Indeed, interaction with a goal-conducive stimulus would generate pleasure states (involving dopamine releases; Berridge, 2009; Berridge et al., 2009; Castro & Berridge, 2014; Ferreri et al., 2019), which would make the stimulus more generally pleasant for future interactions (Smith et al., 2011). Therefore, in the present experiment, our results suggest that intrinsic relevance was updated by an experience in which stimuli were conducive to or obstructed the achievement of a goal. The rationale that intrinsic relevance emerges from goal relevance might raise the question of the interest in appraising intrinsic relevance for the brain because one might argue that its appraisal is a waste of time and cognitive resources. However, considering the brain as a predictive machine (Hohwy, 2013) governed by the principle of minimizing uncertainty (Friston, 2005, 2010), it is advantageous to obtain rapid global information on stimulus relevance while the goal relevance is not yet available. Goal relevance would be a heavier appraisal than intrinsic relevance because it requires more information (Aue et al., 2007; Grandjean & Scherer, 2008; Lanctôt & Hess, 2007; Scherer,

2013). Thus, during the goal relevance appraisal, intrinsic relevance is already appraised to orientate the individual behaviors, which results in a gain of time and resources. The mechanism described here calls into question the notion of valence, widely used in the literature on emotions (Arioli et al., 2021; Lewis et al., 2019; Ossenfort & Isaacowitz, 2021; Yeung & Fernandes, 2021). The notion of valence proposed by Lewin (1938/1951) could be defined by a categorization of stimuli according to their position on a continuum ranging from negative (i.e., repulsion) to positive (i.e., attraction). However, the above mechanism assumes that the position of the stimuli on this continuum can vary according to their intrinsic relevance and their relevance to the goal. Thus, our results support the hypothesis that valence should be viewed as multidimensional based on the intrinsic relevance and the goal relevance of stimuli (Scherer, 2013; Scherer & Moors, 2019; Shuman et al., 2013).

### *Attentional capture and relevance*

Appraisal theories suggest that attention is captured by stimuli based on the combined effects of their intrinsic relevance, i.e., their pleasantness/unpleasantness nature, and their goal relevance, i.e., their conducive/obstructive nature with respect to the currently important goals of the organism (Grandjean & Scherer, 2014). However, these predictions were only tested in the case of relevant stimuli (intrinsically pleasant or unpleasant) overlaid by goal conduction (Fournier & Koenig, 2023, Submitted). In the present study, we proposed to test again this scenario using a gain induction for the goal conduction, but with adding an unavoidable goal obstruction to pleasant and unpleasant stimuli.

First, our results showed that relevant stimuli captured more attention than neutral stimuli. This result was in line with studies showing that attention is captured toward stimuli according to their intrinsic relevance (Brosch & Sharma, 2005; Fox, et al., 2007; LoBue & DeLoache, 2008; Öhman, 1993; Öhman & Mineka, 2001).

Second, our results showed that when contrasting relevant stimuli and neutral stimuli, attentional capture by relevant stimuli increased when they were induced to gain. This result seems to provide additional evidence in favor of the hypothesis that the relevance overlay of intrinsic relevance and goal-conducive relevance increases the attentional bias toward stimuli (Fournier & Koenig, 2023, Under review).

Third, our results showed that neutral stimuli captured more attention than relevant stimuli when neutral stimuli were submitted to the gain induction. This ensured the proper functioning of the goal induction that we manipulated and is consistent with the hypothesis that attention is captured by goal-relevant stimuli if their goal relevance is goal-conducive (Mazzietti et al., 2014; Mazzietti et al., 2015; Müller & Wentura, 2022; Vogt et al., 2011; Vogt et al., 2017).

However, our results revealed that attentional capture by relevant stimuli was canceled when they were induced to automatic loss. This result extends that of a previous study (Müller & Wentura, 2022), which showed an attentional effect for obstructive stimuli only when the obstruction could be avoided. In this regard, we proposed that this result was a consequence of either an insufficiently strong attentional effect or the fact that obstructive stimuli with unavoidable consequences definitely do not capture attention. Our results instead support the second hypothesis. Furthermore, we have shown that these stimuli can negate the initial ability of intrinsically relevant stimuli to capture attention. This could be interpreted as a protection against unnecessary persistence: Neutralizing the relevance of stimuli perceived as causing an obstruction to an unavoidable goal would help focussing processing resources on what can be done to remedy the situation (Rothermund, 2011).

In addition, our results showed that the attentional bias toward relevant stimuli was also reduced when neutral stimuli were induced to automatic loss, making relevant and induced-neutral stimuli captured attention in the same way. This result seems to qualify our previous interpretation: If the induction of an unavoidable obstruction decreases the ability of stimuli to

capture attention, an attentional capture by relevant stimuli should have been observed when neutral stimuli were induced to automatic loss. Some authors proposed that the presence of goal induction eliminates attentional bias generated by intrinsic relevance because all available resources are focused on the processing of goal-relevant stimuli (Correll et al., 2014; Pessoa et al., 2002; Pessoa et al. 2005). However, this interpretation is ruled out by the result showing an attentional bias toward relevant stimuli when stimuli were not induced (when the filler stimuli set was used in the induction task). The results obtained differed in two respects compared to the absence of attentional bias when neutral stimuli were induced to automatic loss. First, neutral stimuli were not induced to automatic loss. Considering that this parameter neutralizes the relevance of neutral stimuli, it is unlikely to explain the cancellation of the attentional bias in favor of relevant stimuli. Second, the stimuli were not presented in the induction task. One plausible interpretation would be that the presentation of the relevant stimuli in the induction task as stimuli that do not win or lose points ultimately attributes them to an irrelevance induction regarding goal relevance. Thus, the overlay of this goal irrelevance would conflict with their intrinsic relevance, so that they no longer captured attention.

This unexpected result provides new insight into the impact of relevance overlay on attention according to the predictive brain framework (Hohwy, 2013). As detailed in the introduction, relevance appraisal of a stimulus consists of a prediction based on prior beliefs emerging from past interactions with stimuli (Barrett, 2017). Thus, confronting the predictions of intrinsic relevance and goal relevance within the overlay provides information on whether to approach or avoid a stimulus (Fournier & Koenig, 2023, Under review). Roughly speaking, the likelihood that a stimulus should be either approached or avoided would be maximized based on whether a stimulus is appraised as both pleasant and goal conducive or both unpleasant and goal obstructive. While the likelihood that a stimulus should be both approached and avoided is maximized based on a stimulus is appraised as both pleasant and goal obstructive or both

unpleasant and goal conducive. All these relevance overlay schemes would increase attentional bias toward stimuli. Indeed, if the likelihood that a stimulus should be either approached or avoided is maximized, the stimulus would capture attention because it likely requires one of the two actions. In contrast, if the likelihood of a stimulus should be both approached and avoided is maximized, the stimulus would capture attention because it requires a further appraisal to reduce uncertainty about what action to take (Friston, 2005, 2010). The present results that attentional bias generated by the intrinsic relevance of stimuli could be canceled by goal irrelevance provide an additional relevance overlay scheme. A scheme in which a stimulus should likely require approach or avoidance because its intrinsic relevance (depending on whether it was pleasant or unpleasant) overlaid by a goal irrelevance, indicating that the stimulus should most likely not require any action. Due to the much more likelihood of the goal irrelevance prediction over the intrinsic relevance prediction, this would cancel the attentional capture toward the stimulus.

This result qualifies the interpretation of the finding that attentional bias toward the relevant stimuli was increased when they were attributed to the gain induction. Indeed, considering that goal irrelevance can overwhelm the impact of intrinsic relevance on attention, we cannot disentangle if the increase in attentional bias toward relevant stimuli reflects the impact of relevance overlay or the overwhelming effect of the gain induction. In other words, the increase in attentional bias toward relevant stimuli reflects the combined likelihood of the intrinsic relevance of stimuli and the gain induction about requiring action. Or only the overwhelming likelihood of the gain induction that the stimuli require to be approached.

The planned analysis on the unpacking of attentional bias depending on facilitated engagement and delayed disengagement could not be conducted. This was because participants' baseline RTs based on the pair contrasting the mixture stimulus with itself were not reliable. Initially, we assumed that mixing all the shapes and colors of relevant and neutral stimuli could



be considered a neutral stimulus. However, it ended up being a complex and degraded stimulus representing a mishmash of pieces of relevant and neutral shapes. We suggest that these longer RTs reflect heavier processing of this stimulus due to its complexity and degradation (Lavie & Fockert, 2003).

In sum, this study provides new insights regarding the impact of relevance overlay on attention. More specifically, it showed that goal relevance can overwhelm the intrinsic relevance impact on attention.

### **Limits and perspective**

A major limitation of the present study relies on the use of only one exemplar of stimuli per intrinsic relevance category (Judd et al., 2012). Because the experimental design required participants to retain the stimuli and the attributed induction, the number of stimuli had to be limited. From our point of view, this does not detract from the value of the study, which provides new insights into the topic of the impact of relevance overlay on attention in specific cases of relevance overlay. In future research, it would be interesting to try to generalize these results using multiple stimuli per intrinsic relevance category (i.e., multiple pleasant and unpleasant stimuli). This could be done by attempting to observe the consistency of the effect of goal relevance induction on attention to categories of stimuli rather than to one exemplar of a category (for instance, presenting multiple pleasant – or unpleasant – stimuli in the induction task and instructing participants that they would earn points if any of the pleasant – or unpleasant – stimuli appear). The constraints of the experimental design mentioned above also prevented us from testing the attentional impact of other types of induction such as automatic gain or avoidable loss. For us, the interest of our study was to show that some inductions can cancel the ability of stimuli to capture attention. Proposing other types of inductions, such as avoidable loss or automatic gain, could certainly allow us to better understand the functioning of this attention cancellation phenomenon. Another limitation of the present study lies in the

inability to analyze the unpacking of attentional bias because the RTs baseline of participants was not reliable. This emphasizes the importance to propose a pair of stimuli not complex or degraded to obtain this baseline.

### **Acknowledgments**

Thanks to Isabella Zsoldos for our inspiring discussions.

### **Declaration of interest statement**

The authors report no conflict of interest. We warrant that participants in this experiment were treated in accordance with the ethical standards of APA. This work was supported by the doctoral grant to Hippolyte Fournier from the Banque Publique d'Investissement (BPI) under grant THERADIA.

### **Data availability statement**

Data, preprocessing, code analyses, and material from the study are available through the Open Science Framework repository: <https://osf.io/ue6st/>. Data were analyzed using R (version 1.4).

### **References**

- Arioli, M., Basso, G., Poggi, P., & Canessa, N. (2021). Fronto-temporal brain activity and connectivity track implicit attention to positive and negative social words in a novel socio-emotional Stroop task. *NeuroImage*, 226, 117580. <https://doi.org/10.1016/j.neuroimage.2020.117580>
- Aue, T., Flykt, A., & Scherer, K. R. (2007). First evidence for differential and sequential efferent effects of stimulus relevance and goal conduciveness appraisal. *Biological Psychology*, 74(3), 347-357. <https://doi.org/10.1016/j.biopsycho.2006.09.001>
- Aupperle, R. L., Melrose, A. J., Francisco, A., Paulus, M. P., & Stein, M. B. (2015). Neural substrates of approach-avoidance conflict decision-making. *Human Brain Mapping*, 36(2), 449-462. <https://doi.org/10.1002/hbm.22639>

- Barrett, L. F. (2017). The theory of constructed emotion: an active inference account of interoception and categorization. *Social cognitive and affective neuroscience*, 12(1), 1-23. <https://doi.org/10.1093/scan/nsx060>
- Berridge, K. C. (2009). 'Liking' and 'wanting' food rewards: brain substrates and roles in eating disorders. *Physiology & behavior*, 97(5), 537-550. <https://doi.org/10.1016/j.physbeh.2009.02.044>
- Berridge, K. C., Robinson, T. E., & Aldridge, J. W. (2009). Dissecting components of reward: 'liking', 'wanting', and learning. *Current Opinion in Pharmacology*, 9(1), 65-73. <https://doi.org/10.1016/j.coph.2008.12.014>
- Brosch, T., Pourtois, G., & Sander, D. (2010). The perception and categorization of emotional stimuli: A review. *Cognition and Emotion*, 24(3), 377-400. <https://doi.org/10.1080/02699930902975754>
- Brosch, T., Sander, D., Pourtois, G., & Scherer, K. R. (2008). Beyond Fear: Rapid Spatial Orienting Toward Positive Emotional Stimuli. *Psychological Science*, 19(4), 362–370. <https://doi.org/10.1111/j.1467-9280.2008.02094.x>
- Brosch, T., Sander, D., & Scherer, K. R. (2007). That baby caught my eye... Attention capture by infant faces. *Emotion*, 7(3), 685–689. <https://doi.org/10.1037/1528-3542.7.3.685>
- Brosch, T., & Sharma, D. (2005). The Role of Fear-Relevant Stimuli in Visual Search: A Comparison of Phylogenetic and Ontogenetic Stimuli. *Emotion*, 5(3), 360–364. <https://doi.org/10.1037/1528-3542.5.3.360>
- Carlson, J. M., & Fang, L. (2020). The stability and reliability of attentional bias measures in the dot-probe task: Evidence from both traditional mean bias scores and trial-level bias scores. *Motivation and Emotion*, 44(5), 657-669. <https://doi.org/10.1007/s11031-020-09834-6>

Castro, D. C., & Berridge, K. C. (2014). Advances in the neurobiological bases for food 'liking versus 'wanting'. *Physiology & Behavior*, *136*, 22-30.

<https://doi.org/10.1016/j.physbeh.2014.05.022>

Chica, A. B., Martín-Arévalo, E., Botta, F., & Lupiáñez, J. (2014). The Spatial Orienting paradigm: How to design and interpret spatial attention experiments. *Neuroscience & Biobehavioral Reviews*, *40*, 35-51. <https://doi.org/10.1016/j.neubiorev.2014.01.002>

Correll, J., Guillermo, S., & Vogt, J. (2014). On the flexibility of attention to race. *Journal of Experimental Social Psychology*, *55*, 74-79. <https://doi.org/10.1016/j.jesp.2014.05.013>

de Jong, P. J., Koster, E. H., van Wees, R., & Martens, S. (2009). Emotional facial expressions and the attentional blink: Attenuated blink for angry and happy faces irrespective of social anxiety. *Cognition and Emotion*, *23*(8), 1640-1652.

<https://doi.org/10.1080/02699930802490227>

Ferreri, L., Mas-Herrero, E., Zatorre, R. J., Ripollés, P., Gomez-Andres, A., Alicart, H., Olivé, G., Marco-Pallarés, J., Antonijoan, R.M., Valle, M., Riba, J., and Rodriguez-Fornells, A. (2019). Dopamine modulates the reward experiences elicited by music. *Proceedings of the National Academy of Sciences*, *116*(9), 3793-3798.

<https://doi.org/10.1073/pnas.181187811>

Forrest, J., Vogt, J., McDonald, C., Searle, C., & Sakaki, M. (2022). Blind to threat: The presence of temporary goals prevents attention to imminent threat already at early stages of attention allocation. *Motivation Science*, *8*(3), 239-251.

<https://doi.org/10.1037/mot0000261>

Fournier, H., & Koenig, O. (2022). Combined effects of intrinsic and goal relevances on attention and action tendency during the emotional episode. *Emotion*, *23*(2), 425-436.

<https://doi.org/10.1037/emo0001081>

- Fournier, H., & Koenig, O. (Submitted). From emotion to action tendency: Attentional effects, and their time course, of agonistic and antagonistic intrinsic and goal relevances.
- Fox, E., Griggs, L., & Mouchlianitis, E. (2007). The detection of fear-relevant stimuli: Are guns noticed as quickly as snakes? *Emotion*, 7(4), 691-696. <https://doi.org/10.1037/1528-3542.7.4.691>
- Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological sciences*, 360(1456), 815-836. <https://doi.org/10.1098/rstb.2005.1622>
- Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, 11(2), 127-138. <https://doi.org/10.1038/nrn2787>
- Fromberger, P., Jordan, K., von Herder, J., Steinkrauss, H., Nemetschek, R., Stolpmann, G., & Müller, J. L. (2012). Initial orienting towards sexually relevant stimuli: Preliminary evidence from eye movement measures. *Archives of sexual behavior*, 41(4), 919-928. <https://doi.org/10.1007/s10508-011-9816-3>
- Grandjean, D., Sander, D., & Scherer, K. R. (2008). Conscious emotional experience emerges as a function of multilevel, appraisal-driven response synchronization. *Consciousness and Cognition*, 17(2), 484-495. <https://doi.org/10.1016/j.concog.2008.03.019>
- Grandjean, D., & Scherer, K. R. (2014). Chapter 2. Theory of cognitive appraisal and dynamics of emotional processes. In Sander, D. (Ed.). *Treatise on the psychology of emotions*, 51-87. <https://doi.org/10.3917/dunod.sande.2014.01.0051>
- Haxby, J. V., Parasuraman, R., Lalonde, F., & Abboud, H. (1993). SuperLab: General-purpose Macintosh software for human experimental psychology and psychological testing. *Behavior Research Methods, Instruments, & Computers*, 25(3), 400-405. <https://doi.org/10.3758/BF03204531>
- Hohwy, J. (2013). Part 1: The mechanism. In Hohwy, J. (Ed.). *The predictive mind*. 13-38. Oxford, UK: Oxford University Press.

- Ito, R., & Lee, A. C. (2016). The role of the hippocampus in approach-avoidance conflict decision-making: evidence from rodent and human studies. *Behavioural Brain Research*, 313, 345-357. <https://doi.org/10.1016/j.bbr.2016.07.039>
- Jiang, Y., Costello, P., Fang, F., Huang, M., & He, S. (2006). A gender-and sexual orientation-dependent spatial attentional effect of invisible images. *Proceedings of the National Academy of Sciences*, 103(45), 17048-17052. <https://doi.org/10.1073/pnas.0605678103>
- Joormann, J., & Gotlib, I. H. (2007). Selective attention to emotional faces following recovery from depression. *Journal of abnormal psychology*, 116(1), 80. <https://doi.org/10.1037/0021-843X.116.1.80>
- Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology*, 103(1), 54-69. <https://doi.org/10.1037/a0028347>
- Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American statistical association*, 90(430), 773-795. <https://doi.org/10.1080/01621459.1995.10476572>
- Koster, E. H., Crombez, G., Verschuere, B., & De Houwer, J. (2004). Selective attention to threat in the dot probe paradigm: Differentiating vigilance and difficulty to disengage. *Behaviour research and therapy*, 42(10), 1183-1192. <https://doi.org/10.1016/j.brat.2003.08.001>
- Kruschke, J. K. (2013). Bayesian estimation supersedes the t-test. *Journal of Experimental Psychology: General*, 142(2), 573. <https://doi.org/10.1037/a0029146>
- Kruschke, J. (2014). Tools in the Trunk. In Kruschke, J. (Ed.). *Doing Bayesian Data Analysis* (2<sup>nd</sup> ed., pp. 721-736). Academic Press. <https://doi.org/10.1016/B978-0-12-405888-0.00025-8>

- Kruschke, J. K. (2018). Rejecting or accepting parameter values in Bayesian estimation. *Advances in Methods and Practices in Psychological Science*, 1(2), 270-280.  
<https://doi.org/10.1177/2515245918771304>
- Lanctôt, N., & Hess, U. (2007). The timing of appraisals. *Emotion*, 7(1), 207–212. <https://doi.org/10.1037/1528-3542.7.1.207>
- Lavie, N., De Fockert, J.W. (2003). Contrasting effects of sensory limits and capacity limits in visual selective attention. *Perception & Psychophysics* 65, 202–212.  
<https://doi.org/10.3758/BF03194795>
- Lewin, K. (1951). *Field theory in social science: Selected theoretical papers* (D. Cartwright, Ed.). Greenwood Press. (Original work published 1938)
- Lewis, E. J., Blanco, I., Raila, H., & Joormann, J. (2019). Does repetitive negative thinking affect attention? Differential effects of worry and rumination on attention to emotional stimuli. *Emotion*, 19(8), 1450-1462. <https://doi.org/10.1037/emo0000535>
- LoBue, V., & DeLoache, J. S. (2008). Detecting the snake in the grass: Attention to fear-relevant stimuli by adults and young children. *Psychological science*, 19(3), 284-289.  
<https://doi.org/10.1111/j.1467-9280.2008.02081.x>
- Lykins, A. D., Meana, M., & Strauss, G. P. (2008). Sex differences in visual attention to erotic and non-erotic stimuli. *Archives of sexual behavior*, 37(2), 219-228.  
<https://doi.org/10.1007/s10508-007-9208-x>
- Mazzietti, A., Sellem, V., & Koenig, O. (2014). From stimulus-driven to appraisal-driven attention: Towards differential effects of goal relevance and goal relatedness on attention? *Cognition and Emotion*, 28(8), 1483-1492.  
<https://doi.org/10.1080/02699931.2014.884488>
- Mazzietti, A., Sellem, V., & Koenig, O. (2015, July 9). To see, or not to see: Dynamic attentional interference by non-consciously perceived relevant stimuli [Conference

- session]. Bi-Annual Conference of the International Society for Research On Emotion, Geneva, Switzerland. <https://www.unige.ch/cisa/isre2015/sites/default/files/Mazzietti.pdf>
- Müller, S., Rothermund, K., & Wentura, D. (2016). Relevance drives attention: Attentional bias for gain-and loss-related stimuli is driven by delayed disengagement. *Quarterly Journal of Experimental Psychology*, 69(4), 752-763.  
<https://doi.org/10.1080/17470218.2015.1049624>
- Müller, P., & Wentura, D. (2022). Undeserved reward but not unavoidable loss biases attention: Personal control moderates evaluative attentional biases in the additional-singleton paradigm. *Quarterly Journal of Experimental Psychology*, 17470218221099125.  
<https://doi.org/10.1177/17470218221099125>
- Öhman, A. (1993). Fear and anxiety as emotional phenomena: Clinical phenomenology, evolutionary perspectives, and information-processing mechanisms. In M. Lewis & J. M. Haviland (Eds.), *Handbook of emotions* (pp. 511–536). The Guilford Press.
- Öhman, A., & Mineka, S. (2001). Fears, phobias, and preparedness: toward an evolved module of fear and fear learning. *Psychological Review*, 108(3), 483-522.  
<https://doi.org/10.1037/0033-295X.108.3.483>
- Ossenfort, K. L., & Isaacowitz, D. M. (2021). Spatial attention to arousing emotional stimuli in younger and older adults. *Motivation and Emotion*, 45(6), 790-797.  
<https://doi.org/10.1007/s11031-021-09899-x>
- Pessoa, L., McKenna, M., Gutierrez, E., & Ungerleider, L. G. (2002). Neural processing of emotional faces requires attention. *Proceedings of the National Academy of Sciences*, 99(17), 11458–11463. <https://doi.org/10.1073/pnas.172403899>
- Pessoa, L., Padmala, S., & Morland, T. (2005). Fate of unattended fearful faces in the amygdala is determined by both attentional resources and cognitive modulation. *NeuroImage*, 28(1), 249–255. <https://doi.org/10.1016/j.neuroimage.2005.05.048>



- Posner, M. I. (1980). Orienting of attention. *Quarterly journal of experimental psychology*, 32(1), 3-25. <https://doi.org/10.1080/00335558008248231>
- Rothermund, K. (2011). Counter-regulation and control-dependency: Affective processing biases in the service of action regulation. *Social Psychology*, 42(1), 56–66. <https://doi.org/10.1027/1864-9335/a000043>
- Scherer, K. R. (2013). The nature and dynamics of relevance and valence appraisals: Theoretical advances and recent evidence. *Emotion review*, 5(2), 150-162. <https://doi.org/10.1177/1754073912468166>
- Scherer, K. R., & Moors, A. (2019). The emotion process: Event appraisal and component differentiation. *Annual Review of Psychology*, 70, 719-745. <https://doi.org/10.1146/annurev-psych-122216-011854>
- Schönbrodt, F. D., & Wagenmakers, E. J. (2018). Bayes factor design analysis: Planning for compelling evidence. *Psychonomic bulletin & review*, 25(1), 128-142. <https://doi.org/10.3758/s13423-017-1230-y>
- Shuman, V., Sander, D., & Scherer, K. R. (2013). Levels of valence. *Frontiers in Psychology*, 4, 261. <https://doi.org/10.3389/fpsyg.2013.00261>
- Spiering, M., & Everaerd, W. (2007). The sexual unconscious. In E. Janssen (Ed.), *The psychophysiology of sex* (pp. 166–184). Indiana University Press.
- Stefan, A. M., Gronau, Q. F., Schönbrodt, F. D., & Wagenmakers, E. J. (2019). A tutorial on Bayes Factor Design Analysis using an informed prior. *Behavior Research Methods*, 51(3), 1042-1058. <https://doi.org/10.3758/s13428-018-01189-8>
- Steimke, R., Nomi, J. S., Calhoun, V. D., Stelzel, C., Paschke, L. M., Gaschler, R., ... & Uddin, L. Q. (2017). Salience network dynamics underlying successful resistance of temptation. *Social cognitive and affective neuroscience*, 12(12), 1928-1939. <https://doi.org/10.1093/scan/nsx123>

- Straube, T., Mentzel, H. J., & Miltner, W. H. (2006). Neural mechanisms of automatic and direct processing of phobogenic stimuli in specific phobia. *Biological psychiatry*, *59*(2), 162-170. <https://doi.org/10.1016/j.biopsych.2005.06.013>
- Sun, L., Ding, C., Xu, M., Diao, L., & Yang, D. (2017). Engagement attentional bias toward value-associated stimuli. *Current Psychology*, *36*(4), 747-754. <https://doi.org/10.1007/s12144-016-9462-y>
- Vogt, J., De Houwer, J., Crombez, G., & Van Damme, S. (2013). Competing for attentional priority: temporary goals versus threats. *Emotion*, *13*(3), 587. <https://doi.org/10.1037/a0027204>
- Vogt, J., Koster, E. H., & De Houwer, J. (2017). Safety first: Instrumentality for reaching safety determines attention allocation under threat. *Emotion*, *17*(3), 528. <https://doi.org/10.1037/emo0000251>
- Vogt, J., Lozo, L., Koster, E. H., & De Houwer, J. (2011). On the role of goal relevance in emotional attention: Disgust evokes early attention to cleanliness. *Cognition and Emotion*, *25*(3), 466-477. <https://doi.org/10.1080/02699931.2010.532613>
- Yeung, R. C., & Fernandes, M. A. (2021). Divided attention at encoding or retrieval interferes with emotionally enhanced memory for words. *Memory*, *29*(3), 284-297. <https://doi.org/10.1080/09658211.2021.1887896>

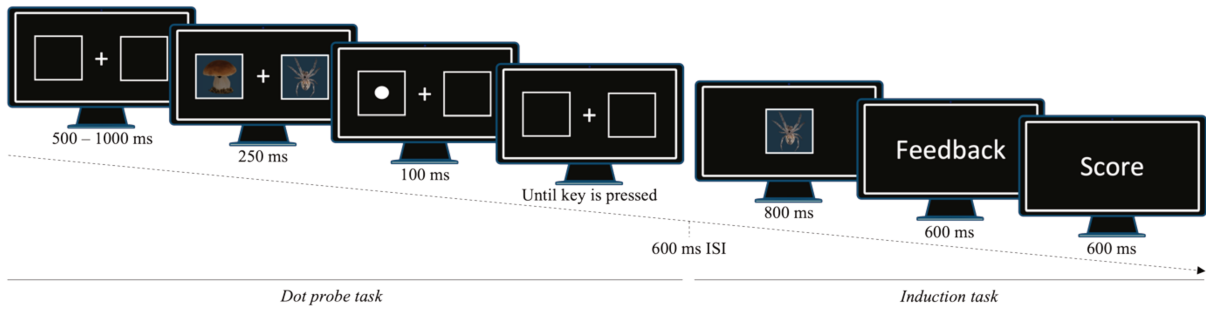
**Table 1**

*Descriptive statistics of the evaluation of the pleasantness or unpleasantness of stimuli according to whether they were induced to gain or to automatic loss, or in the absence of induction.*

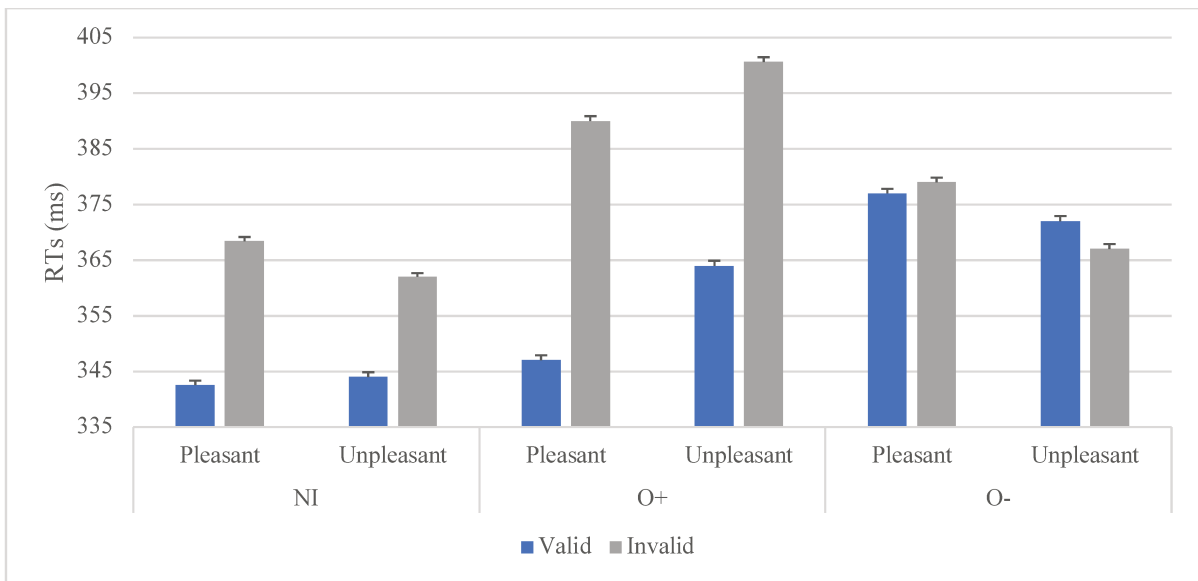
Stimulus	Induction type					
	No induction		Gain		Automatic loss	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Neutral 1	4.48	0.79	5.27	1.22	3.50	1.38
Neutral 2	3.87	0.68	4.97	1.72	3.05	1.11
Pleasant	5.20	1.55	5.98	1.25	4.98	1.56
Unpleasant	2.69	1.39	3.30	1.56	2.50	1.34

*Note. The scale ranges from 1 (very unpleasant) to 7 (very pleasant), with 4 being considered neutral.*

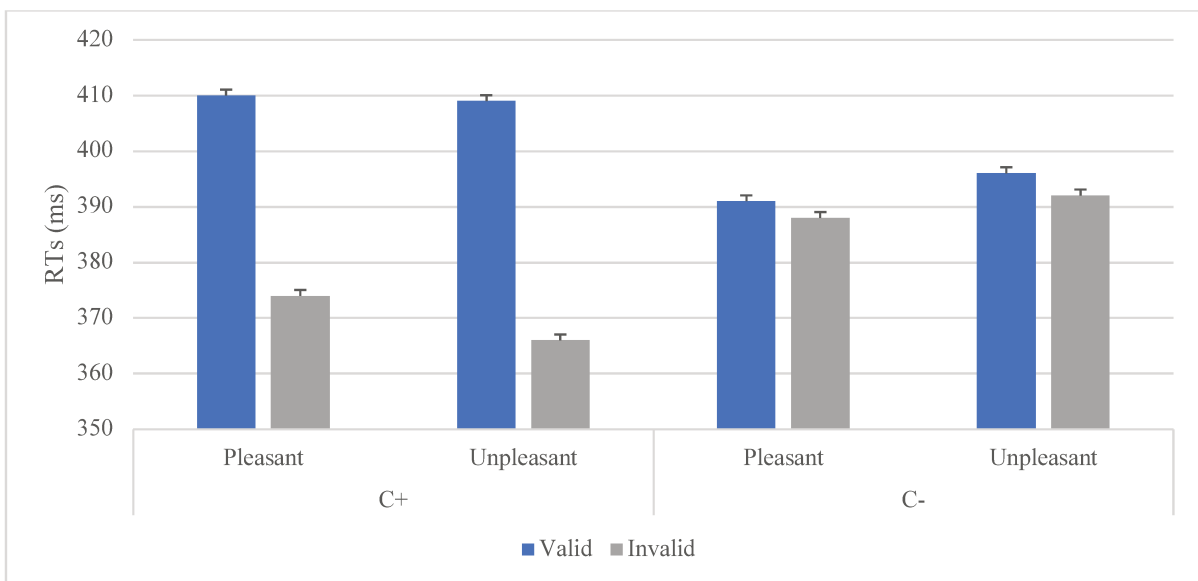
**Figure 1**



**Figure 2**



**Figure 3**



**Figure 1.** *Schematic Overview of the Experimental Design*

*Note.* The first four screens correspond to the dot-probe task: the fixation cross with two empty squares, the cues in the squares, the dot in one of the empty squares, and the response screen with two empty screens. The last three screens correspond to the induction task: a central square with a stimulus requiring pressing a key in the case the stimulus induces conditional gain, the feedback screen, and the screen with the updated score.

**Figure 2.** *Mean response times in the overlay conditions, where intrinsically relevant (pleasant and unpleasant) stimuli were attributed to no induction (NI), to induction of gain (O+) or automatic loss (O-). Each stimulus pair of the dot-probe task consisted of an induced pleasant or unpleasant stimulus and a neutral stimulus. In valid trials, the dot appeared at the location of the relevant stimulus. No induction was used in the NI condition.*

*Note.* Vertical lines depict Standard Errors.

**Figure 3.** *Mean response times in the competition conditions, where neutral stimuli were attributed to induction of gain (C+) or automatic loss (C-). Each stimulus pair of the dot-probe task consisted of an intrinsically relevant (pleasant or unpleasant) stimulus and an induced, neutral stimulus. In valid trials, the dot appeared at the location of the intrinsically relevant stimulus.*

*Note.* Vertical lines depict Standard Errors.

## **Chapitre 6 - Article : “The eye-dot-probe task: An innovative lead to overcome the crisis of the attentional bias toward threat”**

Ce chapitre décrit le quatrième article que nous avons écrit qui est actuellement en soumission dans une revue, dont la référence est la suivante :

Fournier, H., Lefèvre, E., Guerin, E., & Koenig, O. (Submitted). The eye dot-probe task: an innovative lead to overcome the crisis on attentional bias toward threat.

## Résumé

La recherche clinique suggère que les troubles anxieux sont caractérisés par un biais attentionnel pour les stimuli liés à la menace. Néanmoins, ces biais ne sont pas systématiquement observés. Ces résultats mitigés pourraient s'expliquer en partie par la faible fiabilité de la tâche du dot-probe, le paradigme le plus couramment utilisé pour mesurer le biais attentionnel à la menace. Pour remédier à ce problème, plusieurs calculs d'indices attentionnels ont été proposés dans la littérature afin d'augmenter la fiabilité des mesures obtenues avec la tâche de dot-probe. Dans la présente étude, nous avons émis l'hypothèse que la capacité de la tâche de dot-probe à détecter les biais attentionnels vers la menace pouvait être améliorée en demandant aux participants de répondre par des mouvements oculaires plutôt que par des pressions manuelles sur des touches. Cinquante-quatre participants ont effectué une tâche classique de dot-probe (réponse donnée en appuyant sur une touche) et une tâche de eye-dot-probe (réponse donnée par des mouvements oculaires) dans laquelle des visages menaçants et neutres étaient mis en compétition. Huit indices attentionnels ont été calculés pour chaque tâche. Avant les tâches, les participants ont été invités à remplir le questionnaire STAI pour mesurer leur trait d'anxiété trait et leur état d'anxiété. Pour tous les indices attentionnels calculés dans chaque tâche, nous avons vérifié s'ils étaient mieux expliqués par l'état d'anxiété des participants, par leur trait d'anxiété, par la combinaison des deux ou n'étaient expliqués par aucune des deux mesures. Les résultats ont montré que la tâche d'eye-dot-probe semble être le meilleur candidat pour évaluer le biais attentionnel lié à la menace dans l'anxiété. En outre, les résultats soulignent l'importance de prendre en compte les effets combinés du trait d'anxiété et de l'état d'anxiété pour évaluer le biais attentionnel à la menace au moyen de la tâche de dot-probe classique.

**The eye-dot-probe task: An innovative lead to overcome the crisis of the attentional  
bias toward threat.**

Hippolyte Fournier, Élise Lefèvre, Émilie Guerin, and Olivier Koenig

Laboratoire EMC, Université Lumière Lyon 2, Bron, France

**Author note**

Hippolyte Fournier  <https://orcid.org/0000-0002-7861-2387>

Élise Lefèvre  <https://orcid.org/0000-0001-7935-2938>

Olivier Koenig  <https://orcid.org/0000-0002-5124-1889>

We have no conflicts of interest to disclose. We guarantee that participants in this experiment were treated in accordance with the ethical standards of APA. This research was made possible thanks to a doctoral grant to Hippolyte Fournier from the Banque Publique d'Investissement (BPI) under the grant agreement THERADIA.

All data, analysis code and material from the study reported in this article are available through the Open Science Framework repository: <https://osf.io/gdvxp/>.

Correspondence should be addressed to Hippolyte Fournier, Laboratoire EMC, Université Lumière Lyon 2, 5 Avenue Pierre Mendès France, Bron 69676, France. Email: [h.fournier@univ-lyon2.fr](mailto:h.fournier@univ-lyon2.fr)



### **Abstract**

Clinical research suggests that anxiety disorders are characterized by attentional bias to threat-related stimuli. Nonetheless, such biases are not systematically observed. These mixed results could be partially explained by the fleeting reliability of the dot-probe task, the most common paradigm used to measure attentional bias to threat. To address this problem, several attentional indices calculations were proposed in the literature to increase the reliability of the measurement obtained with the dot-probe task. In the present study, we hypothesized that the ability of the dot-probe task to detect attentional biases to threat could be improved by having participants respond with eye movements rather than manual key presses. Fifty-four participants performed a classic dot-probe task (response given by pressing a key) and an eye-dot-probe task (response given by eye movements) in which threatening and neutral faces were put in competition. Eight attentional indices were calculated in each task. Before the tasks, participants were asked to complete the STAI questionnaire to measure their trait anxiety and their state anxiety. For all attentional indices calculated in each task, we tested whether they were best explained by participants' state anxiety level, trait anxiety, both, or neither anxiety level. The findings showed that the eye-dot-probe task appears to be the best candidate for assessing attentional bias to anxiety-related threat. Moreover, the results emphasized the importance of considering the combined effects of both trait and state anxiety to assess the attentional bias to threat in the classic dot-probe task.

### **Public Significance Statement**

This study proposed a new version of a widely used computerized task measuring attentional bias to anxiety-related threat. In this version, responses are given by eye movements rather than pressing a key. This new version provided more robust results compared to the conventional version.

*Keywords:* attentional bias, dot-probe, anxiety, eye-tracker, threat bias

### **The eye-dot-probe task: An innovative lead to overcome the crisis of the attentional bias toward threat**

Over the past decades, it has been repeatedly demonstrated that the emotional nature of stimuli would capture attention in order to facilitate the individual's adaptation to his environment (Bar-Haim et al., 2010). For instance, threat-related stimuli would be prioritized by attention because they would likely require to be avoided to achieve the well-being and survival of individuals (Öhman et al., 2001). However, in some cases, this attentional prioritization can become maladaptive, such as in individuals with anxiety disorders who show a greater attentional bias (Bar-Haim et al., 2007; Waters et al., 2010). According to the cognitive model of anxiety disorders, this exacerbated attentional bias would be manifested in this population by an overestimation of the negative impact or the seriousness of events (Milne et al., 2019). Being focused on the threat stimulus would increase the anxiety level and anxiety would increase the relevance of threat-related events in the environment.

Most studies that have assessed the relationship between attentional bias to threat-related stimuli and anxiety disorder have used the dot-probe task (Macleod et al., 1986). This task is defined by the appearance of three screens: a fixation cross, the simultaneous appearance of a relevant stimulus and a neutral stimulus on either side of the fixation cross, and a dot, replacing one of the two stimuli. Participants are instructed to stare at the fixation cross during the whole task and to indicate as quickly and accurately as possible the side of the dot appearance by pressing a key. Trials are defined as valid if the dot appears on the relevant stimuli side, and as invalid, if the dot appears on the neutral stimuli side. The idea is that when both stimuli appear, attention is captured by the more relevant stimulus. When the dot appears on the relevant stimulus side, attention is already captured on that side and the response time is faster. Attentional Bias Indexes (ABI) are usually calculated to report the results of the dot-probe task.

ABI are calculated by subtracting the valid trials' mean response times from the invalid trials' mean response times (Macleod et al., 1986).

Although many studies have evidenced an attentional bias toward threat-related stimuli in individuals with anxiety disorders (Bar-Haim et al., 2007; Waters et al., 2010), these results have not been consistently observed (Kruijt et al., 2019). Some authors believe that the variability in response times between trials of the dot-probe task does not allow for the calculation of an ABI that truly represents the level of attentional bias to threat (see, e.g., Schmukle, 2005). To address this issue, we propose to target two areas of improvement: increasing the ability of the dot-probe task to evidence the attentional bias toward threat by increasing the effect size and reducing the variability in response times between trials. Regarding the first area of improvement, one can imagine that it can be operationalized thanks to a finer response mode than pressing a key. For instance, by indicating the dot location with the gaze direction. The instructions would be similar to those of the classic dot-probe task except that participants must move their gaze to the location of the dot as quickly as possible when it appears. We believe that using eye movements would result in more reliable response time differences between valid and invalid trials than using response keys. This hypothesis was based on a logical argument that eye movement is directly related to attentional capture. Indeed, the appearance of a relevant stimulus in the visual field would generate shift in gaze and a maintenance of the position of the eyes at the level of its region of appearance (Pool et al., 2019).

Regarding the second area of improvement, other authors have stated that the variability in response times could be explained by the still theoretically unclear relationship between anxiety and attentional bias. For example, some of the literature suggests that anxiety can generate an attentional bias toward threat, but also in the opposite direction (Zvielli et al., 2015). This could be reflected in the temporal response of the classic dot probe task without being considered. To

account for these phenomenon, new ways of calculating attentional indices for the dot-probe task have been proposed. For instance, the Attentional Bias Variability index (ABV; Iacoviello et al., 2014; Naim et al., 2015), was proposed to measure the attentional bias variability through the task across sequential sub-blocks. The Trial Level-Bias Score (TL-BS; Zvielli et al., 2015) was proposed to define five indices calculated from the differences in RT between invalid and nearby valid trials for each participant. Among the five TL-BS indices, two measure the mean TL-BS toward threat and away from threat. The two other indices measure the peak of TL-BS toward threat and away from threat, whereas the last one measures the variability across TL-BS. According to Price et al. (2015) and Molloy and Anderson (2020), ABV and TL-BS are the most reliable index calculations for measuring attentional bias in the dot-probe task.

Another cause of the poor reliability of the measure of the attentional bias to anxiety-related threat in the dot-probe task could be related to the lack of clarity in the type of anxiety studied. Anxiety can be characterized by trait anxiety, i.e., the general tendency to experience anxiety (which defines anxiety disorders), or state anxiety, i.e., the current level of anxiety. Some studies have shown that the attentional bias to threat is related to trait anxiety (Bar-Haim et al., 2007; Waters et al., 2010), while others have reported that the attentional bias was related to state anxiety (Fox & Knight, 2005; Quigley et al., 2012). Still others have failed to show any relationship between attentional bias to threat and trait or state anxiety (Carlson & Fang, 2020). However, to our knowledge, no study has assessed the relationship between the attentional bias to threat and the combined effects of trait anxiety and state anxiety. It may be interesting to consider the combined effects of trait anxiety and state anxiety, as these two pieces of information may be complementary. Thus, a participant with a mean value for both trait and state anxiety might be considered to have an attentional bias toward threat.

The present study aimed to improve the dot-probe task reliability. More specifically, we searched to identify the best way to assess attentional biases toward threat in the dot-probe task

by considering (1) the modality of response (keyboard vs. eye movement); (2) the type of calculation of the attentional index; and (3) the anxiety measure (trait anxiety vs. state anxiety vs. a combination of both). To this end, participants performed two versions of the dot-probe task in which threatening faces and neutral faces were put in competition: a “classic dot-probe task” version, and an “eye-dot-probe task” version. In the eye-dot-probe task, participants were asked to indicate the dot position with an eye movement. The attentional indices discussed above were calculated from the data of the classic dot-probe task and the eye-dot-probe task. Before performing the tasks, participants were asked to complete a State-Trait Anxiety Inventory (STAI; Gauthier & Bouchard, 1993) questionnaire to measure their level of trait anxiety and state anxiety. From the STAI data, a trait anxiety score, a state anxiety score, and a composite trait and state anxiety score were calculated. Using a data-driven method, we investigated whether each attentional index in each task could be best explained by the state anxiety score, the trait anxiety score, or the composite trait and state anxiety score, and in a reliable manner.

## **Method**

### **Participants and procedure**

Fifty-four students (38 women, 16 men;  $M = 20.9$  years,  $SD = 2.2$ ) from Lyon University participated in the study. Twenty-four were undergraduate students (44.44%) and 30 were master's students (55.56 %). The race and ethnicity of the participants were not collected due to the French law concerning the collection of personal data.

Participants had a normal or corrected-to-normal vision. The exclusion criterion for the study was a diagnosis of epilepsy. The minimum sample size to perform linear regression was defined as 25 based on the Akaike Information Criterion (AIC; Burnham & Anderson, 2005), a method considered more robust than  $R^2$  (Jenkins & Quintana-Ascencio, 2020). To propose a sample

size consistent with the study of Molloy and Anderson (2020), the sample size was upped to 65 participants.

A classic dot-probe task and an eye-dot-probe task were proposed to participants. Before performing the tasks, participants filled out a consent form and the STAI (Gauthier & Bouchard, 1993) to obtain a measure of their trait anxiety (STAI-T) and state anxiety (STAI-S). Half of the participants started with the classic dot-probe task and the other half with the eye-dot-probe task. The tasks were programmed using OpenSesame 3.3 software (Mathôt et al., 2012). Data acquisition of eye movements in the eye-dot-probe was recorded using an Eye-link 1000 device with a 1000 Hz sampling rate. A five-point calibration was performed. Participants sat 102 cm away from the screen (1920\*1080 pixels display) in a soundproof room with the lights off. Each trial of the two tasks consisted of the successive appearance of three events presented on a gray background. First, a white fixation cross (10\*10 pixels) with two empty white squares (678\*678 px) centered at 404 pixels left and right of the cross appeared for a randomly defined time between 500 and 1000 ms. Second, a face (677\*677 pixels) appeared in the center of each square for 250 ms. Face pairs contrasted a threatening facial expression with a neutral facial expression of the same individual among 10 selected face pairs from the Chicago face database (Ma et al., 2015; see Table S1 of supplemental material for more details on face pairs). Third, a dot (20\*20 pixels) appeared in the center of one of the squares until the participant responded. In both tasks, participants were instructed to answer as quickly and precisely as possible (Chica et al., 2014).

In the classic dot-probe task, participants were asked to stare at the fixation cross throughout the whole experiment and to respond using a French, AZERTY, keyboard, with the “s” key if the dot appeared in the left square, and with the “m” key if the dot appeared in the right square. Classically, RTs for the classic dot-probe task were defined by the delay between the dot appearance and the key pressing.

In the eye-dot-probe task, participants were asked to stare at the fixation cross and to respond with their eye movements, moving their gaze toward the dot. RTs for the eye-dot-probe task were defined by the delay between the dot appearance and the entry of the gaze into a 44\*44 pixel area of interest (AOI) around the dot. To ensure that participants redirected their gaze to the center of the screen at the end of trials, a new trial started only if participants fixed a 32\*32 pixel AOI around the fixation cross.

In both tasks, each of the ten face pairs appeared 8 times with the threatening face in the left square and 8 times with the threatening face in the right square, for a total number of 160 trials. Of these trials, 80 were valid trials (i.e., the dot replaced the threatening face), and 80 were invalid trials (i.e., dot apparition at the side of the neutral face). The order of trials was randomized for each participant.

### **Data analysis**

Twelve participants were excluded because of a program bug or data recording problems.

Regarding the preprocessing of RTs in the dot-probe task, several methods are commonly used in the literature (Carlson & Fang, 2020; Molloy & Anderson, 2020; Price et al., 2015). One method consists in discarding the extreme RTs defined as below 150ms or 300ms and above 750ms or 1100ms. A second method (Winsorizing method) replaces RTs exceeding the extreme RT thresholds with the extreme RT threshold values (see Price et al., 2015 for an example). Then, some studies refine – or only perform – the preprocessing by discarding outlier RTs defined as RTs deviating by 2 or 3 times the standard deviation from the mean RTs of each validity condition, for each participant.

A data-driven analysis was used to help decide which combination of methods to apply for data preprocessing. For each task, a mixed linear model was fitted for each possible combination of preprocessing methods with RTs as outcome and validity conditions within each participant as random intercepts. The combination of preprocessing methods that best

explained the RTs variance was selected using the Intraclass Correlation Coefficient (ICC; ICC computed from each model was reported in Table S2 of supplemental material).

For the classic dot-probe task, the best preprocessing method combined the removal of extreme RTs defined as less than 250 ms and greater than 900 ms, and then the removal of outlier RTs defined as deviating from the mean by 2 times the standard deviation,  $ICC = 0.78$ . For the eye-dot-probe task, the best preprocessing method combined the removal of extreme RTs defined as less than 200 ms and greater than 750 ms, and then the removal of outlier RTs defined as deviating from the mean by 2 times the standard deviation,  $ICC = 0.72$ .

Next, the ABI of Macleod et al. (1986), the ABV of Iacoviello et al. (2014), the ABV of Naim et al. (2015), and the five indices of TL-BS of Zvielli et al. (2015): Mean TL-BS<sub>negative</sub>, Mean TL-BS<sub>positive</sub>, Peak TL-BS<sub>negative</sub> and Mean TL-BS<sub>positive</sub>, and TL-BS<sub>variability</sub> were calculated for each participant in each task. All calculations were performed following the respective authors' recommendations (see supplemental material for details about calculations).

To assess the relationship between each attentional index calculated from each dot-probe task, each index was entered as an outcome in nine linear models as a function of the inclusion as a predictor of one of nine composite anxiety scores. The aim of this analysis was to identify the likeliest model for each attentional index calculated from each dot-probe task version using the AIC (i.e., the lower the AIC, the more likely the model; Burnham & Anderson, 2004). These composite anxiety scores were: the STAI-T score or the STAI-S score (meaning that the attentional index was best explained by considering only trait anxiety or state anxiety, respectively); the mean of the STAI-T and the STAI-S (meaning that the attentional bias was best explained by considering both trait anxiety and state anxiety); or one of the three previous anxiety scores squared or cubed (meaning that attentional index values increased sharply as the anxiety score increased (and even more so in the case where the anxiety score was cubed). Testing the models in which the anxiety scores were squared or cubed was motivated by the



modeling that only a small difference in the attentional index was expected between low and medium anxiety scores.

Then, the consistency of the best-fit model of each attentional index calculated from each dot-probe task version was tested by forward selection from the null model. This step consisted of comparing the goodness of fit of the best-fit model on the attentional index to the goodness of fit of the null model on the attentional index, in which no predictor was included. To facilitate the interpretation, each score was normalized.

Linear model assumptions were assessed using the Shapiro test for appropriate error normality, the Goldfeld-Quandt test for homoscedasticity, and the Durbin-Watson test for autocorrelation.

### **Transparency and openness**

The level 2 requirement of the Transparency and Openness Guidelines was met in this article. The sample size determination, data exclusion, manipulation, and measures in the study were reported. Raw data, preprocessing code, analysis code, and material are available through the Open Science Framework repository: <https://osf.io/gdvxp/>. This study was not preregistered. Data were analyzed using R (version 1.4).

### **Results and discussion**

The correlation matrix between anxiety score and attentional index is reported in Table S3 for the classic dot-probe task data and in Table S4 for the eye-dot-probe task data. The consistency of the best-fit model for each index of each task version is reported in Table S5.

### **ABI of Macleod et al. (1986)**

Regarding the ABI of Macleod et al. (1986) computed from the data of the classic dot-probe task, the anxiety score included in the best-fit model was the average of the STAI-T and the STAI-S cubed. The relationship between the ABI and the anxiety score was depicted in Figure 1. This model was consistent because the ABI goodness of fit was maximized by the inclusion

of this anxiety score compared to the null model,  $R^2 = .12$ ,  $F(1,40) = 5.70$ ,  $p > .05$ . This anxiety score was positively linked to the ABI,  $\beta = .35$ , 95% CI [.05, .65], and significantly,  $t(41) = 2.38$ ,  $p > .05$ . Thus, the results of the ABI of Macleod et al. (1986) calculated from the classic dot-probe task revealed that the higher the trait anxiety and the state anxiety, the stronger the attentional bias to threat.

Regarding the ABI of Macleod et al. (1986) computed from the data of the eye-dot-probe task, the anxiety score included in the best-fit model was the STAI-T cubed. The relationship between the ABI and the anxiety score was depicted in Figure 2. This model was consistent because the ABI goodness of fit was maximized by the inclusion of this anxiety score compared to the null model,  $R^2 = .14$ ,  $F(1,40) = 6.61$ ,  $p > .05$ . This anxiety score impacted the ABI positively,  $\beta = .38$ , 95% CI [.11, .52], and significantly,  $t(41) = 2.57$ ,  $p > .05$ . In sum, the results of the ABI of Macleod et al. (1986) calculated from the eye-dot-probe task showed that the higher the trait anxiety, the stronger the attentional bias to threat.

The ABI of Macleod et al. (1986) computed from the data of both the classic dot-probe task and the eye-dot-probe task seems to be a good candidate to measure the attentional bias to threat related to anxiety. In both tasks, the greater the anxiety, the greater the attentional bias to threat measured by the ABI. However, the ABI of Macleod et al. (1986) computed from the eye-dot-probe task data seemed to be more consistent compared to the one computed from the classic dot-probe task data. On the one hand, the ABI calculated from the eye-dot-probe task was best explained by the anxiety score than the ABI calculated from the dot-probe task (i.e.,  $\Delta R^2 = .02$ ). On the other hand, Figure 1 and Figure 2 show that the ABI calculated from the eye-dot-probe task is better distributed on the regression line compared to the ABI calculated from the classic dot-probe task. This reflects that the ABI of Macleod et al. (1986) calculated from the eye-dot-probe task data allowed us to better detect the effect of attentional bias on threat compared to the one calculated from the classic dot-probe task data. This superiority could be explained by

the more robust measurement of eye movement responses than keyboard responses. First, the eye-movement record using an eye-tracker is certainly more precise (set at a sampling rate of 1000Hz in the present study) than the keyboard response. Second, the eye-movement response is certainly more precise to capture the attentional bias than the keyboard response. This assumption is motivated by the fact that eye movements are more closely related to the visual attentional bias process than keyboard responses. Indeed, a visual stimulus that captures attention would have a direct effect on the direction of gaze, while it would have no direct impact on finger movement (Pool et al., 2019).

The finding that the ABI of Macleod et al. (1986) appears to be a good index for measuring the attentional bias to threat in the classic dot-probe task is contrary to all past studies (Carlson & Fang, 2020; Kruijt et al., 2019; Molloy & Anderson, 2020; Price et al., 2015). However, unlike these studies, our analyses considered the combined effects of trait anxiety and state anxiety. Our view was based on the assumption that if the attentional bias to threat is increased by an overestimation of the negative impact or severity of events (Milne et al., 2019), this overestimation should be driven by both trait anxiety and state anxiety. In other words, this means that across participants, the increase in attentional bias to threat may be due to either high trait anxiety for some, high state anxiety for others, or both trait and state anxiety. This result is in line with the hypothesis from the literature that the general affective preferences of stimuli (related to trait anxiety) and their conducive nature to current concerns (related to state anxiety) have combined effects on the attentional bias (Fournier & Koenig, 2023).

The consistency of this hypothesis could be criticized knowing that our results showed that the ABI of the eye-dot-probe task was best explained by the trait anxiety alone. Nonetheless, it could be argued that the appraisal of the conducive nature to current concerns of stimuli would have less impact on attentional bias at an early stage of temporality because its appraisal occurs within a long temporal window (see Fournier & Koenig, 2023 for a discussion). Thus, it is

plausible that using eye movement in this type of task allows for the measurement of an earlier attentional bias than that measured using a keypress (Bannerman et al., 2010), in part because the response times are shorter. Although stimuli presentation time was identical between the two versions of the task, it can be hypothesized that the attentional bias measured in the eye-dot-probe task focused on an earlier time window that is more influenced by general affective preferences for stimuli (i.e., trait anxiety).

Finally, in both tasks, the anxiety scores that best explained the ABI were those raised to the cube. This suggests that attentional biases toward threat in both tasks were enhanced for high anxiety scores compared to low and medium scores.

### **Other attentional indices**

Regarding the ABV of Iacoviello et al. (2014) and of Naim et al. (2015), and the TL-BS five indices of Zvielli et al. (2015) calculated from the classic dot-probe task and the eye-dot-probe task, no best-fit model was evidenced as consistent. Indeed, for each of them, no inclusion of anxiety score maximized the attentional index goodness of fit compared to the null model.

These results contrast with those of Molloy and Anderson (2020) and Price et al. (2015) which suggest that ABV of Iacoviello et al. (2014) and of Naim et al. (2015), and TL-BS indices of Zvielli et al. (2015) are the best reliable indices to assess the anxiety-related attentional bias to threat. These indices measuring attentional bias toward threat and away from threat are derived from the assumption that anxiety is characterized by these two types of attentional biases. However, attentional biases toward threat and away from threat related to anxiety would operate in different temporalities. In fact, studies mentioning attentional biases both toward and away from threat (Cooper & Lang, 2006; Mogg & Bradley, 2006) evidenced attentional bias toward threat during short stimulus presentation (i.e., 100ms and 200ms) and attentional bias away from threat during long stimulus presentation (i.e., 500 and 2000ms). Moreover, the other studies that have shown an attentional bias away from threat also used a long stimulus

presentation time (from 500ms to 2500ms; Constans et al., 2004; Litz et al., 1996; Wald et al., 2011). Thus, anxiety would be related to an attentional bias toward threat at short temporality and an attentional bias away from threat at later temporality. Since the stimulus presentation time was 250ms in the current study, there is no theoretical reason to expect an attentional bias away from the threat. Given this, there is no theoretical reason to expect anxiety-related attentional bias variability in the current study. This would explain the non-significant relationships between the ABV of Iacoviello et al. (2014), or of Naim et al. (2015), or of the TL-BS<sub>variability</sub> of Zvielli et al. (2015) and anxiety.

We would like to emphasize that ABV indexes (Iacoviello et al., 2014; Naim et al., 2015) and TL-BS (Zvielli et al., 2015), should be used with caution. Indeed, these indexes do not inform whether the variability in attentional bias is carried by a fluctuation between no bias and bias toward threat, no bias and bias away from threat, or between bias toward threat and bias away from threat. These calculations provide an excellent measure of the stability of the attentional bias within a dot-probe task session. However, ABV indexes do not allow us to conclude whether attentional bias variability has theoretical (e.g., both attentional bias toward threat and away from threat for anxious individuals) or methodological (e.g., less reliable RTs for anxious individuals) implications. In addition, we struggled with the calculation of the TL-BS. In contrast with the ABV of Iacoviello et al. (2014) and Naim et al. (2015), which compute indexes by bins or blocks of 20 invalid and valid trials, the TL-BS of Zvielli et al. (2015) is computed by matching RTs of each invalid trial with a single valid trial. Since the distribution of RTs within the validity conditions of participants is quite sparse in the dot-probe task, such an attentional bias score calculation is problematic from our point of view. Some TL-BS can be calculated from the difference between the RTs at opposite ends of the distribution tails, resulting in a huge attentional bias score that is actually carried by quasi-aberrant RTs. In sum, we recommend using the ABV of Iacoviello et al. (2014) and Naim et al. (2015) as a measure

of attentional bias stability within participants and not to assess the variability between attentional bias to threat and away from threat.

### **Threshold selection recommendation for future studies**

Molloy & Anderson (2020) emphasized the importance of selecting a priori the preprocessing thresholds for dot-probe task studies because, depending on the selection, the results can change drastically, which can lead to a form of p-hacking. However, an a priori selection of thresholds is somewhat problematic because it depends on the parameters and the quality of the response time distribution (Ratcliff, 1993). Indeed, the distribution of response times can vary on kurtosis from one study to another. Thus, we suggest applying a threshold selection fitting with the distribution of the collected data for each study. To properly perform a posteriori selection of extreme and outlier thresholds, we propose to apply a data-driven selection consistent with the dot-probe task assumption and not a theoretical assumption. Here, threshold selection was based on cutoffs that increased the explained variance of response times as a function of the validity conditions within the participants. The explained variance was measured using ICC, following the recommendation of Price et al. (2015), and Molloy and Anderson (2020). We encourage future studies to apply this threshold selection to best capture the attentional bias measured in the dot-probe task.

### **Limitation and conclusion**

The fact that participants performed only one session of the classic dot-probe task and eye-dot-probe task is a major limitation of the study. Indeed, the test-retest reliability of attentional indices is a critical point of the dot-probe literature because of the current crisis about the attentional bias toward anxiety-related threat. (Kruijt et al., 2019). However, the present findings provide encouraging suggestions for the future of the topic. First, we propose to use the eye-dot-probe task rather than the classic dot-probe task. According to our results, responding by eye movements rather than pressing a key better reveals the attentional bias to

anxiety-related threat. Second, we propose to select extreme and outlier thresholds of RTs of the dot-probe task a posteriori, based on a data-driven manner, without relying on theory. Third, we recommend using the ABI of Macleod et al. (1986) rather than the ABV of Iacoviello et al. (2014) or Naim et al. (2015), or the TL-BS of Zvielli et al. (2015). And fourth, we suggest considering the combined effects of trait anxiety and state anxiety to assess the attentional bias toward threat in the classic dot-probe task.

### References

- Bannerman, R. L., Milders, M., & Sahraie, A. (2010). Attentional bias to brief threat-related faces revealed by saccadic eye movements. *Emotion, 10*(5), 733–738. <https://doi.org/10.1037/a0019354>
- Bar-Haim, Y., Lamy, D., Pergamin, L., Bakermans-Kranenburg, M. J., & van IJzendoorn, M. H. (2007). Threat-related attentional bias in anxious and nonanxious individuals: A meta-analytic study. *Psychological Bulletin, 133*(1), 1–24. <https://doi.org/10.1037/0033-2909.133.1.1>
- Bar-Haim, Y., Holoshitz, Y., Eldar, S., Frenkel, T.I., Muller, D., Charney, D. S., Pine, D. S., Fox, N. A., & Wald, I. (2010). Life-threatening danger and suppression of attention bias to threat. *American Journal of Psychiatry, 167*(6), 694-698. <https://doi.org/0.1176/appi.ajp.2009.09070956>
- Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods & Research, 33*(2), 261–304. <https://doi.org/10.1177/0049124104268644>
- Carlson, J. M., & Fang, L. (2020). The stability and reliability of attentional bias measures in the dot-probe task: Evidence from both traditional mean bias scores and trial-level bias scores. *Motivation and Emotion, 44*(5), 657-669. <https://doi.org/10.1007/s11031-020-09834-6>

- Constans, J. I., McCloskey, M. S., Vasterling, J. J., Brailey, K., & Mathews, A. (2004). Suppression of Attentional Bias in PTSD. *Journal of Abnormal Psychology, 113*(2), 315–323. <https://doi.org/10.1037/0021-843X.113.2.315>
- Cooper, R. M., & Langton, S. R. (2006). Attentional bias to angry faces using the dot-probe task? It depends when you look for it. *Behaviour research and therapy, 44*(9), 1321-1329. <https://doi.org/10.1016/j.brat.2005.10.004>
- Fournier, H., & Koenig, O. (2023). Combined effects of intrinsic and goal relevances on attention and action tendency during the emotional episode. *Emotion, 23*(2), 425-436. <https://doi.org/10.1037/emo0001081>
- Fox, L. S., & Knight, B. G. (2005). The effects of anxiety on attentional processes in older adults. *Aging & mental health, 9*(6), 585-593. <https://doi.org/10.1080/13607860500294282>
- Gauthier, J., & Bouchard, S. (1993). A French-Canadian adaptation of the revised version of Spielberger's State–Trait Anxiety Inventory. *Canadian Journal of Behavioural Science, 25*(4), 559–578. <https://doi.org/10.1037/h0078881>
- Iacoviello, B. M., Wu, G., Abend, R., Murrough, J. W., Feder, A., Fruchter, E., Levinstein, Y., Wald, I., Bailey, C. R., Pine, D. S., Neumeister, A., Bar-Haim, Y., & Charney, D. S. (2014). Attention bias variability and symptoms of posttraumatic stress disorder. *Journal of traumatic stress, 27*(2), 232-239. <https://doi.org/10.1002/jts.21899>
- Jenkins, D. G., & Quintana-Ascencio, P. F. (2020). A solution to minimum sample size for regressions. *PloS one, 15*(2), e0229345. <https://doi.org/10.1371/journal.pone.0229345>
- Kruijt, A.-W., Parsons, S., & Fox, E. (2019). A meta-analysis of bias at baseline in RCTs of attention bias modification: No evidence for dot-probe bias towards threat in clinical anxiety and PTSD. *Journal of Abnormal Psychology, 128*(6), 563–573. <https://doi.org/10.1037/abn0000406>

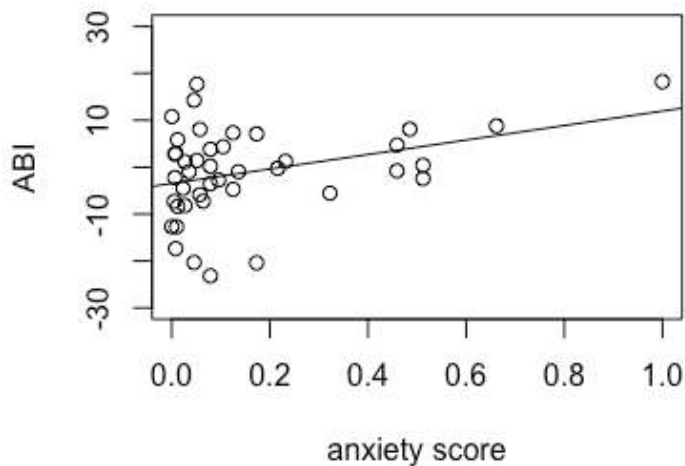


- Litz, B. T., Weathers, F. W., Monaco, V., Herman, D. S., Wulfsohn, M., Marx, B., & Keane, T. M. (1996). Attention, arousal, and memory in posttraumatic stress disorder. *Journal of Traumatic Stress, 9*(3), 497-519. <https://doi.org/10.1007/BF02103661>
- Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior research methods, 47*(4), 1122-1135. <https://doi.org/10.3758/s13428-014-0532-5>
- MacLeod, C., Mathews, A., & Tata, P. (1986). Attentional bias in emotional disorders. *Journal of Abnormal Psychology, 95*(1), 15–20. <http://dx.doi.org/10.1037/0021-843X.95.1.15>
- Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior research methods, 44*(2), 314-324. <https://doi.org/10.3758/s13428-011-0168-7>
- Milne, S., Lomax, C., & Freeston, M. H. (2019). A review of the relationship between intolerance of uncertainty and threat appraisal in anxiety. *The Cognitive Behaviour Therapist, 12*(E38). <https://doi.org/10.1017/S1754470X19000230>
- Molloy, A., & Anderson, P. L. (2020). Evaluating the reliability of attention bias and attention bias variability measures in the dot-probe task among people with social anxiety disorder. *Psychological Assessment, 32*(9), 883-888. <https://doi.org/10.1037/pas0000912>
- Naim, R., Abend, R., Wald, I., Eldar, S., Levi, O., Fruchter, E., Ginat, K., Halpern, P., Sipos, M. L., Adler, A. B., Bliese, P. D., Quartana, P. J., Pine, D. S., & Bar-Haim, Y. (2015). Threat-related attention bias variability and posttraumatic stress. *American Journal of Psychiatry, 172*(12), 1242-1250. <https://doi.org/10.1176/appi.ajp.2015.14121579>
- Öhman, A., Flykt, A., & Esteves, F. (2001). Emotion drives attention: Detecting the snake in the grass. *Journal of Experimental Psychology: General, 130*(3), 466–478. <https://doi.org/10.1037/0096-3445.130.3.466>

- Pool, E. R., Pauli, W. M., Kress, C. S. & O'Doherty, J. P., (2019). Behavioural evidence for parallel outcome-sensitive and outcome-insensitive Pavlovian learning systems in humans. *Nature Human Behavior*, 3(3), 284-296. <https://doi.org/10.1038/s41562-018-0527-9>
- Price, R. B., Kuckertz, J. M., Siegle, G. J., Ladouceur, C. D., Silk, J. S., Ryan, N. D., Dahl, R. E., & Amir, N. (2015). Empirical recommendations for improving the stability of the dot-probe task in clinical research. *Psychological Assessment*, 27(2), 365-376. <https://doi.org/10.1037/pas0000036>
- Quigley, L., Nelson, A. L., Carriere, J., Smilek, D., & Purdon, C. (2012). The effects of trait and state anxiety on attention to emotional images: An eye-tracking study. *Cognition & emotion*, 26(8), 1390-1411. <https://doi.org/10.1080/02699931.2012.662892>
- Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, 114(3), 510–532. <https://doi.org/10.1037/0033-2909.114.3.510>
- Schmukle, S. C. (2005). Unreliability of the dot probe task. *European Journal of Personality*, 19(7), 595-605. <https://doi.org/10.1002/per.554>
- Wald, I., Shechner, T., Bitton, S., Holoshitz, Y., Charney, D. S., Muller, D., Fox, N. A., Pine, D. S., & Bar-Haim, Y. (2011). Attention bias away from threat during life threatening danger predicts PTSD symptoms at one-year follow-up. *Depression and anxiety*, 28(5), 406-411. <https://doi.org/10.1002/da.20808>
- Waters, A. M., Henry, J., Mogg, K., Bradley, B. P., & Pine, D. S. (2010). Attentional bias towards angry faces in childhood anxiety disorders. *Journal of behavior therapy and experimental psychiatry*, 41(2), 158-164. <https://doi.org/10.1016/j.jbtep.2009.12.001>
- Zvielli, A., Bernstein, A., & Koster, E. (2015). Temporal dynamics of attentional bias. *Clinical psychological science*, 3(5), 772–788. <https://doi.org/10.1177/2167702614551572>

**Figure 1**

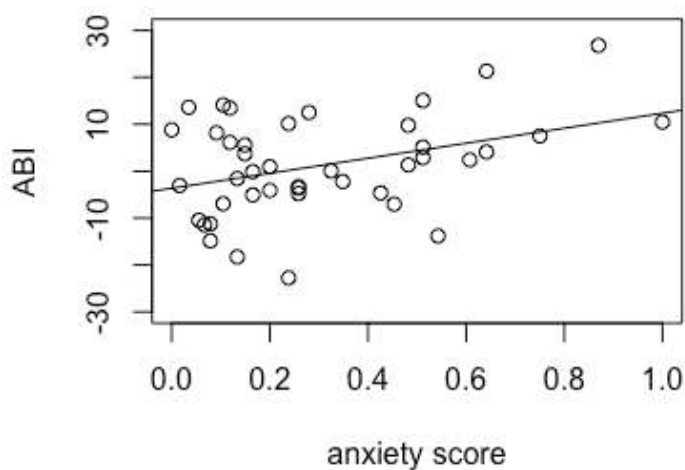
*Scatterplot depicting the relationship between the ABI of Macleod et al., (1986) calculated from the classic dot-probe task data and the anxiety scores*



*Note.* The anxiety score refers to the mean of the STAI-T and the STAI-S cubed. The regression line was reported.

**Figure 2**

*Scatterplot depicting the relationship between the ABI of Macleod et al., (1986) calculated from the eye-dot-probe task data and the anxiety scores*



*Note.* The anxiety score refers to the mean of the STAI-T cubed. The regression line was reported.

## **Chapitre 7 - Article : “Integrated Theory of Appraisal under Causal Hierarchical Inference: an attempt to map appraisal theories to active inference theory”**

Ce chapitre décrit le cinquième article que nous avons écrit qui est actuellement en soumission dans une revue, dont la référence est la suivante :

Fournier, H. & Koenig, O. (Submitted b). Integrated Theory of Appraisal under Causal

Hierarchical Inference: an attempt to map appraisal theories to active inference theory.

## Résumé

La notion d'émotion est utilisée dans de nombreux domaines de recherche et étudiée au moyen de différentes approches depuis des décennies. À l'heure actuelle, il est crucial de confronter ces différentes approches dans le but de tirer profit de possibles connexions. Le présent essai s'inscrit dans cette ligne de pensée, en tentant de rapprocher les hypothèses des théories de l'appraisal de celles de la théorie de l'inférence active. D'une part, les théories de l'appraisal s'inscrivent dans une tradition de psychologie cognitive qui souligne que les interactions des individus avec leur environnement sont construites sur la base d'un processus d'évaluation cognitive. D'autre part, la théorie de l'inférence active suppose que le cerveau fonctionnerait comme une machine d'inférence visant à prédire les différents niveaux de causes hiérarchiques des sensations des individus, façonnant leurs interactions avec l'environnement. L'objectif de cet essai est de tenter d'adapter les processus d'évaluation cognitive en termes de différents niveaux de causes hiérarchiques de la machine d'inférence du cerveau. Nous discutons en particulier des implications d'une telle théorie intégrée de l'appraisal dans le cadre de l'inférence causale hiérarchique. Cette correspondance semble prometteuse et bénéfique à la fois pour les théories de l'appraisal et pour la théorie de l'inférence active en proposant de repenser ou d'affiner certains de leurs postulats.

**Integrated Theory of Appraisal under Causal Hierarchical Inference: an attempt to  
map appraisal theories to active inference theory**

Hippolyte Fournier and Olivier Koenig

Laboratoire EMC, Université Lumière Lyon 2, Bron, France

EMC, Lyon 2 University, Bron, France

**Author note**

Hippolyte Fournier  <https://orcid.org/0000-0002-7861-2387>

Olivier Koenig  <https://orcid.org/0000-0002-5124-1889>

The authors declare that there is no conflict of interest to disclose. Thanks to Eléonore Trân for our discussions and her advice during the drafting process. This article was realized thanks to a doctoral grant to Hippolyte Fournier from the Banque Publique d'Investissement (BPI) under the grant agreement THERADIA.

Correspondence should be addressed to Hippolyte Fournier, Laboratoire EMC, Université Lumière Lyon 2, 5 Avenue Pierre Mendès France, Bron 69676, France. Email:

[h.fournier@univ-lyon2.fr](mailto:h.fournier@univ-lyon2.fr)

### **Abstract**

The notion of emotion has been used in many fields of research for decades. It is now time to confront these different approaches with the objective of drawing benefits from possible connections. The present essay is part of this line of thought, attempting to bring together the assumptions from appraisal theories and from the active inference theory. On the one hand, appraisal theories lie in a cognitive psychology tradition that emphasizes that the interactions of individuals with their environment are built based on a cognitive appraisal process. On the other hand, the active inference theory assumes that the brain would function as an inference machine aimed at predicting the different levels of hierarchical causes of the sensations of individuals, shaping their interactions with the environment. The purpose of this essay is to attempt to fit the cognitive appraisal processes in terms of different levels of hierarchical causes of the brain inference machine. We discuss in particular the implications of such an integrated theory of appraisal under causal hierarchical inference. This matching seems promising and beneficial for both appraisal theories and the active inference theory by proposing to rethink or refine some of their postulates.

*Keywords: emotion, appraisal theories, active inference, predictive coding, free energy,*

## **Integrated Theory of Appraisal under Causal Hierarchical Inference: an attempt to map appraisal theories to active inference theory**

Since the last century, research had developed a massive interest in the study of emotions. Thanks to this, emotion literature has known a huge step forward based on many areas of research, involving different prisms of study rooted in different theoretical traditions. The risk of these considerable parallel advances is to no longer be able to converge on a consensual point. One of the major challenges of the current research is to build a multidisciplinary approach to emotion based on the building blocks of each area, to push the boundaries of understanding emotion (Scherer, 2019). The purpose of the present article is to make an effort in this direction, by trying to bring appraisal theories closer to active inference theory. On the one hand, appraisal theories are a long theoretical tradition of cognitive psychology, defining the individual's interactions in their environment as well as emotion based on a cognitive appraisal process. On the other hand, active inference theory is a very influential recent neuroscientific framework, defining the individual's interactions in their environment and emotion based on a paradigm in which the brain is considered as an inference machine (Bottemanne et al., 2022; Friston, 2012a; Hohwy, 2013). In this essay, we first provided an overview of the assumptions of appraisal theories and active inference theory. Then, we tried to integrate them into a single framework. Finally, we discussed the implications of such an integrated appraisal-active inference framework.

### **Overview of appraisal theories**

First of all, appraisal theories should not be simply reduced to theories of emotion. The core of these theories lies in studying the environmental interactions of individuals through an appraisal-driven perspective fitting in the embodied cognition framework (Shapiro, 2019). In this context, the place reserved for affective phenomena – including emotion – refers to the exchanges between the appraisal process and the organism's subsystems involved in the



individual's environmental interactions (Scherer, 2005b). Thus, the research tradition of the appraisal framework focuses primarily on understanding the nature and the organization of the appraisal process as well as its relationship to the organism's subsystems.

### **The nature and organization of the appraisal process**

In the terms of “appraisal theories”, the appraisal process should be understood as a series of evaluative operations aimed at giving meaning to the environment for the individuals. To this end, the appraisal process would be based on the past sensory experience that individuals have forged from their bodies (Prinz, 2003). In other words, the perception of the environment and the resulting actions would be partly mediated by the appraisal process according to the body in which it is embodied. For instance, a hill would be perceived as steeper when individuals are carrying a heavy backpack, as the action of climbing it would be appraised as more costly due to the extra weight on their body (Bhalla & Proffitt, 1999). The appraisal process would occur at different levels of processing and consciousness. More precisely, appraisals would operate in part outside the field of consciousness, from the sensory-motor level to the conceptual level (see Grandjean et al., 2008 and Scherer, 2005a for a discussion).

Appraisal theorists tend to agree that the appraisal process would be organized according to four major criteria outlined in the Component Process Model (CPM; Scherer, 1999; Scherer, 2009). These criteria would be the relevance, the implication, the coping, and the normative significance. The purpose of relevance appraisal would be to establish the hierarchy of importance of the stimuli in relation to the individual's goals/needs, ranging from irrelevant to relevant. The purpose of the implication appraisal would be to measure the consequences of the stimuli on the individual's short- and long-term goals/needs, ranging from obstructive to conducive. The purpose of the coping appraisal would be to measure whether an adjustment or adaptation is possible in the face of stimuli and their consequences, ranging from low to high. Finally, the purpose of the normative significance appraisal would be to measure whether the

stimulus significance is in accordance with the internal or external standards of the individual, ranging from violation to agreement. In sum, environment perception and action selection would be mediated by the stimulus appraisal based on their relevance, their implication, the coping in front of their implication, and their normative significance in terms of an individual's goals and needs.

While appraisal theorists agree that these appraisal criteria range from lower to higher levels of processing (i.e., from relevance to normative significance), their prerequisites of occurrence is a matter of debate (Grandjean & Scherer, 2014). Based on the premise of associative networks, some authors suggest that appraisals would function in terms of memory activation by common features of the stimuli (Forgas & Smith, 2007; Smith & Lazarus, 1990). Thus, the temporal sequence of the criteria of the appraisal process would be flexible and may even occur simultaneously. Other authors assume that the temporal sequence of appraisal is fixed so that the high-level appraisal criteria cannot be processed before low-level criteria, although they may occur in parallel (Grandjean & Scherer, 2008; Sander et al., 2005; Scherer, 1984). The rationale underlying this assumption is that low-level appraisal criteria should occur first, as their results are a precondition of the smooth running of high-level appraisal criteria. Still, others fall somewhere in between these two views, assuming that the temporal sequence of appraisal criteria would be flexible, except for the first sub-criterion within the relevance appraisal criterion, namely the novelty (Ellsworth, 1991). This position emphasizes the existence of a micro-sequence in which appraising the novelty of stimuli is a necessary condition for the triggering of the other appraisal criteria.

### ***Focus on the functioning of the relevance appraisal criterion***

Among all appraisal criteria, particular interest was given to relevance. This interest lies in the fact that relevance would be the main interface criterion between individuals and their environment. More precisely, the hierarchy of the importance of stimuli established by the

relevance appraisal criterion would allow individuals to allocate their attention to their environment in an automatic way (Scherer & Brosch, 2009). This phenomenon, defined as emotional attention (Vuilleumier, 2005), would aim to indicate which stimuli need to be processed first because interaction with them is likely necessary to achieve the well-being and survival of individuals (Scherer & Brosch, 2009). Thus, the study field of emotional attention provides insight into both the deployment of attention as well as the rule governing the relevance of the stimulus importance hierarchy established by the relevance criterion. The relevance appraisal criterion would be divided into three sub-criteria (Scherer, 2009). The first criterion, defined as “novelty”, would aim to appraise whether stimuli are sudden, familiar, or predictable. For instance, stimuli that suddenly appear in the visual field would capture attention because they are relevant based on the novelty appraisal (e.g., Sussman et al., 2003; Yantis & Jonides, 1984). The second criterion, “intrinsic relevance”, would appraise whether the stimuli are generally unpleasant (probably requiring avoidance) or pleasant (probably requiring approach) regardless of the current state of the organism. Thus, stimuli appraised as generally unpleasant such as snakes, spiders, or firearms would capture attention because they should probably be avoided (e.g., Brosch & Sharma, 2005; Fox et al., 2007; LoBue & DeLoache, 2008; Öhman et al., 2001; Öhman & Mineka, 2003). The same goes for stimuli appraised as generally pleasant such as baby faces, happy faces, or erotic because they should probably be approached (e.g., Brosch et al., 2007, 2008; Fromberger et al., 2012; Pool et al., 2016). The third criterion, “goal relevance”, would appraise whether the stimuli can obstruct (probably requiring avoidance) or conduct (probably requiring approach) the achievement of the currently important goals/needs. Thus, stimuli appraised as obstructive to the achievement of a goal or need, such as loss-related stimuli in the context of a game, would capture attention because they should probably be avoided (Müller et al., 2016; Sun et al., 2017). In contrast, stimuli appraised as conducive to the achievement of a goal or need, such as a bottle of water in the context of

thirst, stimuli related to cleanliness after disgust induction, or stimuli related to gain in the context of a game, because they should probably be approached (Mazzietti et al., 2014; Vogt et al., 2011, 2013).

The way in which the novelty criterion combines with the intrinsic relevance and the goal relevance criteria to establish the stimulus relevance hierarchy – and generate attentional capture – seems to be self-evident and does not suffer from criticism in the literature. Indeed, it seems easy to assimilate that the relevance of a stimulus fulfilling both the novelty and intrinsic relevance appraisal criteria would have a prominent place in the stimulus importance hierarchy (van Peer et al., 2014). However, the question of the combined effects of intrinsic relevance and goal relevance has been raised: What about the place in the hierarchy of stimuli fulfilling both intrinsic relevance and goal relevance criteria? Recent studies have focused on this question through the attentional impact of the combined effects of intrinsic relevance and goal relevance. First, it has been shown that stimuli (i.e., foods and drinks) appraised as pleasant and goal conducive (i.e., satisfying hunger or thirst) would have multiplicative effects on attentional capture (Fournier & Koenig, 2023). These multiplicative effects mean that foods (and drinks) appraised as both moderately pleasant and satisfying hunger captured more attention than foods (and drinks) appraised only as strongly pleasant or strongly satisfying hunger (or thirst). The authors concluded that attentional capture would be increased toward stimuli because the two relevances suggest that stimuli should likely be approached, especially if the two relevance levels are close. In addition, it has been shown that this increase in attentional capture toward stimuli appraised to be both pleasant and goal conducive would be equivalent if stimuli were appraised to be both unpleasant and goal-conducive (Fournier & Koenig, Under review). According to the authors, since stimuli should likely be avoided and approached by being appraised as both unpleasant and goal-conducive, an increase in attentional capture would emphasize the need for a more thorough appraisal in deciding which action to take. Finally, it

has been evidenced that attentional capture toward stimuli appraised as unpleasant or pleasant could be canceled once they were appraised as irrelevant to the goal (Fournier & Koenig, Submitted a). In this context, authors observed that the initial need to avoid (or approach) the unpleasant (or pleasant) stimuli was canceled by the lack of need for action resulting from the induction of goal irrelevance, which resulted in attention no longer being captured by stimuli. Based on these findings, the relevance hierarchy would be established as a function of whether intrinsic relevance and goal relevance of stimuli are congruent (e.g., both generating approach tendency) or conflicting (i.e., generating approach and avoidance tendencies). In the case of congruence, this would reflect an increase in the need for action (especially when the relevance levels are close), while in the case of conflict, this would reflect a need for deeper appraisal. In addition, one relevance could overwhelm the other, if it is strongly more likely.

Although not in the tradition of appraisal theories, several pieces of the literature suggest that early attention would be driven more by intrinsic relevance, while later attention would be driven more by goal relevance (Castellanos et al., 2009; Field & Cox, 2008; Forrest et al., 2022; Nijs et al., 2010; Werthmann et al., 2011). These results would reflect that appraising intrinsic relevance would occur faster than goal relevance because its appraisal would require less information (Aue et al., 2007; Grandjean & Scherer, 2008; Lanctôt & Hess, 2007; Scherer, 2013). Thus, the relevance hierarchy established at an early stage would give more weight to the intrinsic relevance of the stimuli, while more weight would be given to the goal relevance at a later stage.

### **Affective phenomena through appraisal theories**

Appraisal theories are based on an epistemological current that emphasizes the importance to think about all models of behavior, brain, and mind considering the role of affect (Dukes et al., 2021). In this framework, affective phenomena refer to multi-component processes that intensify or attenuate bodily responses, thereby shaping individuals' environmental interactions

(Massumi, 2002). The appraisals would be the conductor of this multi-component process, aiming to regulate the state of the organism's subsystems in order to facilitate the realization of adapted behaviors of individuals in their environment (Frijda, 2004; Sander, 2013). For instance, stimulus appraisal would generate an appropriate physiological response (e.g., skin conductance; van Reekum et al., 2004), motor expression (e.g., facial expressions; Scherer et al., 2019), or action tendency (e.g., avoidance or approach; Scarantino, 2017). In turn, the internal states of the organism's subsystems could influence the appraisal process by biasing the appraisal of criteria (Grandjean & Scherer, 2014). This claim can be illustrated through the somatic marker hypothesis assuming that the interoceptive states would play a role in decision-making (Bechara et al., 2005; Harrison et al., 2016; Poppa & Bechara, 2018; Sugawara et al., 2020). A well-known case study from this research tradition highlighted that physiological responses triggered by the appraisal of negative consequences of stimuli could influence decision-making related to those stimuli in subsequent interactions (see Bechara et al., 2005). Such a bidirectional loop between the appraisal process and the regulation of an organism's subsystem states during episodes of an individual's environmental interactions would be underlaid by the so-called "reappraisal" (Ellsworth & Dougherty, 2016; Gross, 2015; Uusberg et al., 2023). For a given episode starting with the perception of an event, the appraisal process would be used to measure the event implications as well as the appropriate behaviors to the response of this event. Depending on the result of the appraisal process, the organism's subsystem states would be modified to facilitate the realization of these appropriate behaviors. During this process, a reappraisal would occur to refine the measurement of the event implications and the selection of appropriate behaviors that follow. The bidirectional loop between the reappraisal process and the regulation of the organism's subsystem states would be carried out until the environmental interaction is appraised as completed.

A singularity of appraisal theories that undoubtedly places them within the affective framework is that they assume that the non-affective episode does not differ from the affective episode in nature (Scherer, 2005a, 2005b). That is, all episodes of environmental interactions would likely involve synchronization between appraisals and organism's subsystems activations, whether considered as affective or not. For example, an individual with a small amount of hunger is likely to appraise a food as relevant, which could trigger both a mild salivation and a slight approach tendency, although this episode would not be considered purely affective. Thus, the distinction between non-affective episodes and affective episodes, such as emotion or mood, depends on a set of conditions being met or not. On the one hand, an episode should be considered as an emotion when: 1) it is triggered by a stimulus; 2) the stimulus has been appraised as relevant; 3) the synchronization of the organism's subsystems is important; 4) the behavioral impact is very high (e.g., interrupting ongoing behaviors); and 5) the episode is very intense and brief (Sander, 2013; Scherer, 2005b). On the other hand, an episode should be considered as a mood when: 1) it is not clearly triggered by a stimulus; 2) the synchronization of the organism's subsystems is low; 3) the behavioral impact is high (e.g., by conflicting with other ongoing behaviors); 4) and the episode is moderately intense and is durable (Scherer, 2005b; Siemer, 2009). In the case these conditions are not filled, the episode would not be considered as affective. Appraisal theorists emphasize the importance of not confusing the affective state with affect (Folkman & Lazarus, 1985; Frijda, 1986; Moors et al., 2013; Scherer, 1993). While the affective state would reflect the subjective feeling shaped by the organism's subsystems states during the episode, the affect refers to the whole process involved in the episode. The affective state is thought to be primarily non-conscious and emerges to consciousness if the activation of the organism's subsystems is high and reaches critical thresholds of synchronization (Dan Gläuser & Scherer, 2008; Grandjean et al., 2008; Scherer, 2009).

According to appraisal theories, the appraisal process would be causal in the elicitation and differentiation of emotions (Lazarus, 1991; Moors, 2013; Moors et al., 2013; Smith & Kirby, 2009). If patterns of appraisal criteria are intended to produce the most appropriate response, similar patterns would probably generate similar schemes of organism's subsystems activations, thereby similar emotional experiences. So, emotions would be associated with specific patterns of appraisal criteria. For example, a pattern of appraisal criteria consisting of goal obstruction and low coping potential would likely generate fear, whereas it would likely generate anger if the coping potential is high (Scherer et al., 2013, 2019; Scherer & Ellgring, 2007). In sum, there would be a necessary and sufficient relationship between the patterns of appraisal criteria and the emotional response (Siemer et al., 2007): On the one hand, a given pattern of appraisal criteria would generate a given scheme of synchronization of activations of the organism's subsystems. On the other hand, a given scheme of synchronization of activations of the organism's subsystems would be the consequence of a given pattern of appraisal criteria.

In sum, appraisal theories provide a powerful framework of cognitive psychology for understanding the interactions between individuals and their environment and for delineating the boundaries of affective phenomena as well as their nature.

### **Overview of active inference theory**

The core of active inference theory lies in defining the brain functions in terms of predictions and prediction error exchanges (Bottemanne et al., 2022; Friston, 2012a; Hohwy, 2013). In this context, individuals are considered as self-organized systems whose environmental interactions are described through a hierarchical prediction process that meets the purpose of maintaining their homeostasis – or minimizing their free energy (Kirchhoff et al., 2018; Seth, 2013). The recent craze for this research field focuses primarily to rethink the existing models of behavior, brain, and mind into this perspective. This leads to a universal framework for the study of human beings, unifying the fields of psychology, neuroscience, philosophy, and mathematics.



## **The architecture and the organization of the brain system**

The active inference theory is based on the predictive coding hypothesis applied to a predictive brain paradigm (Bottemanne et al., 2022; Friston, 2012a; Hohwy, 2013). According to this hypothesis, the brain would function as an inference machine building a generative model of the world. This model would aim at predicting the current and future state of the sensory inputs as well as the different levels of hierarchical causes that can explain their state. As an example, based on the state of the sensory inputs from a perception of a visual scene, the hierarchical causes would refer to the local image features (e.g., luminance, color), to the shape of the objects (e.g., a circle), to the class of the object (e.g., a balloon; Kersten et al., 2004). To this end, the brain would implement two models. On the one hand, the forward model – constituting the top-down connections – would code the predictions about the different levels of hierarchical causes of the sensory input states. These predictions would be performed by Bayesian inference (Colombo & Seriès, 2012; Durante & Dunson, 2018; Lin & Garrido, 2022) encoding the probability distributions of the multiple hypotheses representing the plausible beliefs about the hierarchical causes (Seth & Friston, 2016). On the other hand, the backward model – constituting the bottom-up connections – would code the prediction errors about the causes of the sensory input states. The prediction error should be understood as the part of an inferred cause that does not correctly explain the states of the sensory inputs (e.g., predicting the class of an object such a balloon knowing that its inferred shape is a square). Such a conceptualization of the brain system involves reversing the traditional view between bottom-up and top-down messages (Friston, 2018). Indeed, rather than defining top-down messages as executive control and bottom-up messages as environment perception (Eysenck et al., 2007; Mogg & Bradley, 2016, 2018; Yantis, 1998), top-down messages would reflect the prediction, while bottom-up messages would reflect the prediction error (Friston, 2008; Grossberg, 2009; Hesselmann et al., 2010; Rauss & Pourtois, 2013).

One assumption of the brain implementation according to the predictive coding hypothesis involves two postulates (Friston, 2005, 2018). First, the different levels of the hierarchical causes would be assigned to specific brain layers (e.g., the object shape, and the object class). The plausible hypotheses regarding the cause assigned to each layer would be inferred within it. Second, the higher hierarchical levels of causes would aim to predict the causes of the lower levels (through the forward model). In turn, the lower hierarchical levels of causes would aim to correct the prediction about them made based on the causes of the higher levels (through the backward model). Thus, the belief about a given level of causes – or the state of its assigned layer – would result from the selection of the most likely hypothesis among those inferred as plausible within the layer.

It was proposed to apply the free energy principle to account for the selection of the most likely hypothesis regarding the different levels of hierarchical causes (Friston, 2005, 2010, 2012a). First, it implies that the brain would manipulate a statistical quantity of prediction error called free energy (Friston, 2012b; Smith et al., 2019). With respect to the hypothesis selection process about a given cause, two computations of free energy quantity would be performed for each plausible hypothesis. The first computation results from the difference between each hypothesis and the prediction sent by the forward model. This free energy quantity would reflect the hypothetical prediction error quantity that the backward model will send to the higher level of causes if this hypothesis is selected. The second computation of free energy quantity results from the difference between each hypothesis and the prediction error sent by the backward model. This free energy quantity would reflect the hypothetical quantity of prediction error that the backward model will return to the current cause level (from the lower level of cause) if this hypothesis is selected.

A second assumption of the application of the free energy principle to the hypothesis selection process invites us to consider the brain layers as self-organized systems (Kirchhoff et

al., 2018). More precisely, it is assumed that the processes occurring within layers would be governed by an internal rule, namely the maintenance of their homeostasis. In this context, the brain layer homeostasis refers to maintaining functional integrity by resisting disorder. This would be done by minimizing the prediction error or, more precisely, the free energy (Friston, 2010). Thus, among the plausible hypotheses concerning the different levels of hierarchical causes, the selection process would be performed on the basis of those that minimize the quantity of free energy. It is important to note that the computation of free energy quantity (Smith et al., 2021) would be weighted by the degree of confidence granted to the predictions. In other words, if a plausible hypothesis minimizes the free energy relative to a highly confident prediction sent by the forward model, it could be selected whatever the free energy quantity computed in relation to the prediction error sent by the backward model. In summary, the predictive coding hypothesis governed by the free energy principle invites us to consider the brain itself as a self-organized system composed of nested self-organized subsystems (i.e., the brain layers). At the level of the whole brain, maintaining homeostasis – or minimizing free energy – aims to maximize the accuracy of the generative model of the world predicting the state of the sensory inputs and their different levels of hierarchical causes (Friston, 2012b; Friston et al., 2010; Kirchhoff et al., 2018).

### **The brain inference machine and environmental interactions**

Active inference theory extends the implications of the predictive coding hypothesis through free energy minimization to a broader level of abstraction, namely the set of interactions of individuals with their environment (Friston et al., 2017). This is part of an embodied perspective in which the inferences – or cognition – are nested in a body that is itself nested in an environment (Friston et al., 2013). In other words, the brain generative model would aim to maintain the homeostasis of the body's subsystems (i.e., their structural and functional integrity). To this end, the model would also implement a process of interoceptive inference

aiming to predict the current and future interoceptive states as well as their different levels of hierarchical causes (Barrett & Simmons, 2015; Seth, 2013, 2015; Seth & Friston, 2016). Interoceptive state refers to the experience of the activation states of the organism's subsystems (Craig, 2002, 2003). For example, it would predict that a previously robust effort (a level of hierarchical cause) can explain a surprising state of hunger (i.e., an inferred interoceptive state). In this case, the surprising state of hunger would be generated by updating the inferred interoceptive state from an afferent message (prediction error) of a drop in sugar blood level (Seth, 2015). In this context, actions would be integrated into the brain generative model as autonomic reflexes inferred as correcting prediction errors in the different levels of hierarchical causes that lead to undesirable states (Adams et al., 2013; Buckley et al., 2017; Friston, 2011; Pezzulo et al., 2015). The term "action" is used here in a broad sense, including external actions (i.e., behavioral sequences) and internal actions (i.e., regulation of the states of the organism's subsystems). Following the previous example, consuming food (external action) and/or metabolizing fat reserves (internal action) could be inferred as allowing to compensate for the drop in blood sugar level (prediction error) that led to the hunger state (undesired interoceptive state). However, the generative model can also update its predictions, by defining the hunger state as viable for a while by lowering the viability thresholds of blood sugar levels. The decision about the resulting scheme (between updating predictions and/or triggering internal/external action) would be made by selecting the hypotheses within the different levels of hierarchical causes that best minimize the free energy (Friston, 2005, 2010, 2012a). In brief, active inference theory assumes that the perception and meaning of the environment, as well as the resulting actions, are the result of a complex process of inference governed by the principle of free energy minimization. Although the notion of free energy minimization refers to minimizing prediction error at a neurocomputational level, it ultimately leads to achieving the individual's homeostasis (i.e., their well-being and survival).

### **Toward an integrated framework of appraisal and active inference**

The use of the postulates of appraisal theories to account for the phenomenon of emotion in active inference theory has already been discussed in the literature. It was first done in the context of an active interoceptive inference proposal (Clark et al., 2018; Seth, 2013, 2015; Seth & Friston, 2016). In this context, an emotional state emerges from the result of prediction errors about interoceptive inference. More precisely, emotional states would refer to the multicomponent state generated by the update of the brain generative model regarding the different levels of hierarchical causes of a surprising interoceptive state. According to this perspective, the active interoceptive inference hypothesis would extend appraisal theories to active inference theory. However, it seems that appraisal theories were confused with the peripheral theory (see James, 1894/1994) and the two-factor theory (Schachter & Singer, 1962). Indeed, appraisal theories have been amalgamed with a definition of emotion in which they "... depend on cognitive interpretations of physiological changes" (Seth, 2013, p.1). Thus, this statement neglects the causal nature of appraisal in the triggering of emotions and the fact that emotions are not states but episodes (Folkman & Lazarus, 1985; Frijda, 1986; Moors et al., 2013; Scherer, 1993).

Later, a possible connection between active inference theory and appraisal theories (their recent form) has been proposed, while being subject to criticism (Barrett, 2017, p. 7). The author claimed that in the active inference theory, "The meaning of a sensory event includes visceromotor and motor action plans to deal with that event ... meaning does not trigger action but results from it", while "appraisal theories ... assume that a response derives from a stimulus that is evaluated for its meaning". According to the author, these two statements would make appraisal theories highly doubtful because the former theory assumes that the stimulus meaning derives from the action, whereas the latter assumes that the action derives from the stimulus meaning. However, we claim that there can be no debate on the sequence because visceromotor

and motor action plans necessarily include prior beliefs about the stimuli meaning. Regardless of the level of processing that drives action toward stimuli, the achievement of biological needs or conceptual goals requires prior beliefs about the conduciveness nature of stimuli to achieve goals and needs and thus requires meaning. Thus, meaning certainly derives from action, but action also derives from meaning, so questioning the primacy of action or meaning is like questioning the primacy of the egg or the chicken. Barrett (2017, p. 7) further pointed out that: “a brain implements its internal model with ‘concepts’ that ‘categorize’ sensations to give them meaning ... Predictions are concepts ... Completed predictions are categorizations ... Appraisals as descriptions of the world ... are produced by categorization with concepts”. It is precisely this point that allows us to establish a connection between appraisal theories and active inference theory, appraisals are part of the predictions that give meaning to the world.

Such a concept of appraisal, in the sense of a prediction process that gives meaning to the world – and triggers the emotional episode – has been integrated into the active inference theory through the Three-Process Model (TPM; Smith, 2020). According to the TPM, the emotional episode would begin when the brain generative model would build a multi-level situation representation of the event. This involves predicting the different levels of its hierarchical causes from sensory inputs, such as local image features, to object classes. (5/1/23 9:31:00 AM. Then, the event would be appraised in a probabilistic fashion according to its meaning regarding the well-being and survival of individuals. It is important to note that this appraisal process is explicitly stated in the TPM as one proposed by appraisal theories (Scherer, 1999; Scherer, 2009). Finally, the scheme of cognitive disposition (e.g., attentional capture), the activation of the organism’s subsystems, and the behavioral sequence predicted to best minimize free energy in the face to the event implication inferred from the appraisal process would be triggered.

### **Implications and future directions of an integrated framework of appraisal and active inference**

## **The nature and organization of the appraisal process through an integrated theory of appraisal and active inference**

Considering the functioning of the appraisal process within the framework of the active inference theory provides new insights for appraisal theories. In the remainder of the paper, we propose to refer to the “Integrated Theory of Appraisal under Causal Hierarchical Inference” (ITACHI) to mention the framework integrating appraisal theories with active inference theory. First, ITACHI invites a rethink of the appraisal process in the appraising brain (i.e., the brain implementation of the appraisal process; Brosch & Sander, 2013; Sander et al., 2018). More precisely, the appraisal process would be considered as a process predicting (Barrett, 2017; Smith, 2020) the different levels of hierarchical causes of the implication of events concerning the homeostasis of the individual (Friston, 2012b; Friston et al., 2010; Kirchhoff et al., 2018). The implication of the events would consist of their impact on the current and future inferred interoceptive states, taking into account potential actions predicted to correct undesirable states (Friston et al., 2017; Seth, 2013, 2015; Seth & Friston, 2016). To this end, the appraisal would be performed by Bayesian inferences (Colombo & Seriès, 2012; Durante & Dunson, 2018; Lin & Garrido, 2022) encoding the probability distributions of the multiple hypotheses representing the plausible beliefs – or hypotheses – about the different levels of these hierarchical causes.

In terms of the organization of the appraisal process, ITACHI involves defining each appraisal criterion as a hierarchical level of the cause of the implication of the events for the individual’s homeostasis (Friston, 2005). As a reminder, these different levels of hierarchical causes would be, from the lowest to the highest level: The relevance, the implication, the coping, and the normative significance (Scherer, 1999; Scherer, 2009). Each appraisal criterion would be assigned to a specific brain layer (Friston, 2005). The plausible hypotheses regarding the appraisal criteria would be inferred within the brain layer assigned to them. On the one hand, the higher-level appraisal criteria would predict the lower-level appraisal criteria through

the forward model. On the other hand, the lower-level appraisal criteria would correct the predictions sent from the higher-level appraisal criteria through the backward model. Thus, each appraisal criterion hypothesis (or belief) would be selected as the one that minimizes the free energy in its correspondence with the prediction and the prediction error sent by the higher- and lower-level appraisal criteria, respectively (Friston, 2005, 2018). Such a functioning of the appraisal process seems rather suitable. Indeed, the proposal that the relevance of stimuli would be predicted based on their implication seems to make sense. So does the proposal that the implication of stimuli would be predicted based on the possible coping of individuals in the face of their consequences. The same is true for the proposal that the possible coping of individuals with the stimulus consequences would be predicted based on their normative significance. The statement that the inverse path (i.e., from lower- to higher-level appraisal criteria) would aim to correct the prediction errors of higher-level appraisal criteria seems also consistent. It is important to note that we described here a proposition of the simplest formalism about an organization of appraisal criteria in the framework of the active inference theory. We do not deny that the implementation of appraisal criteria in such an architecture is surely more complex (e.g., implication appraisal criteria would receive direct predictions from both the coping and normative significance appraisal criteria). That said, considering the nature and the organization of the appraisal process according to this view allows us to refine the understanding of its functioning.

For instance, ITACHI provides new elements to the debate on the conditions of occurrence of appraisal criteria (Ellsworth, 1991; Forgas & Smith, 2007; Grandjean & Scherer, 2008; Lazarus, 1991; Scherer, 1984). First, the functioning of the appraisal process according to ITACHI might appear to be consistent with the assumption that the appraisal process would be flexible and may even occur in one step (Forgas & Smith, 2007; Smith & Lazarus, 1990). When the multi-level situation representation of the event is achieved, it would be possible for prior



beliefs – or hypotheses – to emerge simultaneously about the appraisal criteria, in a manner similar to the operation of associative networks. This seems to be in contradiction with the assumption that lower-level appraisal criteria should occur first, as their results are a precondition of the smooth running of higher-level appraisal criteria (Grandjean & Scherer, 2008; Sander et al., 2005; Scherer, 1984). However, retrieving prior beliefs about appraisal criteria would also consist of a process of inference. Because higher-level appraisal criteria probably require more information – or at least more complex information – than lower-level criteria (Aue et al., 2007; Grandjean & Scherer, 2008; Lanctôt & Hess, 2007; Scherer, 2013), retrieving prior beliefs about them would likely take longer. Unless there is no uncertainty in the appraisal process – an unlikely situation – it seems more plausible that beliefs about higher-level appraisal criteria emerge after those of lower-level criteria, although all their inferences may begin simultaneously. The same applies to the novelty appraisal sub-criterion (Ellsworth, 1991).

***Focus on the functioning of the relevance appraisal criterion through an integrated framework of appraisal and active inference***

ITACHI also provides a better understanding of how the stimulus relevance hierarchy is established according to the sub-criteria of relevance appraisal, thereby accounting for the deployment of emotional attention (Scherer, 2009; Scherer & Brosch, 2009; Vuilleumier, 2005). On the one hand, it allows for a precise account of how the other appraisal criteria shape the relevance criterion, based on the predictions sent by the forward model from the implication criterion. On the other hand, it is particularly useful for refining the functioning of the combined effects of intrinsic relevance and goal relevance. In the case of congruence (e.g., when pleasant and goal-conducive appraisals emphasize a need for approach), the increase in attentional capture toward stimuli – reflecting a prominent place in the hierarchy – would reflect an increase in prediction confidence. More precisely, the hypothesis that stimuli need to be

approached would minimize free energy, especially when intrinsic relevance and goal relevance appraisals are close (Fournier & Koenig, 2023). In case of conflict (e.g., when unpleasant and goal-conducive appraisals emphasize both a need to avoid and to approach), the increase in attentional capture toward stimuli would reflect a need to reduce the level of uncertainty (Fournier & Koenig, Under review). This corresponds to the direct translation of the free energy principle dictating that organisms seek to minimize free energy (Friston, 2005, 2010). Then, in case of one relevance overwhelm the other (e.g., when goal irrelevance cancels the initial intrinsic relevance of stimuli), the cancellation of the attentional capture by the stimuli would reflect a gap between the predictions of the two relevance appraisals (Fournier & Koenig, submitted a). Specifically, intrinsically relevant stimuli initially appraised as requiring action may ultimately be appraised as requiring no action if they are appraised as goal irrelevant and if the confidence of the latter appraisal is higher.

Finally, the assumption that early attention would be driven more by intrinsic relevance, while later attention would be driven more by goal relevance can also be articulated in terms of prediction uncertainty (Castellanos et al., 2009; Field & Cox, 2008; Forrest et al., 2022; Nijs et al., 2010; Werthmann et al., 2011). Since the prediction of goal relevance would not be completed at early temporality unlike that of intrinsic relevance (Aue et al., 2007; Grandjean & Scherer, 2008; Lanctôt & Hess, 2007; Scherer, 2013), it means that its prediction is likely more uncertain. Thus, the greater impact of intrinsic relevance on attention at early temporality may reflect that greater weight would be given to it because its prediction is likely to be more confident than that of goal relevance at that stage of temporality. As time goes by, the prediction of goal relevance is refined. The goal relevance prediction is likely to be more credible than that of intrinsic relevance once it is sufficiently completed because it is more in tune with the current goals and needs of individuals. Hence, the greater impact of goal relevance on attention

at later temporality would reflect that a greater weight is given to it once its inferential process is sufficiently advanced.

### **Affective phenomena through an integrated appraisal-active inference framework**

The considerations of ITACHI seem consistent with the definition of how the appraisal process would mediate the episodes of an individual's environmental interactions proposed by appraisal theories. Indeed, both appraisal theories (Frijda, 2004; Sander, 2013) and the TPM (Smith, 2020) assume that the appraisal process conducts the regulation of the organism's subsystem states to facilitate the realization of adapted behaviors in the face of the event implications. Furthermore, ITACHI extends some concepts proposed by appraisal theories. First, it qualifies both the regulation of the organism's subsystem states and the adapted behaviors as autonomic reflexes fulfilling predictions inferred as minimizing free energy based on the appraisal process (Adams et al., 2013; Buckley et al., 2017; Friston, 2011; Pezzulo et al., 2015). Second, it invites to consider the process of reappraisal (Ellsworth & Dougherty, 2016; Gross, 2015; Uusberg et al., 2023) refining the measure of the event implications and the selection of adapted actions as a loop of bidirectional exchanges between predictions and prediction errors. More precisely, the reappraisal process would be a search for alignment (i.e., free energy minimization) between the forward and the backward models about the update of the hypotheses about the appraisal criteria as the episode proceeds. For instance, the update about the inferred implication of the event would generate prediction errors regarding the inferred coping of its consequences, which would lead to updating this inferred coping, and then generating prediction errors to other levels of appraisal criteria until a satisfactory alignment is found.

ITACHI can also allow refining the proposal of appraisal theories to distinguish affective episodes from non-affective episodes according to condition filling (Scherer, 2005a, 2005b). For instance, one of the appraisal theories' challenges is distinguishing emotional episodes from

non-emotional episodes based on a sufficiently important synchronization of the organism's subsystems and behavioral impacts. The difficulty arises insofar as these processes are performed using information of a different nature (e.g., heart rate for peripheral responses, action units for facial expression, neural messages for cognitive processes). Although not yet addressed in the present essay, the literature on active inference theory focuses largely on the development of statistical formulas representing these activities through unified information, namely free energy (Buckley et al., 2017; Friston, 2010; Friston et al., 2017; see Smith et al., 2022 for a step-by-step tutorial to build a model from empirical data under active inference). Thus, a mathematical application of the assumptions of appraisal theories could lead in the long run to distinguish emotional episodes from non-emotional ones based on thresholds of synchronization of the organism's subsystems and behavioral impacts through free energy formulas. The same goes for distinguishing emotional episodes from mood.

Applying ITACHI could also be beneficial in addressing the challenge of emotion differentiation. In this regard, active inference theory assumes that emotions could be differentiated according to how the meaning of events inferred from their different level of hierarchical causes would lead to varying the free energy over time with respect to the inferred interoceptive states (Clark et al., 2018; Joffily & Coricelli, 2013; Smith, 2020). First, this involves that negative emotions would be associated with a prediction of free energy maximization over time (i.e., a prediction of an increase in the degree of uncertainty regarding the maintenance of homeostasis about interoceptive states over time). In contrast, positive emotions would be associated with a prediction of free energy minimization (i.e., a prediction of an increase in the degree of confidence regarding the maintenance of homeostasis about interoceptive states over time). Second, emotions of the same valence (i.e., negative or positive) could be differentiated by considering the temporal dynamics of the free energy between the starting and arrival ending quantities within the prediction.

Some authors assume that emotions can be mapped as curves (rather than points) in a Cartesian plane, along the two dimensions of inferred free energy and time (Joffily & Coricelli, 2013). Regarding a positive emotion such as happiness, its curve would be described as an already low quantity of free energy that continues to decrease little by little over time. In other words, this represents a confident prediction about the maintenance of homeostasis of future interoceptive states, the confidence of which is inferred to increase little by little over time. The curve of another positive emotion such as hope would be described as a large quantity of free energy that decreases more and more strongly with time. That is, it refers to an uncertain prediction about the maintenance of homeostasis of future interoceptive states, whose uncertainty is inferred to decrease more and more strongly over time. Regarding a negative emotion such as sadness, its curve would be described as an already high quantity of free energy that continues to increase little by little over time. So, unhappiness would reflect an already uncertain prediction about the maintenance of homeostasis of future interoceptive states, whose uncertainty is inferred to increase little by little over time. In contrast, the curve of another negative emotion such as fear would be described as a low quantity of free energy that increases more and more strongly with time. That is, it corresponds to a confident prediction about the maintenance of homeostasis of future interoceptive states, whose uncertainty is inferred to increase more and more strongly over time.

However, the differentiation of emotions by considering only two dimensions is subject to criticism in the literature. For instance, it has been pointed out regarding the widely used two-dimensional model of valence and arousal (Russell, 1980, 2009), which struggles to separate fear and anger, both of which are defined as highly aroused and negative emotions (Fontaine et al., 2007). This criticism could be also formulated with respect to the proposal of emotion differentiation based on the two dimensions of inferred free energy and time (Joffily & Coricelli, 2013). Although this proposal takes into account supplemental information (i.e.,

mapping emotions as curves rather than points), drawing the curve of anger could also be problematic. No anger curve was proposed to our knowledge. One could assume that the anger curve would be drawn in the same way as the fear curve, i.e., a prediction associated with a small quantity of free energy that increases more and more with time. This problem of differentiating fear and anger has been resolved by taking into account more than two dimensions as proposed by appraisal theories (i.e., the appraisal criteria Scherer, 2005b). In this context, fear and anger are differentiated according to the coping appraisal with the consequences of the events, which is low for fear and high for anger. Nevertheless, appraisal theories also struggle to differentiate emotions in some cases, for example when only a few appraisal dimensions – or criteria – are involved in the appraisal process (Moors et al., 2013). A way to gain precision would be to take into account both the different appraisal criteria involved in emotion and how the free energy is predicted to vary over time for these criteria. In other words, this would refer to differentiate emotions according to how the free energy is predicted to vary over time within the different levels of hierarchical causes of the event meaning (i.e., the appraisal criteria). In sum, consideration of ITACHI would allow taking advantage of both appraisal theories and the active inference theory to better differentiate emotions.

### **Conclusion**

The present essay outlined an integrated theory of appraisal and active inference defined as ITACHI (i.e., Integrated Theory of Appraisal under Causal Hierarchical Inference). ITACHI delineates individual's environmental interactions and affective phenomena based on the confronting assumptions of appraisal theories and active inference theory. The starting point of this confrontation was to match the process of criteria appraisal proposed by appraisal theories with that of the hierarchical causes of the brain generative model inferring meaning from the environment proposed by the active inference theory. This matching seems very coherent and

could enrich both theoretical fields by suggesting some answers to key questions. For example, it can help distinguish non-affective episodes from affective episodes such as emotion or mood. Or it can provide us with more precise assumptions to address the challenge of emotion differentiation. Although this paper is only a draft proposal for an integrated appraisal-active inference framework called ITACHI, it opens interesting perspectives for future research on affective phenomena and, more broadly, on the interactions of the individual with his environment.

### References

- Adams, R. A., Shipp, S., & Friston, K. J. (2013). Predictions not commands: Active inference in the motor system. *Brain Structure and Function*, *218*(3), 611-643.  
<https://doi.org/10.1007/s00429-012-0475-5>
- Aue, T., Flykt, A., & Scherer, K. R. (2007). First evidence for differential and sequential efferent effects of stimulus relevance and goal conduciveness appraisal. *Biological Psychology*, *74*(3), 347-357. <https://doi.org/10.1016/j.biopsycho.2006.09.001>
- Barrett, L. F. (2017). The theory of constructed emotion: An active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience*, *12*(1), 1-23.  
<https://doi.org/10.1093/scan/nxs060>
- Barrett, L. F., & Simmons, W. K. (2015). Interoceptive predictions in the brain. *Nature Reviews Neuroscience*, *16*(7), 419-429. <https://doi.org/10.1038/nrn3950>
- Bechara, A., Damasio, H., Tranel, D., & Damasio, A. R. (2005). The Iowa Gambling Task and the somatic marker hypothesis: Some questions and answers. *Trends in Cognitive Sciences*, *9*(4), 159-162. <https://doi.org/10.1016/j.tics.2005.02.002>
- Bhalla, M., & Proffitt, D. R. (1999). Visual-Motor Recalibration in Geographical Slant Perception. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(4), 1076-1096. <https://doi.org/10.1037/0096-1523.25.4.1076>

- Bottemanne, H., Longuet, Y., & Gauld, C. (2022). L'esprit predictif : Introduction à la théorie du cerveau bayésien. *L'Encéphale*, 48(4), 436-444.  
<https://doi.org/10.1016/j.encep.2021.09.011>
- Brosch, T., & Sander, D. (2013). Comment: The Appraising Brain: Towards a Neuro-Cognitive Model of Appraisal Processes in Emotion. *Emotion Review*, 5(2), 163-168.  
<https://doi.org/10.1177/1754073912468298>
- Brosch, T., Sander, D., Pourtois, G., & Scherer, K. R. (2008). Beyond Fear : Rapid Spatial Orienting Toward Positive Emotional Stimuli. *Psychological Science*, 19(4), 362-370.  
<https://doi.org/10.1111/j.1467-9280.2008.02094.x>
- Brosch, T., Sander, D., & Scherer, K. R. (2007). That baby caught my eye... Attention capture by infant faces. *Emotion*, 7(3), 685-689. <https://doi.org/10.1037/1528-3542.7.3.685>
- Brosch, T., & Sharma, D. (2005). The Role of Fear-Relevant Stimuli in Visual Search. *Emotion*, 5(3), 360-364.
- Buckley, C. L., Kim, C. S., McGregor, S., & Seth, A. K. (2017). The free energy principle for action and perception: A mathematical review. *Journal of Mathematical Psychology*, 81, 55-79. <https://doi.org/10.1016/j.jmp.2017.09.004>
- Castellanos, E. H., Charboneau, E., Dietrich, M. S., Park, S., Bradley, B. P., Mogg, K., & Cowan, R. L. (2009). Obese adults have visual attention bias for food cue images: Evidence for altered reward system function. *International Journal of Obesity*, 33(9), 1063-1073. <https://doi.org/10.1038/ijo.2009.138>
- Clark, J. E., Watson, S., & Friston, K. J. (2018). What is mood? A computational perspective. *Psychological Medicine*, 48(14), 2277-2284. <https://doi.org/10.1017/S0033291718000430>
- Colombo, M., & Seriès, P. (2012). Bayes in the Brain—On Bayesian Modelling in Neuroscience. *The British Journal for the Philosophy of Science*, 63(3), 697-723.  
<https://doi.org/10.1093/bjps/axr043>



- Craig, A. D. (2002). How do you feel? Interoception: the sense of the physiological condition of the body. *Nature Reviews Neuroscience*, 3(8), 655-666. <https://doi.org/10.1038/nrn894>
- Craig, A. D. (2003). Interoception: The sense of the physiological condition of the body. *Current Opinion in Neurobiology*, 13(4), 500-505. [https://doi.org/10.1016/S0959-4388\(03\)00090-4](https://doi.org/10.1016/S0959-4388(03)00090-4)
- Dan Glaser, E. S., & Scherer, K. R. (2008). Neuronal Processes Involved in Subjective Feeling Emergence: Oscillatory Activity During an Emotional Monitoring Task. *Brain Topography*, 20(4), 224-231. <https://doi.org/10.1007/s10548-008-0048-3>
- Dukes, D., Abrams, K., Adolphs, R., Ahmed, M. E., Beatty, A., Berridge, K. C., Broomhall, S., Brosch, T., Campos, J. J., Clay, Z., Clément, F., Cunningham, W. A., Damasio, A., Damasio, H., D'Arms, J., Davidson, J. W., de Gelder, B., Deonna, J., de Sousa, R., ... Sander, D. (2021). The rise of affectivism. *Nature Human Behaviour*, 5(7), 816-820. <https://doi.org/10.1038/s41562-021-01130-8>
- Durante, D., & Dunson, D. B. (2018). Bayesian Inference and Testing of Group Differences in Brain Networks. *Bayesian Analysis*, 13(1). <https://doi.org/10.1214/16-BA1030>
- Ellsworth, P. C. (1991). Some Implications of Cognitive Appraisal Theories of Emotion. In *International review of studies on emotion* (K. T. Strongman, Vol. 1, p. 143-161). Wiley.
- Ellsworth, P. C., & Dougherty, A. (2016). Appraisals and Reappraisals in the Courtroom. *Emotion Review*, 8(1), 20-25. <https://doi.org/10.1177/1754073915601227>
- Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance : Attentional control theory. *Emotion*, 7(2), 336-353. <https://doi.org/10.1037/1528-3542.7.2.336>
- Field, M., & Cox, W. (2008). Attentional bias in addictive behaviors : A review of its development, causes, and consequences. *Drug and Alcohol Dependence*, 97(1-2), 1-20. <https://doi.org/10.1016/j.drugalcdep.2008.03.030>

- Folkman, & Lazarus, R. S. (1985). If it changes it must be a process : Study of emotion and coping during three stages of a college examination. *Journal of Personality and Social Psychology*, 48(1), 150-170. <https://doi.org/10.1037/0022-3514.48.1.150>
- Fontaine, J. R. J., Scherer, K. R., Roesch, E. B., & Ellsworth, P. C. (2007). The World of Emotions is not Two-Dimensional. *Psychological Science*, 18(12), 1050-1057. <https://doi.org/10.1111/j.1467-9280.2007.02024.x>
- Forgas, J. P., & Smith, C. (2007). Affect and emotion. In *The SAGE handbook of social psychology* (M. A. Hogg&J. Cooper, p. 146-175).
- Forrest, J., Vogt, J., McDonald, C., Searle, C., & Sakaki, M. (2022). Blind to threat : The presence of temporary goals prevents attention to imminent threat already at early stages of attention allocation. *Motivation Science*, 8(3), 239-251. <https://doi.org/10.1037/mot0000261>
- Fournier, H., & Koenig, O. (2023). Combined effects of intrinsic and goal relevances on attention and action tendency during the emotional episode. *Emotion*, 23(2), 425-436. <https://doi.org/10.1037/emo0001081>
- Fournier, H. & Koenig, O. (Submitted a). From emotion to action tendency: Time course of attentional effects of agonistic and antagonistic overlay of intrinsic and goal relevances.
- Fournier, H. & Koenig, O. (Submitted b). Attentional impact of relevance overlay: Effects of goal conduction and unavoidably goal obstruction inductions on pleasant and unpleasant stimuli.
- Fox, E., Griggs, L., & Mouchlianitis, E. (2007). The detection of fear-relevant stimuli : Are guns noticed as quickly as snakes? *Emotion*, 7(4), 691-696. <https://doi.org/10.1037/1528-3542.7.4.691>
- Frijda, N. H. (1986). *The emotions* (Maison des sciences de l'homme). Cambridge University Press.

- Frijda, N. H. (2004). Emotions and action. In *Feelings and emotions : The Amsterdam symposium* (In A. S. R. Manstead, N. Frijda. A. Fischer (Ed.), p. 158-173). Cambridge university Press.
- Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1456), 815-836.  
<https://doi.org/10.1098/rstb.2005.1622>
- Friston, K. (2008). Hierarchical Models in the Brain. *PLoS Computational Biology*, 4(11), e1000211. <https://doi.org/10.1371/journal.pcbi.1000211>
- Friston, K. (2010). The free-energy principle : A unified brain theory? *Nature Reviews Neuroscience*, 11(2), 127-138. <https://doi.org/10.1038/nrn2787>
- Friston, K. (2011). What Is Optimal about Motor Control? *Neuron*, 72(3), 488-498.  
<https://doi.org/10.1016/j.neuron.2011.10.018>
- Friston, K. (2012a). The history of the future of the Bayesian brain. *NeuroImage*, 62(2), 1230-1233. <https://doi.org/10.1016/j.neuroimage.2011.10.004>
- Friston, K. (2012b). A Free Energy Principle for Biological Systems. *Entropy*, 14(11), 2100-2121. <https://doi.org/10.3390/e14112100>
- Friston, K. (2018). Does predictive coding have a future? *Nature Neuroscience*, 21(8), 1019-1021. <https://doi.org/10.1038/s41593-018-0200-7>
- Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., & Pezzulo, G. (2017). Active Inference : A Process Theory. *Neural Computation*, 29(1), 1-49.  
[https://doi.org/10.1162/NECO\\_a\\_00912](https://doi.org/10.1162/NECO_a_00912)
- Friston, K. J., Daunizeau, J., Kilner, J., & Kiebel, S. J. (2010). Action and behavior : A free-energy formulation. *Biological Cybernetics*, 102(3), 227-260.  
<https://doi.org/10.1007/s00422-010-0364-z>
- Friston, K., Schwartenbeck, P., FitzGerald, T., Moutoussis, M., Behrens, T., & Dolan, R. J.

- (2013). The anatomy of choice : Active inference and agency. *Frontiers in Human Neuroscience*, 7. <https://doi.org/10.3389/fnhum.2013.00598>
- Fromberger, P., Jordan, K., von Herder, J., Steinkrauss, H., Nemetschek, R., Stolpmann, G., & Müller, J. L. (2012). Initial Orienting Towards Sexually Relevant Stimuli : Preliminary Evidence from Eye Movement Measures. *Archives of Sexual Behavior*, 41(4), 919-928. <https://doi.org/10.1007/s10508-011-9816-3>
- Grandjean, D., Sander, D., & Scherer, K. R. (2008). Conscious emotional experience emerges as a function of multilevel, appraisal-driven response synchronization. *Consciousness and Cognition*, 17(2), 484-495. <https://doi.org/10.1016/j.concog.2008.03.019>
- Grandjean, D., & Scherer, K. R. (2008). Unpacking the cognitive architecture of emotion processes. *Emotion*, 8(3), 341-351. <https://doi.org/10.1037/1528-3542.8.3.341>
- Grandjean, D., & Scherer, K. R. (2014). Chapitre 2. Théorie de l'évaluation cognitive et dynamique des processus émotionnels [Chapter 2. Theory of cognitive appraisal and dynamics of emotional processes]. In *Traité de Psychologie des émotions [Treatise on the psychology of emotions]* (In D. Sander, p. 51-87). Dunod. <https://www.cairn.info/traite-de-psychologie-des-emotions--9782100705344-page-51.htm?contenu=resume>
- Gross, J. J. (2015). Emotion Regulation : Current Status and Future Prospects. *Psychological Inquiry*, 26(1), 1-26. <https://doi.org/10.1080/1047840X.2014.940781>
- Grossberg, S. (2009). Cortical and subcortical predictive dynamics and learning during perception, cognition, emotion and action. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1521), 1223-1234. <https://doi.org/10.1098/rstb.2008.0307>
- Harrison, N. A., Voon, V., Cercignani, M., Cooper, E. A., Pessiglione, M., & Critchley, H. D. (2016). A Neurocomputational Account of How Inflammation Enhances Sensitivity to Punishments Versus Rewards. *Biological Psychiatry*, 80(1), 73-81. <https://doi.org/10.1016/j.biopsych.2015.07.018>

- Hesselmann, G., Sadaghiani, S., Friston, K. J., & Kleinschmidt, A. (2010). Predictive Coding or Evidence Accumulation? False Inference and Neuronal Fluctuations. *PLoS ONE*, 5(3), e9926. <https://doi.org/10.1371/journal.pone.0009926>
- Hohwy, J. (2013). *Into the predictive mind* (Hohwy, J.). Oxford, UK: Oxford University Press.
- James, W. (1994). The Physical Basis of Emotion. *Psychological Review*, 101(2), 205-210. (Original work published 1884). <https://doi.org/10.1037/0033-295X.101.2.205>
- Joffily, M., & Coricelli, G. (2013). Emotional Valence and the Free-Energy Principle. *PLoS Computational Biology*, 9(6), e1003094. <https://doi.org/10.1371/journal.pcbi.1003094>
- Kersten, D., Mamassian, P., & Yuille, A. (2004). Object Perception as Bayesian Inference. *Annual Review of Psychology*, 55(1), 271-304. <https://doi.org/10.1146/annurev.psych.55.090902.142005>
- Kirchhoff, M., Parr, T., Palacios, E., Friston, K., & Kiverstein, J. (2018). The Markov blankets of life : Autonomy, active inference and the free energy principle. *Journal of The Royal Society Interface*, 15(138), 20170792. <https://doi.org/10.1098/rsif.2017.0792>
- Laanctôt, N., & Hess, U. (2007). The timing of appraisals. *Emotion*, 7(1), 207-212. <https://doi.org/10.1037/1528-3542.7.1.207>
- Lazarus, R. S. (1991). *Emotion and adaptation*. Oxford University Press.
- Lin, C.-H. S., & Garrido, M. I. (2022). Towards a cross-level understanding of Bayesian inference in the brain. *Neuroscience & Biobehavioral Reviews*, 137, 104649. <https://doi.org/10.1016/j.neubiorev.2022.104649>
- LoBue, V., & DeLoache, J. S. (2008). Detecting the Snake in the Grass : Attention to Fear-Relevant Stimuli by Adults and Young Children. *Psychological Science*, 19(3), 284-289. <https://doi.org/10.1111/j.1467-9280.2008.02081.x>
- Massumi, B. (2002). The autonomy of affect. In *Parables for the Virtual, Movement, Affect*,

*Sensation* (p. 25-48). Duke university press.

- Mazzietti, A., Sellem, V., & Koenig, O. (2014). From stimulus-driven to appraisal-driven attention : Towards differential effects of goal relevance and goal relatedness on attention? *Cognition and Emotion*, 28(8), 1483-1492. <https://doi.org/10.1080/02699931.2014.884488>
- Mogg, K., & Bradley, B. P. (2016). Anxiety and attention to threat : Cognitive mechanisms and treatment with attention bias modification. *Behaviour Research and Therapy*, 87, 76-108. <https://doi.org/10.1016/j.brat.2016.08.001>
- Mogg, K., & Bradley, B. P. (2018). Anxiety and Threat-Related Attention : Cognitive-Motivational Framework and Treatment. *Trends in Cognitive Sciences*, 22(3), 225-240. <https://doi.org/10.1016/j.tics.2018.01.001>
- Moors, A. (2013). On the Causal Role of Appraisal in Emotion. *Emotion Review*, 5(2), 132-140. <https://doi.org/10.1177/1754073912463601>
- Moors, A., Ellsworth, P. C., Scherer, K. R., & Frijda, N. H. (2013). Appraisal Theories of Emotion : State of the Art and Future Development. *Emotion Review*, 5(2), 119-124. <https://doi.org/10.1177/1754073912468165>
- Müller, S., Rothermund, K., & Wentura, D. (2016). Relevance drives attention : Attentional bias for gain- and loss-related stimuli is driven by delayed disengagement. *Quarterly Journal of Experimental Psychology*, 69(4), 752-763. <https://doi.org/10.1080/17470218.2015.1049624>
- Nijs, I. M. T., Muris, P., Euser, A. S., & Franken, I. H. A. (2010). Differences in attention to food and food intake between overweight/obese and normal-weight females under conditions of hunger and satiety. *Appetite*, 54(2), 243-254. <https://doi.org/10.1016/j.appet.2009.11.004>
- Öhman, A., Flykt, A., & Esteves, F. (2001). Emotion drives attention : Detecting the snake in the grass. *Journal of Experimental Psychology: General*, 130(3), 466-478.

<https://doi.org/10.1037/0096-3445.130.3.466>

Öhman, A., & Mineka, S. (2003). The Malicious Serpent : Snakes as a Prototypical Stimulus for an Evolved Module of Fear. *Current Directions in Psychological Science*, *12*(1), 5-9.

<https://doi.org/10.1111/1467-8721.01211>

Pezzulo, G., Rigoli, F., & Friston, K. (2015). Active Inference, homeostatic regulation and adaptive behavioural control. *Progress in Neurobiology*, *134*, 17-35.

<https://doi.org/10.1016/j.pneurobio.2015.09.001>

Pool, E., Brosch, T., Delplanque, S., & Sander, D. (2016). Attentional bias for positive emotional stimuli : A meta-analytic investigation. *Psychological Bulletin*, *142*(1), 79-106.

<https://doi.org/10.1037/bul0000026>

Poppa, T., & Bechara, A. (2018). The somatic marker hypothesis : Revisiting the role of the ‘body-loop’ in decision-making. *Current Opinion in Behavioral Sciences*, *19*, 61-66.

<https://doi.org/10.1016/j.cobeha.2017.10.007>

Prinz, J. (2003). Emotion, psychosemantics, and embodied appraisals. In A. Hatzimoysis (Éd.), *Philosophy and the emotions* (p. 69-86). Cambridge University Press.

Rauss, K., & Pourtois, G. (2013). What is Bottom-Up and What is Top-Down in Predictive Coding? *Frontiers in Psychology*, *4*. <https://doi.org/10.3389/fpsyg.2013.00276>

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161-1178. <https://doi.org/10.1037/h0077714>

Russell, J. A. (2009). Emotion, core affect, and psychological construction. *Cognition & Emotion*, *23*(7), 1259-1283. <https://doi.org/10.1080/02699930902809375>

Sander, D. (2013). Models of emotion : The affective neuroscience approach. In *The Cambridge handbook of human affective neuroscience* (J. Armony & P. Vuilleumier, p. 5-53).

Sander, D., Grandjean, D., & Scherer, K. R. (2005). A systems approach to appraisal

- mechanisms in emotion. *Neural Networks*, 18(4), 317-352.  
<https://doi.org/10.1016/j.neunet.2005.03.001>
- Sander, D., Grandjean, D., & Scherer, K. R. (2018). An Appraisal-Driven Componential Approach to the Emotional Brain. *Emotion Review*, 10(3), 219-231.  
<https://doi.org/10.1177/1754073918765653>
- Scarantino, A. (2017). Do Emotions Cause Actions, and If So How? *Emotion Review*, 9(4), 326-334. <https://doi.org/10.1177/1754073916679005>
- Schachter, S., & Singer, J. (1962). Cognitive, social, and physiological determinants of emotional state. *Psychological Review*, 69(5), 379-399. <https://doi.org/10.1037/h0046234>
- Scherer, K. R. (1984). On the nature and function of emotion : A component process approach. In *Approaches to emotion* (K. R. Scherer & P. Ekman, p. 293-317). Taylor & Francis.
- Scherer, K. R. (1993). Studying the emotion-antecedent appraisal process : An expert system approach. *Cognition & Emotion*, 7(3-4), 325-355.  
<https://doi.org/10.1080/02699939308409192>
- Scherer, K. R. (1999). Appraisal theory. In *Handbook of cognition and emotion* (T. Dalglish & M. J. Power, p. 637-663). John Wiley & Sons.
- Scherer, K. R. (2005a). Unconscious Processes in Emotion : The Bulk of the Iceberg. In L. F. Barrett, P. M. Niedenthal, & P. Winkielman (Éds.), *Emotion and consciousness* (p. 312-334). Guilford Press.
- Scherer, K. R. (2005b). What are emotions? And how can they be measured? *Social Science Information*, 44(4), 695-729. <https://doi.org/10.1177/0539018405058216>
- Scherer, K. R. (2009). The dynamic architecture of emotion : Evidence for the component process model. *Cognition & Emotion*, 23(7), 1307-1351.  
<https://doi.org/10.1080/02699930902928969>



- Scherer, K. R. (2013). The Nature and Dynamics of Relevance and Valence Appraisals : Theoretical Advances and Recent Evidence. *Emotion Review*, 5(2), 150-162.  
<https://doi.org/10.1177/1754073912468166>
- Scherer, K. R. (2019). Studying appraisal-driven emotion processes : Taking stock and moving to the future. *Cognition and Emotion*, 33(1), 31-40.  
<https://doi.org/10.1080/02699931.2018.1510380>
- Scherer, K. R., & Brosch, T. (2009). Culture-specific appraisal biases contribute to emotion dispositions. *European Journal of Personality*, 23(3), 265-288.  
<https://doi.org/10.1002/per.714>
- Scherer, K. R., & Ellgring, H. (2007). Are facial expressions of emotion produced by categorical affect programs or dynamically driven by appraisal? *Emotion*, 7(1), 113-130.  
<https://doi.org/10.1037/1528-3542.7.1.113>
- Scherer, K. R., Ellgring, H., Dieckmann, A., Unfried, M., & Mortillaro, M. (2019). Dynamic Facial Expression of Emotion and Observer Inference. *Frontiers in Psychology*, 10, 508.  
<https://doi.org/10.3389/fpsyg.2019.00508>
- Scherer, K. R., Mortillaro, M., & Mehu, M. (2013). Understanding the Mechanisms Underlying the Production of Facial Expression of Emotion : A Componential Perspective. *Emotion Review*, 5(1), 47-53. <https://doi.org/10.1177/1754073912451504>
- Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Sciences*, 17(11), 565-573. <https://doi.org/10.1016/j.tics.2013.09.007>
- Seth, A. K. (2015). The Cybernetic Bayesian Brain The Cybernetic Bayesian Brain : From Interoceptive Inference to Sensorimotor Contingencies: From Interoceptive Inference to Sensorimotor Contingencies. *Open MIND*. <https://doi.org/10.15502/9783958570108>
- Seth, A. K., & Friston, K. J. (2016). Active interoceptive inference and the emotional brain. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1708),

20160007. <https://doi.org/10.1098/rstb.2016.0007>
- Shapiro, L. (2019). *Embodied Cognition* (2<sup>e</sup> éd.). Routledge.  
<https://doi.org/10.4324/9781315180380>
- Siemer, M. (2009). Mood Experience : Implications of a Dispositional Theory of Moods. *Emotion Review*, 1(3), 256-263. <https://doi.org/10.1177/1754073909103594>
- Siemer, M., Mauss, I., & Gross, J. J. (2007). Same situation--Different emotions : How appraisals shape our emotions. *Emotion*, 7(3), 592-600. <https://doi.org/10.1037/1528-3542.7.3.592>
- Smith, C. A., & Kirby, L. D. (2009). Putting appraisal in context : Toward a relational model of appraisal and emotion. *Cognition & Emotion*, 23(7), 1352-1372.  
<https://doi.org/10.1080/02699930902860386>
- Smith, C., & Lazarus, R. S. (1990). Emotion and adaptation. In *Handbook of personality : Theory and research* (L. A. Pervin, Vol. 21, p. 609-637).
- Smith, R. (2020). The three-process model of implicit and explicit emotion. In *Neuroscience of Enduring Change : Implications for Psychotherapy* (Richard D. Lane&Lynn Nadel, Vol. 3, p. 25-55). Oxford University Press.
- Smith, R., Badcock, P., & Friston, K. J. (2021). Recent advances in the application of predictive coding and active inference models within clinical neuroscience. *Psychiatry and Clinical Neurosciences*, 75(1), 3-13. <https://doi.org/10.1111/pcn.13138>
- Smith, R., Friston, K. J., & Whyte, C. J. (2022). A step-by-step tutorial on active inference and its application to empirical data. *Journal of Mathematical Psychology*, 107, 102632.  
<https://doi.org/10.1016/j.jmp.2021.102632>
- Smith, R., Parr, T., & Friston, K. J. (2019). Simulating Emotions : An Active Inference Model of Emotional State Inference and Emotion Concept Learning. *Frontiers in Psychology*, 10, 2844. <https://doi.org/10.3389/fpsyg.2019.02844>

- Sugawara, A., Terasawa, Y., Katsunuma, R., & Sekiguchi, A. (2020). Effects of interoceptive training on decision making, anxiety, and somatic symptoms. *BioPsychoSocial Medicine*, *14*(1), 7. <https://doi.org/10.1186/s13030-020-00179-7>
- Sun, L., Ding, C., Xu, M., Diao, L., & Yang, D. (2017). Engagement Attentional Bias toward Value-Associated Stimuli. *Current Psychology*, *36*(4), 747-754. <https://doi.org/10.1007/s12144-016-9462-y>
- Sussman, E., Winkler, I., & Schröger, E. (2003). Top-down control over involuntary attention switching in the auditory modality. *Psychonomic Bulletin & Review*, *10*(3), 630-637. <https://doi.org/10.3758/BF03196525>
- Uusberg, A., Yih, J., Taxer, J. L., Christ, N. M., Toms, T., Uusberg, H., & Gross, J. (2023). Appraisal shifts during reappraisal. *Emotion, Advance online publication*. <https://doi.org/10.1037/emo0001202>
- van Peer, J. M., Grandjean, D., & Scherer, K. R. (2014). Sequential unfolding of appraisals : EEG evidence for the interaction of novelty and pleasantness. *Emotion*, *14*(1), 51-63. <https://doi.org/10.1037/a0034566>
- van Reekum, C., Johnstone, T., Banse, R., Etter, A., Wehrle, T., & Scherer, K. (2004). Psychophysiological responses to appraisal dimensions in a computer game. *Cognition & Emotion*, *18*(5), 663-688. <https://doi.org/10.1080/02699930341000167>
- Vogt, J., De Houwer, J., Crombez, G., & Van Damme, S. (2013). Competing for attentional priority : Temporary goals versus threats. *Emotion*, *13*(3), 587-598. <https://doi.org/10.1037/a0027204>
- Vogt, J., Lozo, L., Koster, E. H. W., & De Houwer, J. (2011). On the role of goal relevance in emotional attention : Disgust evokes early attention to cleanliness. *Cognition and Emotion*, *25*(3), 466-477. <https://doi.org/10.1080/02699931.2010.532613>
- Vuilleumier, P. (2005). How brains beware : Neural mechanisms of emotional attention.

- Trends in Cognitive Sciences*, 9(12), 585-594. <https://doi.org/10.1016/j.tics.2005.10.011>
- Werthmann, J., Roefs, A., Nederkoorn, C., Mogg, K., Bradley, B. P., & Jansen, A. (2011). Can(not) take my eyes off it : Attention bias for food in overweight participants. *Health Psychology*, 30(5), 561-569. <https://doi.org/10.1037/a0024291>
- Yantis, S. (1998). Control of visual attention. In *Attention* (H. Pashler, p. 223-256). Hove, England: Psychology Press.
- Yantis, S., & Jonides, J. (1984). Abrupt Visual Onsets and Selective Attention : Evidence From Visual Search. *Journal of Experimental Psychology: Human Perception and Performance*, 10(5), 601-621. <https://doi.org/10.1037/0096-1523.10.5.601>

## Chapitre 8 - Projet THERADIA

Le projet THERADIA a été proposé afin de développer, à terme, une solution pour accompagner les patients souffrant d'altération des fonctions cognitives dans la réalisation d'entraînements cognitifs informatisés (ECI) à domicile visant à remédier à cette altération. À cette fin, le projet a pour objectif de développer un assistant virtuel qui pourra médier les interactions entre les patients, l'ECI, et les acteurs qui gravitent autour des patients (i.e., thérapeutes et aidants). Notre implication dans ce projet s'est concentrée sur la participation au développement de la phase expérimentale visant à mettre en œuvre cet assistant virtuel.

### 1. Introduction

Une partie de la médecine actuelle s'inscrit dans une volonté de conduite du changement vers l'approche dites des « 4P » : Préventive, Personnalisée, Participative et Prédicative. Dans ce contexte, les thérapies numériques, pouvant relever le challenge de répondre à ces quatre critères, y occupent une place de choix. On peut par exemple citer le cas de l'ECI, qui consiste à réaliser de manière répétée des exercices standardisés sur ordinateur afin de maintenir et/ou d'améliorer le fonctionnement cognitif (Hill et al., 2017). L'ECI est principalement proposé en réponse aux pathologies telles que les maladies neurodégénératives, les troubles du développement neurologique ou certains troubles psychiatriques (e.g., schizophrénie, dépression, ...), dans lesquelles les fonctions cognitives sont altérées.

#### a. Altération des fonctions cognitives et état des traitements actuels

À propos de l'altération des fonctions cognitives, une attention particulière a été portée à la démence, le syndrome qui lui est le plus associé. Ce syndrome, chronique et progressif, est caractérisé par une détérioration de la mémoire, du raisonnement, du langage et des fonctions exécutives. Pour l'heure, 55 millions de personnes seraient atteintes de démence à travers le

monde et l'Organisation Mondiale de la Santé (2021) prédit une augmentation du nombre de cas jusqu'à 139 millions d'ici 2050. La démence est un problème majeur de santé publique, car à un stade avancé, l'autonomie et la qualité des vies des individus déclinent fortement. Ce syndrome atteint majoritairement les individus dans le vieillissement et particulièrement dans le cas de maladie d'Alzheimer (associée à environ 60 à 70% des cas), bien que d'autres pathologies puissent en être responsables (e.g., maladie de Parkinson, démence à corps de Lewy, démence vasculaire ; Albert et al., 2011). Actuellement, les symptômes de la maladie d'Alzheimer et de la démence ne rencontrent pas de traitements médicamenteux satisfaisants. Par exemple, l'administration d'inhibiteurs de cholinestérase et de mémantine est associée à de lourds effets secondaires (e.g., troubles digestifs) et ne présente que de modestes améliorations à court terme du fonctionnement cognitif (Lissek & Suchan, 2021 ; Tricco et al., 2013 ; Vaci et al., 2021 ; Wong, 2016).

### **b. L'ECI en réponse à l'altération des fonctions cognitives : des perspectives intéressantes**

Face aux problèmes évoqués précédemment concernant les traitements médicamenteux visant à lutter contre l'altération des fonctions cognitives, l'ECI a connu un intérêt grandissant dans la littérature ces dernières années (Zucchella et al., 2018). L'ECI s'inscrit dans le cadre de la médecine des 4P, car : a) elle a pour objectif de maintenir le fonctionnement cognitif (« Prévenir ») ; b) elle propose des programmes d'entraînement adaptés à chaque patient en ciblant par exemple une fonction cognitive spécifique comme la mémoire ou l'attention (« Personnalisée ») ; c) elle offre un entraînement qui rend les patients acteurs de leur thérapie (« Participative ») ; et d) elle permet le suivi des résultats des patients au fur et à mesure des séances et fournit une information sur l'évolution future potentielle de l'altération des fonctions cognitives (« Prédictive »). Les études s'intéressant à l'efficacité de l'ECI en réponse à l'altération des fonctions cognitives montrent des résultats encourageants, comme cela a été

montré chez les individus âgés et en bonne santé (Lampit et al., 2014), atteints de la maladie de Parkinson (Gavelin et al., 2022 ; Leung et al., 2015) ou de troubles cognitifs légers (Hill et al., 2017; Hu et al., 2021). Les troubles cognitifs légers réfèrent à un état transitif situé entre le vieillissement normal et la démence, dans lequel l'altération des fonctions cognitives est supérieure à celle attendue lors d'un vieillissement normal, sans pour autant empêcher l'autonomie des individus (Belleville et al., 2016).

### **c. Les problèmes inhérents à l'ECI**

Bien que l'utilisation de l'ECI en réponse à l'altération des fonctions cognitives semble être intéressante, elle rencontre certaines limites. Premièrement, l'ECI semble inefficace pour lutter contre le déclin cognitif lorsque la démence est cliniquement diagnostiquée (Hill et al., 2017). Aussi, ce traitement doit être prescrit en amont, lors de la détection de troubles cognitifs légers (Albert et al., 2011; Belleville et al., 2016). Deuxièmement, l'efficacité de l'ECI dans la préservation et/ou l'amélioration des fonctions cognitives n'est pas systématiquement rapportée dans la littérature (Butler et al., 2018 ; Gaitán et al., 2013 ; Kallio et al., 2017). Ce manque de consensus pourrait être expliqué par des problèmes méthodologiques (e.g., taille d'échantillon insuffisante, manque de clarté dans les méthodes de randomisation ; Clare & Woods, 2004 ; Gates et al., 2019 ; Kallio et al., 2017 ; Zucchella et al., 2018). De plus, la réalisation de l'ECI à domicile, sans supervision, serait inefficace sur l'amélioration des fonctions cognitives par rapport à une réalisation supervisée en groupe (Lampit et al., 2014). Cette absence d'efficacité pourrait émerger d'une plus faible implication, dans la mesure où la réalisation à domicile n'inclut pas d'interaction sociale, ni de soutien ou d'intervention de la part du thérapeute (e.g., encouragements, explications personnalisées des exercices). Or, l'efficacité de l'ECI semble être en grande partie tributaire de la motivation des patients résultant de leur implication (Jaeggi et al., 2014). Ce point constitue un problème majeur dans les perspectives d'utilisation de l'ECI.

En effet, l'argument fort de l'ECI selon lequel les patients peuvent suivre ce type de thérapie depuis leur domicile, sans nécessité la présence d'un thérapeute, semble être à nuancer.

L'utilisation d'une assistance artificielle pourrait être une solution concevable afin de conserver une forme d'interaction sociale dans l'ECI lors de son utilisation en autonomie à domicile. De fait, il a été montré que les technologies d'assistance informatisée, comme les robots sociaux, étaient particulièrement appréciées par les populations âgées (Góngora Alonso et al., 2019). Également, l'assistance fournie par ce type de robots lors de la réalisation d'entraînements cognitifs semble augmenter l'efficacité de la remédiation sur la mémoire et les fonctions exécutives des individus âgés souffrant de troubles cognitifs légers (Park et al., 2021). Cependant, l'utilisation systématique de robots sociaux accompagnant les thérapies par ECI est utopique pour l'heure, car une telle mise en œuvre engagerait un coût financier bien trop important.

## **2. Présentation du projet THERADIA**

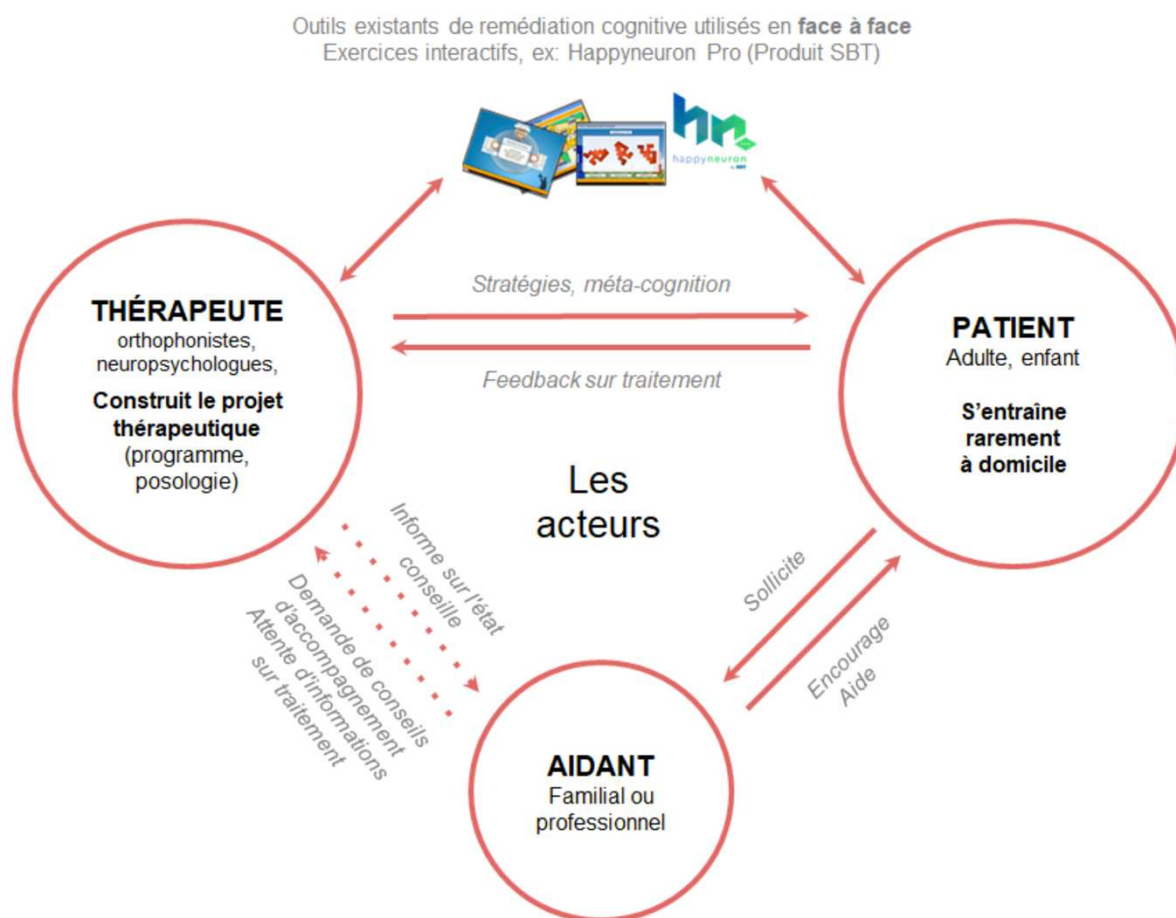
Le projet THERADIA a pour but de développer un assistant virtuel directement implémenté au sein des programmes d'ECI, afin de proposer une solution d'accompagnement viable tant sur le plan financier que de la faisabilité. Cet assistant virtuel sera représenté sous la forme d'un avatar féminin affiché à l'écran, capable d'échanger verbalement avec les patients. L'implémentation de cet assistant virtuel vise à répondre à la problématique de l'efficacité de la remédiation cognitive par ECI mentionnée précédemment dans sa réalisation à domicile. Pour ce faire, l'assistant virtuel doit être en mesure de conseiller les patients, de s'adapter à leurs ressentis et d'ajuster de manière adéquate les exercices proposés durant l'ECI. De plus, l'utilisation de l'assistant virtuel intégré à l'ECI a pour ambition de faciliter les interactions des acteurs qui gravitent autour du patient. En d'autres termes, il s'agit pour l'assistant virtuel d'informer le thérapeute et les aidants, qu'ils soient des proches ou des professionnels, par le biais de bilans issus des séances, quant à l'observance de l'ECI ainsi que les états du patient qui



seront recueillis pendant les séances. Les modèles actuels et prévisionnels (à la suite de l'implémentation de l'assistant virtuel) du fonctionnement des échanges entre le thérapeute, les aidants, le patient et l'ECI sont représentés dans la Figure 1 et la Figure 2, respectivement.

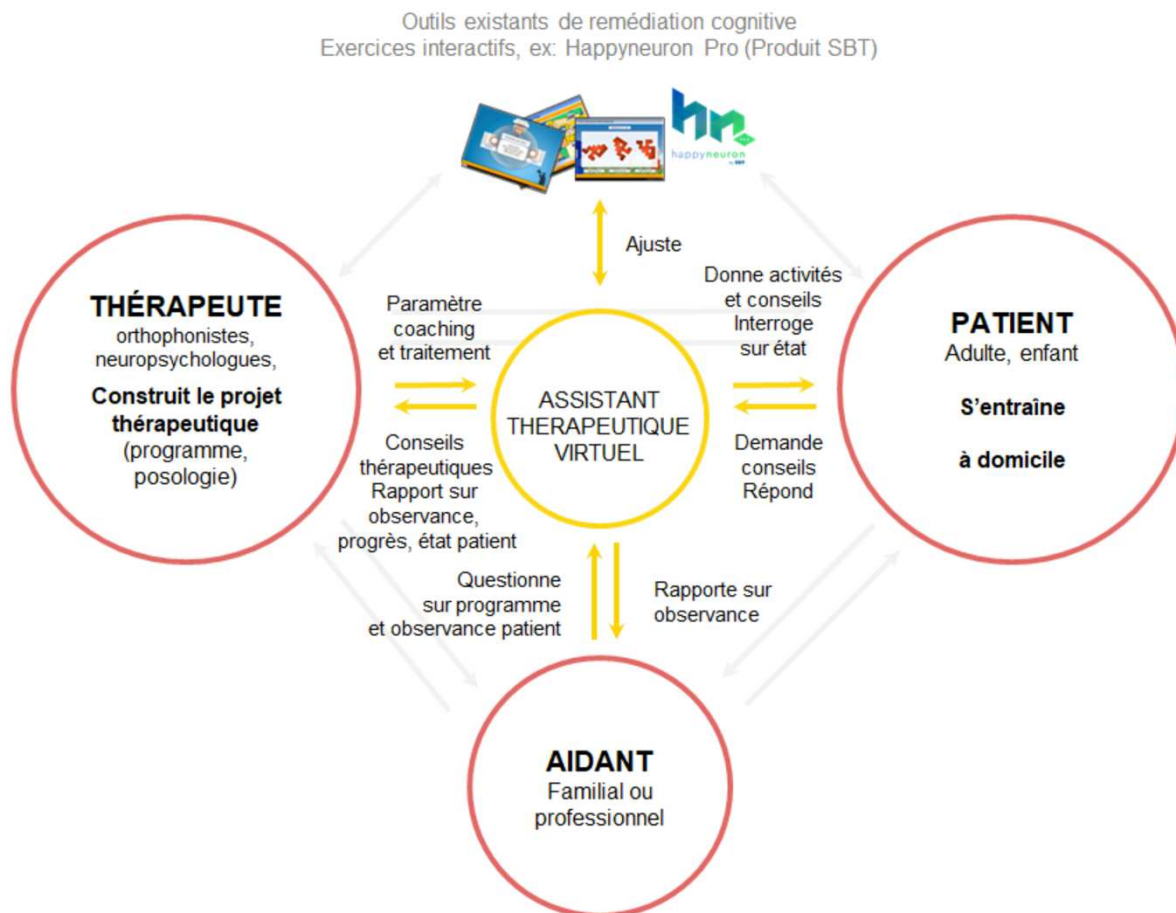
### Figure 1

*Modèle actuel du fonctionnement des échanges entre le thérapeute, les aidants, le patient et l'ECI.*



**Figure 2**

*Modèle prévisionnel du fonctionnement des échanges entre le thérapeute, les aidants, le patient et l'ECI grâce à l'implémentation de l'assistant virtuel.*



### a. L'organisation du projet et ses partenaires

Ce projet a démarré en 2020 et est prévu sur 5 ans. Il a réuni un consortium de partenaires académiques et industriels visant à développer l'assistant thérapeutique virtuel (Tarpin-Bernard et al., 2021). Ce projet a été financé par la Banque Publique d'Investissement (BPI) dans le cadre de la convention de subvention THERADIA, par l'Association Nationale de la Recherche et de l'Innovation (ANRRI), par l'Agence Nationale de la Recherche (ANR) ainsi que la Recherche et de la Technologie (ANRT), sous la convention de subvention n° 2019/0729.

### i. Les partenaires du projet

Le consortium du projet THERADIA s'organise autour de 6 partenaires, décrits dans le Tableau 1, dont les compétences sont complémentaires pour le développement de l'assistant thérapeutique dans le cadre de l'ECI.

**Tableau 1**

*Les partenaires du projet THERADIA, leur type de structure et leurs compétences pour le projet*

Partenaire	Type	Compétences
SBT	PME	Thérapies digitales, remédiation cognitive
PERTIMM	PME	Agents conversationnels
ATOS	Groupe	Détection émotions, SI santé
EMC	Laboratoire	Cognition, émotion, validation
LIG	Laboratoire	Inférence des états mentaux, méthodologie expérimentale pour la collecte et l'annotation de données multimodales, génération de résumé
GISPA-lab	Laboratoire	Génération de comportements multimodaux interactifs, synthèse de parole expressive, styles

#### 1. SBT Human(s) Matter

La société SBT Human(s) Matter est une entreprise qui propose des ECI à travers sa filiale Happyneuron, dont les programmes de remédiation sont multiples, appliqués dans le cadre de troubles dys, d'AVC, de la maladie d'Alzheimer et également de troubles psychiatriques.

#### 2. PERTIMM

La société PERTIMM a créé la plateforme Viki.ai, basée sur l'intelligence artificielle et le traitement automatique du langage naturel. Cette plateforme permet de développer des agents conversationnels dans de nombreux domaines, y compris celui de la santé.

### 3. ATOS

Le groupe ATOS est spécialisé dans la transformation digitale et particulièrement dans l'intelligence artificielle et l'intégration de systèmes.

### 4. EMC

L'EMC est un laboratoire de psychologie cognitive. Au sein de l'EMC, l'équipe MEA, impliquée dans le projet THERADIA, s'intéresse à l'étude des mécanismes cognitifs tels que les processus mnésiques, émotionnels et attentionnels. Le laboratoire EMC travaille également sur la prise en charge du déclin cognitif dû au vieillissement normal et pathologique via l'entraînement cognitif et physique.

### 5. LIG

Le LIG est un laboratoire d'informatique. Son équipe GETALP, qui participe au projet THERADIA, aborde les aspects théoriques et méthodologiques de la communication et du traitement de l'information multilingue à l'aide de l'intelligence artificielle. Par exemple, un intérêt particulier est porté à la détection automatique d'états mentaux, à la collecte de données multimodales du comportement humain et à la génération automatique de résumés.

### 6. GIPSA-lab

Le GIPSA-lab est un laboratoire qui s'intéresse à l'étude des signaux et systèmes complexes. Plus particulièrement, son équipe CRISSP, associée au projet THERADIA, se concentre sur le développement de technologies vocales et leur utilisation dans des systèmes interactifs.

## **b. Les problématiques liées au développement de l'assistant virtuel**

Dans le cadre du développement de l'assistant virtuel accompagnant l'ECI, plusieurs problématiques ont été ciblées.

### **i. Le visuel de l'assistant virtuel**

Dans un premier temps, le développement de l'aspect de l'assistant virtuel sous la forme d'un avatar a été réalisé. Cette partie a été sous-traitée par la société Dynamixyz, spécialisée

dans la capture de mouvements faciaux. Le visuel 3D de l'avatar est représenté dans la Figure 3.

### **Figure 3**

*Visuel de l'avatar de l'assistant virtuel développé par la société Dynamixyz*



#### **ii. Construction des dialogues**

Afin que l'assistant virtuel puisse accompagner le patient durant les séances, cinq types d'interactions pertinentes ont été définies : a) le dialogue d'accueil proposé en amont de la réalisation des exercices qui a pour objectif de collecter des données relatives au patient telles que son humeur ou sa motivation ; b) le dialogue d'introduction des exercices, proposé avant chaque nouvel exercice, qui vise à présenter le fonctionnement de chaque exercice que le patient s'apprête à réaliser ou à lui rappeler ses performances précédentes en cas de répétition ; c) le dialogue en cours des exercices qui ambitionne de traiter les états du patient qui pourraient nécessiter un arrêt de la tâche ; d) le dialogue de feedback qui se produit à chaque fin d'exercice et qui vise à commenter (positivement) la performance du patient pour renforcer sa motivation ; et enfin e) le dialogue de fin qui se déroule à l'issue de l'ECI et qui a pour but de féliciter le patient pour ses réalisations durant la séance et d'organiser de nouveaux rendez-vous.

### **iii. L'intégration d'une intelligence émotionnelle à l'assistant virtuel**

Dans l'objectif de maintenir l'efficacité de la remédiation cognitive apportée par l'ECI avec l'ajout de l'assistant virtuel, il est nécessaire que les interactions de l'avatar avec le patient soient adaptées. Pour cela, trois axes ont été identifiés comme nécessaires, à savoir, la capacité à exprimer de l'empathie envers les patients ; la capacité à transmettre des émotions positives aux patients, ainsi que la capacité à détecter les états mentaux des patients et à adapter sa posture de façon appropriée.

### **iv. La génération de comportements multimodaux interactifs**

Afin de rendre les interactions fluides, l'assistant virtuel doit être en mesure d'adapter ses interactions selon les différentes situations. À cette fin, la génération de comportements de l'assistant doit être développée dans un paradigme dit d'agent actif. Ce paradigme correspond à un système de synthèse de comportements capable de délivrer des énoncés dans un style d'expression qui dépend de ceux de son interlocuteur.

### **v. La génération de comptes-rendus synthétiques pertinents**

Afin que l'assistant virtuel soit en mesure de délivrer des bilans aux thérapeutes à la suite des séances, il est essentiel d'identifier, d'agréger, de sélectionner et de structurer les informations à leur communiquer. Ces informations nécessitent ensuite d'être structurées dans un texte cohérent.

### **vi. L'optimisation du traitement**

L'assistant virtuel doit également être en mesure d'adapter les types d'exercices proposés aux patients en fonction de leurs besoins (e.g., proposer des exercices permettant d'améliorer la mémoire dans le cas de pathologies affectant cette fonction cognitive). Il est ainsi indispensable de proposer des niveaux de difficulté adéquats aux patients selon leur niveau de compétence.

### **c. Phase expérimentale répondant à ces problématiques**

Dans le but de répondre aux problématiques évoquées précédemment, une phase expérimentale a été réalisée à l'aide du paradigme du magicien d'Oz. Ce paradigme consistait à faire croire à des participants qu'ils interagissaient avec l'assistant virtuel, tandis qu'une expérimentatrice le contrôlait à distance. Dans le cadre de notre étude, les participants devaient réaliser des séances d'ECI accompagnés par un magicien d'Oz. La vidéo et l'audio de chaque séance étaient à la fois enregistrés pour le magicien d'Oz et les participants. Cette étude avait pour objectif de fournir un corpus de données nécessaire au développement de l'assistant virtuel. Concernant le développement de l'avatar, les données visuelles et audio issues des prises de parole de l'expérimentatrice avaient pour objectif de créer des modèles de synthèse vocale et d'expression faciale visant à rendre l'avatar autonome. Concernant la construction des dialogues et la génération d'interactions fluides, l'étude avait pour objectif de tester et d'améliorer des scripts de dialogues en situation réelle, écrits par les scénaristes du projet, dans le but de les implémenter ultérieurement dans un paradigme d'agent actif. Pour ce qui est de doter l'avatar d'une intelligence émotionnelle, les données visuelles et audio ont été recueillies afin, à terme, de créer des modèles informatiques visant à détecter les émotions des patients. Dans la mesure où nous avons majoritairement contribué à cet axe de recherche, une section sera consacrée à ce dernier axe plus loin dans cet écrit. Afin de générer une panoplie d'émotions la plus large possible lors des séances, des inductions ont été attribuées à certains participants en « truquant » les explications du magicien d'Oz (e.g., expliquer que le niveau d'un exercice est facile alors qu'il est en réalité très difficile), dans le but de générer des réactions de surprise, voire de frustration. Les données ont également servi à développer l'architecture des comptes rendus synthétiques pertinents pour les thérapeutes et les aidants. Enfin, un autre objectif de l'étude était de tester la manière dont le niveau de difficulté devait être ajusté par l'assistant virtuel.

### 3. Méthode de la phase expérimentale

#### a. Participants

Trois groupes de participants ont pris part à l'étude : deux groupes d'adultes sans pathologie, l'un formé de participants âgés de 60 ans et plus, l'autre de participants jeunes de 18 à 30 ans, et un groupe d'adultes âgés de 72 à 79 ans, atteints de troubles cognitifs légers. Bien que l'ECI proposé dans le projet THERADIA cible en priorité les adultes âgés, les jeunes adultes peuvent également souffrir de troubles cognitifs dans certaines situations (e.g., à la suite d'un AVC, ou d'une maladie psychiatrique). Afin que l'assistant virtuel puisse, à terme, être implémenté dans des ECI proposés à ce genre de population, nous nous sommes intéressés à développer des dialogues adaptés pour les individus jeunes, dont les capacités, besoins et compétences informatiques diffèrent de ceux des individus âgés. Au total, 113 participants ont pris part à l'étude, dont 52 jeunes adultes (18-27 ans ;  $M_{\text{âge}} = 21,2$  ;  $SD = 2,9$  ; femmes = 35), 52 adultes âgés (60-79 ans ;  $M_{\text{âge}} = 67,9$  ;  $SD = 5,1$  ; femmes = 40) et 9 adultes âgés atteints de troubles cognitifs légers (72-79 ans ;  $M_{\text{âge}} = 75,78$  ;  $SD = 2,1$  ; femmes = 2).

Tous les participants étaient francophones, avaient une vision et une audition normales ou corrigées à la normale, et ont donné leur consentement avant de commencer l'expérience. En ce qui concerne le niveau d'éducation, 60 % des jeunes adultes avaient obtenu le baccalauréat, 31 % avaient obtenu un diplôme de licence, 6 % avaient obtenu un master ou un diplôme de niveau supérieur, et 2 % n'avaient pas obtenu de diplôme. Concernant les adultes âgés, 40 % d'entre eux étaient titulaires d'un master ou d'un diplôme de niveau supérieur, 33 % d'un diplôme de licence et 27 % du baccalauréat. Quant aux adultes âgés atteints de troubles cognitifs légers, 22% avaient obtenu un diplôme de niveau master ou plus, 22% un diplôme de niveau licence, et 56% un diplôme de niveau CAP, équivalent ou inférieur.

Les jeunes adultes ont été recrutés sur le campus de l'Université Lumière Lyon 2, par le biais d'annonces ainsi que par une diffusion sur les réseaux sociaux tels que Facebook. Pour les



adultes âgés, deux annonces ont été publiées dans des journaux régionaux : *Le Progrès* et *Le Dauphine Libéré*. Une campagne de recrutement de personnes âgées a également été menée en s'adressant aux personnes inscrites dans un programme d'enseignement ouvert aux personnes de tous âges ("Université Tous Âges") rattaché à l'Université Lumière Lyon 2. Les participants âgés atteints de troubles cognitifs légers ont été recrutés par le biais de médecins ou de neuropsychologues ayant proposé à leurs patients de participer à l'expérience. A la fin de l'expérience, chaque participant a reçu une carte cadeau de 20€ en guise de récompense. Cette étude a été approuvée par le Comité d'éthique de l'Université de Grenoble Alpes (CERGA-Avis-2021-1).

### **b. Matériel**

Le matériel a nécessité la sélection de quatre exercices de la plateforme d'ECI HappyNeurons. Deux dispositifs informatiques ont été utilisés pour le magicien d'Oz et les participants comprenant : un ordinateur Dell 24', une webcam et un casque avec microphone intégré. Un iPhone X a été utilisé en complément pour l'enregistrement audio et vidéo du côté participant. L'avatar de l'assistant virtuel a été développé par la société Dynamixyz, et contrôlé par une expérimentatrice du projet. Des questionnaires interrogeant différentes dimensions psychologiques, affectives et cognitives des participants ont été utilisés, ainsi qu'un questionnaire visant à évaluer le protocole (i.e., l'assistant virtuel et l'ECI). Ces questionnaires sont reportés dans le Tableau 2. Les passations se sont déroulées dans deux salles expérimentales du laboratoire EMC.

**Tableau 2**

*Questionnaires utilisés pour évaluer les différentes dimensions psychologiques, affectives et cognitives des participants rangés par dimensions et sous-dimensions étudiées*

<b>Dimension</b>	<b>Sous-dimension</b>	<b>Questionnaire</b>
Fonctionnement psycho-affectif	Expérience affective globale	Positive and Negative Affect Schedule (PANAS ; Watson et al., 1988)
	État affectif actuel	Version modifié du Self Assessment Manikin (SAM ; Bradley & Lang, 1994)
		Brief Mood Introspection Scale (BMIS ; Mayer & Gaschke, 1988)
	Anxiété	State-Trait Anxiety Inventory (STAI-Y; Gauthier & Bouchard, 1993)
Personnalité	Extraversion Agréabilité Conscience professionnelle Stabilité émotionnelle Ouverture d'esprit	Ten-Item Personality Inventory (TIPI ; Gosling et al., 2003)
Motivations personnelles	Motivation intrinsèque Motivation extrinsèque Amotivation	Global Motivation Scale (GMS-28 ; Guay et al., 2003)
Habitudes et capacité cognitives	Habitudes et capacité cognitives	Questionnaire réalisé au sein du projet
Fonctionnement cognitif (questionnaire seulement proposé aux adultes âgés)	Fonctionnement cognitif global	Mini-Mental State Examination (MMSE ; Folstein et al., 1975)
	Fonctions exécutives	Trail Making Test (TMT ; Reitan, 1958)
		Frontal Assessment Battery (FAB ; Dubois et al., 2000)
Mémoire	5 Mots de Dubois (5MD ; Dubois et al., 2002)	
Évaluation du protocole	Avatar ECI	Debriefing réalisé au sein du projet

### **c. Design**

L'ECI consistait en 4 exercices sélectionnés sur la plateforme HappyNeuronPro à réaliser deux fois chacun. Les exercices sélectionnés sollicitaient différentes fonctions cognitives telles que la mémoire, le langage, l'attention et la planification.

Pendant la séance d'ECI, le participant était guidé par l'assistant virtuel et pouvait interagir avec lui. L'ECI a été réalisé via un logiciel développé à cet effet par la société Atos qui permettait d'alterner entre les apparitions de l'assistant virtuel et celles des exercices. Bien qu'il fût expliqué aux participants que l'assistant virtuel était piloté par une intelligence artificielle, une expérimentatrice le pilotait à distance. Plus précisément, l'expérimentatrice se trouvait dans une autre pièce et le participant n'était pas informé de son existence (voir Figure 4A et 4B pour un aperçu du dispositif).

L'expérimentatrice était assise devant un ordinateur Dell identique à celui du participant. A l'aide d'une caméra haute définition, nous avons utilisé la solution de capture de mouvements faciaux proposée par la société Dynamixyz pour transformer, en temps réel, le visage de l'expérimentatrice en un avatar féminin en 3D. L'avatar apparaissait sur un fond 3D simulant le cabinet d'un professionnel de santé, similaire à ceux des neuropsychologues ou des orthophonistes qui pratiquent habituellement la remédiation cognitive. L'image de l'avatar était transmise en temps réel sur l'écran du participant, via le logiciel développé par Atos. Une webcam transmettait également le visage du participant en temps réel sur l'écran de l'expérimentatrice, afin que celle-ci puisse suivre les mouvements et le regard du participant pendant les discussions pour les rendre plus naturelles. La vidéo et l'audio de l'expérimentatrice et du participant étaient enregistrées. L'expérimentatrice et le participant communiquaient via des casques avec microphones intégrés.

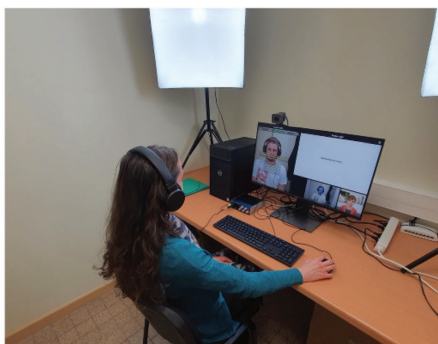
Le discours de l'assistant virtuel était scénarisé (scripté) et apparaissait sur l'écran de l'expérimentatrice qui pouvait ainsi le lire et le faire défiler (voir Figure 4C pour une vue

détaillée de l'écran de l'expérimentatrice). La trame principale du discours de l'assistant était donc identique d'un participant à l'autre, mais, en cas de besoin, le dispositif permettait à l'expérimentatrice d'intervenir librement à tout moment de la session pour aider les participants en cas de question ou de difficulté. En cas de problème technique ne pouvant être résolu par l'assistant virtuel, l'expérimentatrice prévenait un second expérimentateur afin que celui-ci intervienne.

#### Figure 4

*Dispositif du magicien d'Oz et visuel de l'écran de l'expérimentatrice*

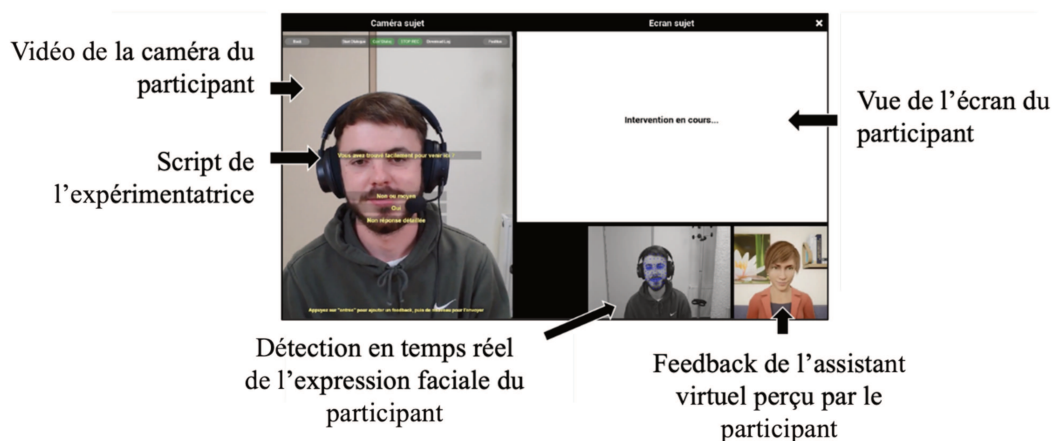
*A. Box de l'expérimentatrice*



*B. Box du participant*



*C. Visuel de l'écran de l'expérimentatrice*



#### **d. Procédure**

Avant de se rendre à l'Université Lumière Lyon 2 pour effectuer la ou les séances d'ECI assisté par l'assistant virtuel (le magicien d'Oz), les participants étaient attribués à un groupe. La différence entre les groupes résidait dans le fait que des inductions étaient associées à certains exercices ou non et que les participants allaient réaliser une ou deux séances. Vingt participants jeunes et 20 adultes âgés étaient attribués aux groupes sans induction et ont réalisé 2 séances de l'expérience espacées d'environ une semaine. Trente-deux participants jeunes et 32 adultes âgés ont été attribués aux groupes avec induction et n'ont réalisé qu'une seule séance de l'expérience. Les inductions n'étaient pas proposées aux participants âgés atteints de troubles cognitifs légers et ces participants n'ont réalisé qu'une seule séance. Les participants assignés aux groupes d'induction étaient répartis en 4 sous-groupes (A, B, C et D) afin de contrebalancer les différentes inductions selon un plan en carré latin. Les différents types d'induction attribués aux sous-groupes sont mentionnés dans le Tableau 3.

La durée moyenne des séances était d'environ 2 heures, mais il se pouvait parfois qu'une séance s'étende jusqu'à 4 heures pour les participants les plus âgés. Après avoir rempli le formulaire de consentement, les participants répondaient à tous les questionnaires mentionnés dans le tableau Tableau 2 (à l'exception du questionnaire de debriefing). Une pause était proposée à la fin de cette première partie, et les participants étaient informés qu'ils pouvaient en reprendre une à tout moment. Ensuite, les participants étaient installés devant l'ordinateur et équipés du casque avec microphone pour réaliser l'ECI en compagnie de l'assistant. Pour cette deuxième partie de l'expérience, l'expérimentateur quittait le box expérimental et laissait les participants effectuer les sessions seuls.

L'assistant virtuel apparaissait alors à l'écran, posait quelques questions aux participants (e.g., nom, âge) et engageait la discussion sur des sujets plus personnels (e.g., travail, loisir). Ce premier dialogue était rédigé de manière à mettre les participants à l'aise et à les habituer à

interagir avec l'assistant. L'assistant expliquait l'intérêt de l'ECI pour l'entraînement des fonctions cognitives et présentait le déroulement de la session en posant régulièrement des questions aux participants.

Avant chaque exercice, l'assistant virtuel présentait la tâche et expliquait de manière interactive quelles fonctions cognitives allaient être entraînées. Ensuite, l'assistant disparaissait pour la durée de l'exercice, mais pouvait réapparaître pour intervenir si le participant rencontrait des difficultés à terminer l'exercice. Après chaque exercice, l'assistant virtuel demandait aux participants leur ressenti sur celui-ci, puis leur donnait un feedback sur leur performance, et parfois des conseils pour s'améliorer et des encouragements pour l'exercice suivant. Chaque exercice était effectué deux fois, de sorte que les participants réalisaient 8 exercices au total. Le niveau de difficulté de la deuxième répétition de l'exercice était adapté en fonction de la performance de la première fois. Pour les groupes avec induction, un exercice sur deux était associé à une induction. Ces inductions consistaient à manipuler le niveau de difficulté de la tâche annoncé par l'expérimentateur, le niveau de difficulté réel et le feedback. Par exemple, l'assistant virtuel allait expliquer qu'un exercice serait facile alors que son niveau était réglé sur difficile, puis donner un feedback très positif, quelle que soit la performance des participants (les différents types d'induction en fonction des groupes sont reportés dans le Tableau 3).

Après le dernier exercice, l'assistant virtuel demandait aux participants comment s'était déroulée la séance, s'ils l'avaient appréciée, quels exercices ils avaient préférés et pourquoi. L'assistant remerciait ensuite les participants avant de mettre fin à l'ECI.

Dans la dernière partie de l'expérience, les participants répondaient au questionnaire de debriefing de l'assistant virtuel et de la session avec l'aide de l'expérimentateur. Dans le cas d'une seconde séance (selon le groupe assigné aux participants), le déroulement était identique. Enfin, les participants ont été informés du dispositif du Magicien d'Oz et ont été invités à

rencontrer l'expérimentatrice (cette phase ne se déroulait que lors de la deuxième séance pour les participants qui ont réalisé deux séances).

**Tableau 3**

*Les différents types d'induction associés aux exercices selon le groupe de participants*

<b>Groupe</b>	<b>Ordre d'exécution</b>	<b>Tâche présentée</b>	<b>Tâche réelle</b>	<b>Type de feedback</b>	<b>Feedback</b>
A	1	Facile	Facile	Adéquation	Bien joué
B	1	Facile	Facile	Adéquation	Excellent, félicitations
C	1	Facile	Facile	Truqué à la baisse	Peut mieux faire
D	1	Facile	Facile	Truqué à la baisse	Très critique
D	3	Facile	Difficile	Adéquation	Peut mieux faire
A	3	Facile	Difficile	Adéquation	Très critique
B	3	Facile	Difficile	Truqué à la hausse	Bien joué
C	3	Facile	Difficile	Truqué à la hausse	Excellent, félicitations
C	2	Difficile	Facile	Adéquation	Bien joué
D	2	Difficile	Facile	Adéquation	Excellent, félicitations
A	2	Difficile	Facile	Truqué à la baisse	Peut mieux faire
B	2	Difficile	Facile	Truqué à la baisse	Très critique
B	4	Difficile	Difficile	Adéquation	Peut mieux faire
C	4	Difficile	Difficile	Adéquation	Très critique
D	4	Difficile	Difficile	Truqué à la hausse	Bien joué
A	4	Difficile	Difficile	Truqué à la hausse	Excellent, félicitations

*Note. La colonne Tâche présentée fait référence à la manière dont l'assistant virtuel présentait la tâche ; la colonne Tâche réelle réfère au niveau de difficulté réel de l'exercice ; la colonne Type de feedback fait référence à la manière dont le feedback de l'assistant virtuel était truqué ; La colonne Feedback au participant précise la manière dont le feedback de l'assistant virtuel était truqué*



## **4. Usage du corpus de données récoltées durant la phase d'expérimentation : focus sur la partie détection d'émotion**

La récolte des données de la phase d'expérimentation est terminée à l'heure actuelle (avril 2023). Les partenaires du consortium ont, dès lors, pu commencer à travailler sur le corpus de données récoltées selon leurs intérêts respectifs. Nous proposons à présent de nous concentrer sur la partie visant à construire les modèles de détection de l'émotion des participants, partie sur laquelle nous avons collaboré. L'objectif de cette partie était de construire une base de données, dans laquelle les séquences audio-visuelles d'interactions récoltées durant les séances étaient évaluées par des annotateurs, afin d'entraîner les modèles de détection de l'émotion. Pour l'heure, ce travail a été réalisé sur les données des séances réalisées par les participants âgés (sans troubles cognitifs légers).

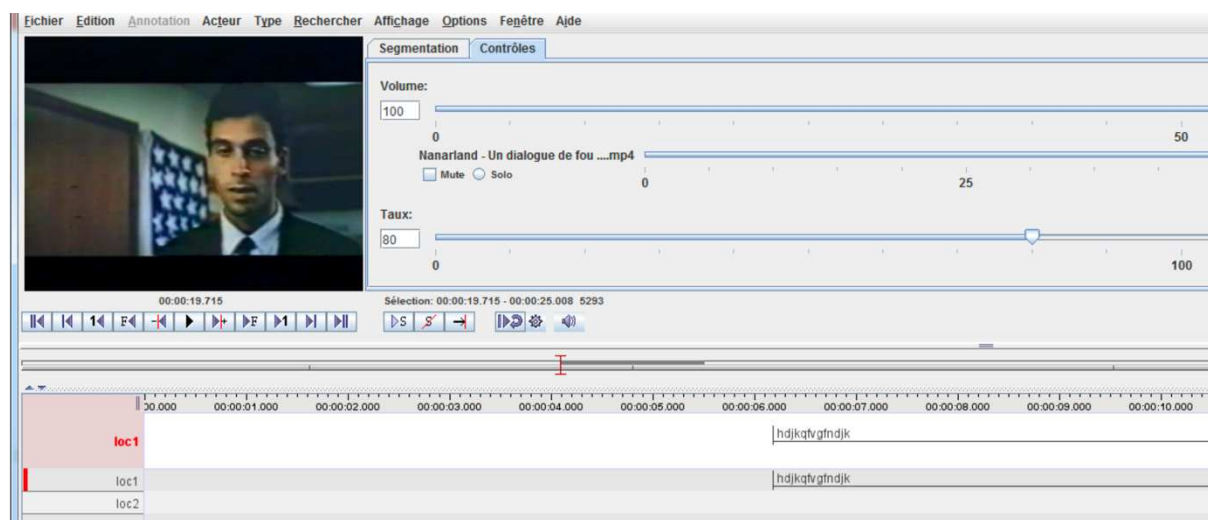
### **a. Segmentation des séquences audio-visuelles**

Dans un premier temps, tous les enregistrements vidéo des séances concernant les 52 adultes âgés sans troubles cognitifs légers ont été segmentés en extraits dans lesquels les échanges observés ont été retranscrits. Ce travail a seulement été réalisé sur les enregistrements audiovisuels des participants (et non sur ceux de l'expérimentatrice) dans lesquels la caméra était centrée sur leur visage. La segmentation/retranscription a été réalisée à l'aide du logiciel ELAN pour un total d'une centaine d'heures d'enregistrement (voir Figure 5 pour un aperçu du fonctionnement du logiciel ELAN). Ce travail a été réalisé par des étudiants en master de sciences du langage de l'Université de Grenoble dans le cadre d'un stage au LIG. La segmentation des enregistrements consistait en la séparation de chaque proposition les unes des autres. La notion de proposition fait référence aux suites continues ou discontinues de mots qui portaient une même information. La segmentation était effectuée sur la base des échanges verbaux ainsi que d'indices comportementaux/non verbaux rattachés ou non à la proposition.

En d'autres termes, la prosodie et les expressions faciales ne servaient pas de support temporel à la segmentation, mais d'éléments contextuels, qui étaient intégrés à la proposition s'ils étaient en relation avec la proposition ou à proximité temporelle immédiate. Par exemple, si un participant affichait une expression faciale notable juste avant de s'exprimer, cette expression devait être contenue dans la proposition. Cela était également le cas si l'expression faciale intervenait après l'expression verbale. Chaque prise de parole durant les propositions était retranscrite à l'écrit, ainsi que les expressions non verbales (i.e., prosodiques, faciales ou posturales à l'aide de l'abréviation <nv>).

### Figure 5

*Capture d'écran du logiciel d'annotation ELAN utilisé pour segmenter et retranscrire les extraits vidéos*



### b. Annotation des extraits audiovisuels selon l'émotion

Parmi la centaine d'heures d'enregistrements audiovisuels divisés en propositions, 2 557 extraits ont été sélectionnés pour leur caractère émotionnel. Ces extraits avaient une durée moyenne de 7.24 secondes ( $SD = 5.02$ ) et totalisaient environ une dizaine d'heures d'enregistrement. L'objectif concernant ces extraits était de les faire évaluer par des annotateurs selon une approche catégorielle et une approche dimensionnelle de l'émotion.

## **i. Sélection des dimensions de l'émotion à annoter**

### **1. Approche catégorielle : sélection des labels**

Une première approche proposée aux annotateurs, pour évaluer l'émotion exprimée par les participants dans les extraits, était une approche par labels. Pour ce faire, il était question de sélectionner un ensemble de labels émotionnels pertinents dans le contexte d'interaction de l'ECI assisté par l'avatar. Ce contexte concernait deux aspects distincts selon notre point de vue : la réalisation d'exercices et les interactions avec l'avatar. Aussi, nous nous sommes intéressés à la littérature portant sur les émotions dans le cadre de l'apprentissage scolaire. Ce cadre théorique nous semblait être le plus proche du contexte de ce projet puisque dans les deux situations, des individus se trouvent dans un contexte d'apprentissage assisté par un superviseur (l'avatar dans le cadre de l'ECI, un professeur dans le cadre scolaire). Ces émotions, dites d'accomplissement, peuvent être divisées en plusieurs classes (Pekrun, 2006, 2017; Pekrun & Stephens, 2010, 2012) : a) les émotions d'activité ; b) les émotions d'issue rétrospective ; c) les émotions d'issue prospective ; d) les émotions épistémiques ; et e) les émotions sociales.

Ces émotions d'accomplissement font référence aux émotions en lien direct avec l'activité. Cela comprend les émotions expérimentées durant l'accomplissement d'activité ou bien à l'issue de l'activité, qu'elle soit rétrospective ou prospective.

- a) Les émotions d'activité émergent selon l'évaluation de l'activité et non de l'issue. En d'autres termes, ces émotions ne sont pas impactées par le fait qu'un résultat, qu'il soit passé ou imaginé, soit un échec ou une réussite. Ces émotions se concentrent majoritairement sur l'évaluation de la nouveauté et du potentiel de maîtrise relatif à l'activité en cours d'accomplissement. Par exemple, une activité évaluée comme suffisamment inhabituelle et maîtrisable pourra générer de l'intérêt (Silvia, 2010).
- b) Les émotions d'issue rétrospective émergent à la suite de l'obtention d'un résultat, qu'il soit un échec ou une réussite. Celles-ci vont être différenciées en partie sur la base de

l'évaluation de l'attribution causale du résultat (e.g., se sentir responsable d'un échec ou bien l'attribuer à quelqu'un ou quelque chose d'autre). Ainsi, si un individu s'attribue la cause d'une réussite, il pourra ressentir de la fierté (Pekrun, 2006).

- c) Les émotions d'issue prospective émergent à la suite de l'évaluation du potentiel de maîtrise face à l'accomplissement d'une activité future. Plus précisément, les émotions d'issue prospective sont différenciées selon que l'issue de l'activité est évaluée comme donnant lieu à un échec ou une réussite. Ces émotions se distinguent également en fonction de l'évaluation de l'impact des moyens possibles pour éviter l'échec ou conduire à la réussite ainsi que du niveau de probabilité qui est associé à chacun de ces moyens. Par exemple, un individu qui infère le résultat d'un futur accomplissement comme étant une réussite à l'aide de ses propres moyens pourra expérimenter de la joie anticipée (Pekrun, 2006).
- d) Les émotions épistémiques font référence aux émotions en lien direct avec les dimensions de l'apprentissage et des activités cognitives qui génèrent des connaissances. Ces émotions se distinguent des émotions d'accomplissement car elles sont générées par une forme de métacognition sur la perception que l'apprenant a de l'accomplissement. Par exemple, un même label émotionnel peut être associé à une émotion d'accomplissement ou une émotion épistémique selon l'objet déclencheur de l'émotion. Prenons le cas de la frustration expérimentée par un apprenant qui se heurte à une résolution de problème. Cette frustration sera considérée comme une émotion épistémique si elle émerge du fait que l'apprenant ne s'évalue pas en mesure de mettre en œuvre le raisonnement nécessaire à la résolution du problème ou bien à une émotion d'accomplissement s'il évalue ce blocage comme un échec qu'il n'arrive pas à maîtriser.

- e) Les émotions sociales font référence aux émotions dans lesquelles l'objet déclencheur implique d'autres individus. Par exemple, le fait d'expérimenter de la fierté en étant félicité à la suite d'une réussite.

La littérature sur les émotions liées à l'apprentissage mentionne également une classe d'émotions dites « du thème ». Ces émotions font référence aux émotions générées directement par le matériel pédagogique. Pekrun et Stephens (2012) font par exemple référence à la frustration expérimentée par un enfant américain après avoir appris que Pluton était considérée comme une planète naine (Broughton et al., 2013). Cependant, il ne nous semblait pas pertinent de nous concentrer sur ce type d'émotions dans le cadre de l'ECI.

Sur la base de ces différentes classes d'émotions qui peuvent être impliquées dans un contexte d'apprentissage, nous avons sélectionné 23 labels d'émotions qui nous sont apparus les plus probables d'être expérimentés par les participants dans le cadre de l'ECI assisté par l'avatar. Ces 23 labels sont reportés dans le Tableau 4 accompagnés des types d'émotions auxquels ils peuvent être associés, ainsi que de la définition qui a été fournie aux annotateurs les concernant.

#### **Tableau 4**

*Les différents labels sélectionnés pour la phase d'annotation et proposés aux annotateurs, avec les types d'émotion qui leur sont associés et leur définition.*

<b>Label</b>	<b>Type d'émotion</b>	<b>Définition proposée aux annotateurs</b>
Anxieux	Épistémique/D'issue prospective	Qui exprime de l'inquiétude et une forte tension nerveuse.
Coupable	D'issue rétrospective	Qui ressent un sentiment de faute, que celle-ci soit réelle ou imaginaire
Déçu	D'issue rétrospective	Sentiment résultant du fait qu'un évènement n'a pas généré les conséquences attendues pour lesquelles on se réjouissait
Désespéré	D'issue prospective	Sentiment résultant de la perte totale de tout espoir qui conduit à ne plus savoir comment faire évoluer les choses positivement
En colère	Social/D'issue rétrospective/D'activité	État traduisant un vif mécontentement, suite à un évènement désagréable et associé à une

		tendance à agir de façon brutale, vive ou agressive
Ennuyé	D'activité	État de lassitude morale, produit par un manque d'intérêt ("qui s'ennuie...")
Frustré	D'activité/Épistémique	Sentiment résultant de la privation d'un bien ou d'un avantage attendu ou encore de l'impossibilité à montrer un atout personnel
Gêné	D'issue prospective	Sentiment ressenti lorsqu'on se trouve dans une situation embarrassante ou difficile qui génère une contrainte
Honteux	D'issue rétrospective	Sentiment pénible d'être perçu (réellement ou potentiellement) comme inférieur ou indigne
Méprisant	Social	Sentiment par lequel on considère quelqu'un comme indigne de toute estime
Triste	D'issue rétrospective	État affectif pénible d'insatisfaction qui conduit au repli sur soi et supprime l'envie d'agir
Vexé	Social/D'issue rétrospective	Sentiment ressenti lorsqu'on est blessé(e) dans son amour propre
Confiant	D'issue prospective/D'activité	Sentiment d'assurance par rapport à la réalisation d'un événement ou l'atteinte d'un but ou d'un état à venir
Curieux	D'activité/Épistémique	Qui a le désir de connaître, d'en savoir plus ou d'apprendre
Détendu	D'activité/D'issue rétrospective/prospective	Qui est dans un état de calme, de bien-être, qui est relaxé(e), toute tension étant relâchée
Étonné	D'activité/Épistémique	État de surprise créé par la présence d'un événement inattendu
Excité	D'activité	État caractérisé par une activité mentale vive, allant dans le sens d'un fort intérêt pour quelque chose
Fier	D'issue rétrospective	Qui ressent un sentiment de satisfaction lié à sa réussite
Impatient	D'activité	Qui a envie que quelque chose se passe ou qui a envie d'agir alors qu'il faut attendre
Intéressé	D'activité/Épistémique	Qui est prêt à porter son attention sur quelque chose qui le ou la passionne ou captive
Joyeux	D'issues rétrospective et prospective	Qui éprouve du plaisir
Plein d'espoir	D'issue prospective	Sentiment résultant de la possibilité d'atteindre un objectif désiré ou que quelque chose de désirée se passe
Satisfait	D'issue rétrospective	Sentiment résultant de l'atteinte de l'objectif fixé ou du fait que quelque chose d'agréable qui était attendue soit arrivée

## 2. Approche dimensionnelle : critères d'évaluation des théories de l'appraisal

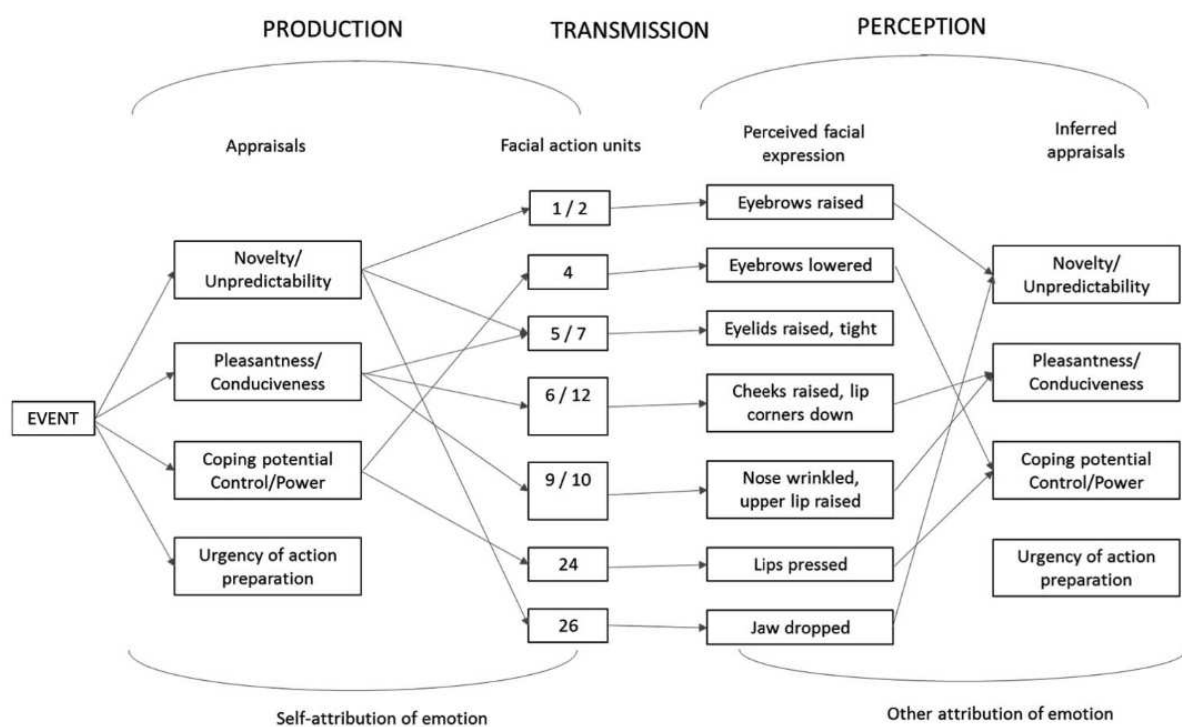
Comme mentionné dans l'introduction de cette thèse, les théories de l'appraisal proposent que les émotions sont définies comme des épisodes durant lesquels les sous-systèmes de l'organisme se synchronisent et/ou se couplent afin de produire une réaction adaptée à un événement évalué comme pertinent pour les buts et besoins des individus (Grandjean & Scherer, 2014). Dans ce cadre, les émotions pourraient être différenciées sur la base des patterns du processus d'évaluation cognitive réalisé par les individus, concernant l'événement déclencheur (Lazarus, 1991 ; Moors, 2013 ; Moors et al., 2013 ; Smith & Kirby, 2009). En effet, l'attribution de labels émotionnels permettant de qualifier les épisodes émergerait sur la base de la partie consciente du processus d'évaluation cognitive et de la synchronisation des sous-systèmes de l'organisme (Grandjean et al., 2008). Lorsqu'il s'agit de reconnaître une émotion produite par un autre individu, les théoriciens de l'appraisal proposent que les indices comportementaux exprimés seraient utilisés afin d'inférer sur le processus d'évaluation cognitive qui a pu générer ces indices expressifs (Scherer, 2013). En d'autres termes, si les êtres humains sont capables de reconnaître et distinguer les émotions, c'est parce qu'ils imaginent le processus d'évaluation cognitive qui a pu générer l'épisode émotionnel à partir des expressions motrices qu'ils observent. Cela peut être illustré par le Modèle Tripartite d'Expression et de Perception de l'Émotion (TEPE) décrit dans la Figure 6 (Scherer, 2013; Scherer et al., 2018). Bien que le modèle TEPE se focalise sur l'inférence du processus d'évaluation cognitive à travers l'expression faciale, il est entendu qu'un tel phénomène d'inférence pourrait être engagé dans tout type d'expression motrice (e.g., prosodie ou posture). Prenons un exemple décrivant la reconnaissance de l'émotion d'un émetteur par un observateur sur la base de l'expression faciale. Dans le cadre du modèle TEPE, l'émetteur de l'émotion va évaluer un événement selon les différents critères d'évaluation cognitive du MPC (i.e., pertinence, implication, potentiel de maîtrise, signification normative). Selon l'évaluation de ces critères, une activation

synchronisée des sous-systèmes de l'organisme sera générée chez l'émetteur (dont des expressions faciales). Ce phénomène représente l'épisode émotionnel dont la partie accessible à la conscience de l'émetteur lui permettra de qualifier l'émotion qu'il ressent. L'observateur lui, inférerait les différentes évaluations qu'a pu faire l'émetteur sur la base des unités d'action de l'expression faciale que l'émetteur a générées. Ainsi, l'observateur attribuerait l'émotion ressentie par l'émetteur sur la base de ces inférences.

### Figure 6

*Le modèle tripartite d'expression et de perception des émotions adapté de (Scherer et al., 2018)*

*« Appraisal-driven facial actions as building blocks for emotion inference »*



Dans le cadre du projet THERADIA, nous nous sommes basés sur les hypothèses du modèle TEPE afin de proposer aux annotateurs d'évaluer les émotions selon différents critères d'évaluation qu'ils attribueraient aux participants lors du visionnage des extraits audiovisuels. Les critères que nous avons sélectionnés sont ceux utilisés par les théoriciens de l'appraisal dans des études similaires (Scherer et al., 2013, 2018, 2019). Parmi eux, se trouvent la



nouveauté, l'agrément intrinsèque (i.e., la pertinence intrinsèque), la conductivité aux buts et le potentiel de maîtrise. Nous avons également demandé aux participants d'évaluer le niveau d'arousal (ou d'éveil) des participants, afin de qualifier l'intensité de l'émotion qu'ils observaient. Les différents critères étaient définis comme suit aux participants :

- Éveil : à quel point le participant semble endormi ou éveillé ?
  - Évaluer si le participant est plutôt endormi ou éveillé durant l'extrait vidéo.
    - Endormi : Le participant semble endormi, quelle qu'en soit la raison. Par exemple, il peut être fatigué ou il peut trouver l'exercice trop facile ou ennuyeux.
    - Éveillé : Le participant semble très éveillé, quelle qu'en soit la raison. Par exemple, il vient d'apprendre une excellente nouvelle
- Nouveauté : à quel point le participant ressent ce qui se passe comme prédictible ou inattendu ?
  - Évaluer si ce que ressent le participant lors de l'extrait vidéo est dû à un évènement considéré par lui-même comme plus ou moins prédictible ou inattendu. Par exemple, les choses se passent comme le participant s'y attendait, quelle qu'en soit la raison : ses performances à l'exercice sont conformes à ses prévisions, ou les évènements qui surviennent sont en total accord avec ses attentes. Ou, au contraire, les choses se passent de façon inattendue, quelle qu'en soit la raison : ses performances à l'exercice sont différentes de ses prévisions, ou les évènements qui surviennent sont totalement inattendus.
    - Prédictible : les choses se passent comme le participant s'y attendait, quelle qu'en soit la raison. Par exemple, ses performances à l'exercice sont conformes à ses prévisions ou les événements qui surviennent sont en total accord avec ses attentes.

- Inattendu : les choses semblent se passer de façon inattendue, quelle qu'en soit la raison. Par exemple, ses performances à l'exercice sont différentes de ses prévisions, ou les événements qui surviennent sont totalement inattendus.
- Agrément intrinsèque : À quel point le participant exprime du plaisir ou du déplaisir ?
  - Évaluer si le participant exprime du déplaisir ou du plaisir lors de l'extrait de la vidéo. La difficulté de cette évaluation est de ne pas la confondre avec la conductivité au but, qui évalue si le participant semble vouloir ce qui se passe ou non en fonction des buts qu'il poursuit. Par exemple, le participant peut ressentir du déplaisir à effectuer un exercice dont il connaît l'importance pour s'améliorer.
    - Déplaisir : le participant exprime du déplaisir.
    - Plaisir : le participant exprime du plaisir.
- Conductivité aux buts : à quel point ce qui se passe semble être ce que veut le participant ?
  - Évaluer si le participant semble vouloir ce qui se passe ou non. La difficulté de cette évaluation est de ne pas la confondre avec l'agrément intrinsèque, qui évalue si le participant exprime du plaisir ou du déplaisir. En effet, le participant peut vouloir une situation sans pour autant ressentir du plaisir.
- Potentiel de maîtrise : à quel point le participant semble pouvoir faire face à ce qui se passe ?
  - Pas du tout : le participant a le sentiment que rien ne peut être fait. Par exemple, il apprend avoir raté un exercice et se sent incapable de faire mieux.
  - Tout à fait : Le participant a le sentiment de pouvoir surmonter les conséquences négatives de l'événement. Par exemple, en cas d'échec à un exercice, il sait qu'il pourra faire mieux la prochaine fois.

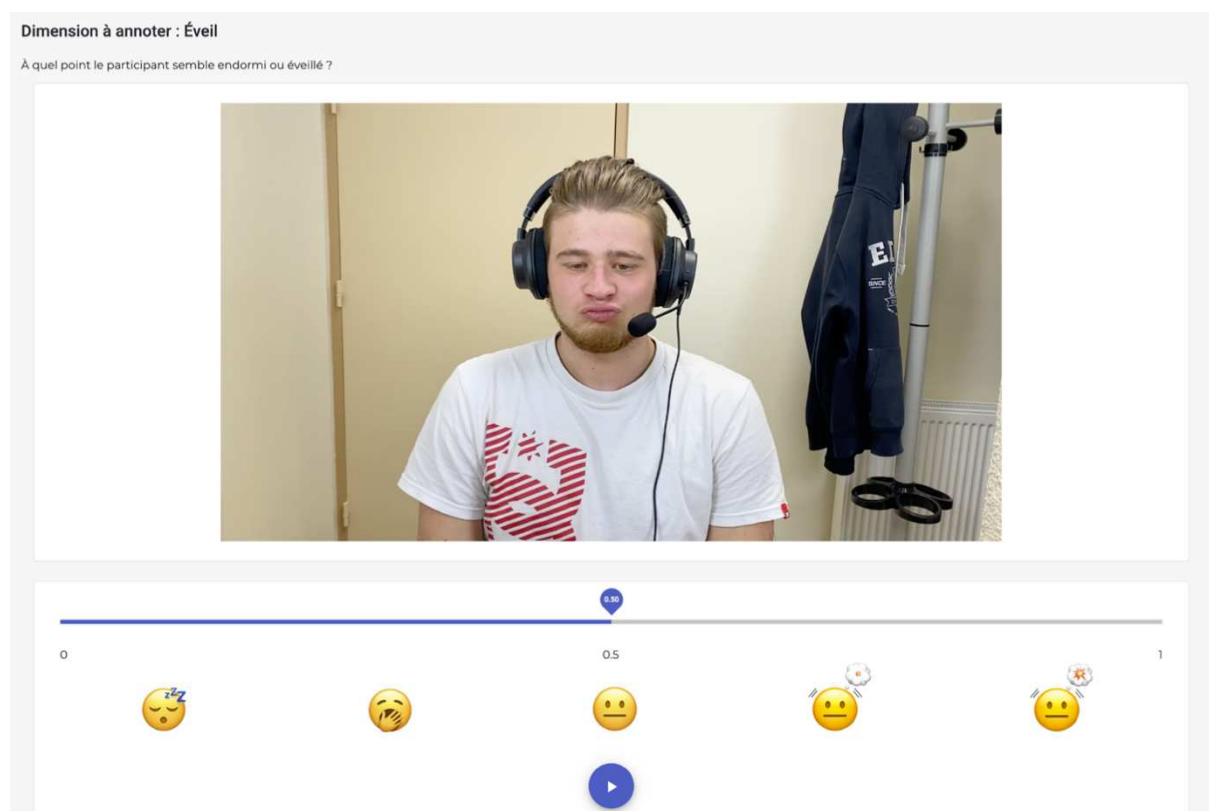
## ii. Annotations avec le logiciel Annot'

Six annotateurs ont été recrutés pour annoter les 2 557 extraits audiovisuels issus des séances des participants âgés, sans troubles cognitifs, réalisant l'ECI accompagnés de l'avatar. L'annotation s'est déroulée sur la plateforme informatique Annot' développée dans le cadre du projet THERADIA par le LIG. Chaque extrait vidéo était annoté de 3 manières différentes :

- Une annotation dimensionnelle continue dans le temps de chaque critère d'évaluation mentionné précédemment, sur un axe allant de 0 à 1 avec un pas de 0.01, comme illustré dans la Figure 7. Cette annotation était réalisée tout au long du visionnage de l'extrait vidéo. À l'issue de l'annotation du critère, un feedback de l'évaluation réalisée apparaissait à l'écran, comme décrit dans la Figure 8. L'annotateur pouvait soit valider soit recommencer l'annotation. Chaque extrait audiovisuel était visionné une fois par critère, pour un total de 5 visionnages pour ce type d'annotation.
- Une fois la courbe d'annotation temporelle du critère validée, il était demandé aux annotateurs de fournir une valeur moyenne du critère d'évaluation représentant, pour lui, l'ensemble de l'extrait audiovisuel (voir Figure 9).
- À l'issue de l'annotation dimensionnelle de tous les critères, l'annotateur visionnait à nouveau l'extrait audiovisuel et devait décider du ou des labels émotionnels qui pouva(en)t le représenter. Plusieurs labels pouvaient être sélectionnés pour le même extrait. Chaque label était annoté sur une échelle allant de 0 à 100, comme illustré dans la Figure 10.

**Figure 7**

*Visuel du logiciel Annot' lors de l'évaluation continue dans le temps du critère d'éveil*

**Figure 8**

*Visuel de la courbe d'annotation continue dans le temps d'un critère d'évaluation dans le logiciel Annot'*

**Courbe d'annotation temporelle**

Recommencer

Valider

## Figure 9

*Visuel de l'annotation moyenne d'un critère d'évaluation dans le logiciel Annot'*

Resumé 0 / 3

Dimension à résumer : Éveil

Pour résumer cette séquence, la personne semble être :

0 0.5 1

🤤 🤔 😐 🌟 😟

[soumettre](#)

## Figure 10

*Visuel de l'annotation des labels dans le logiciel Annot'*

0:00 / 0:03

100

0

[soumettre](#)

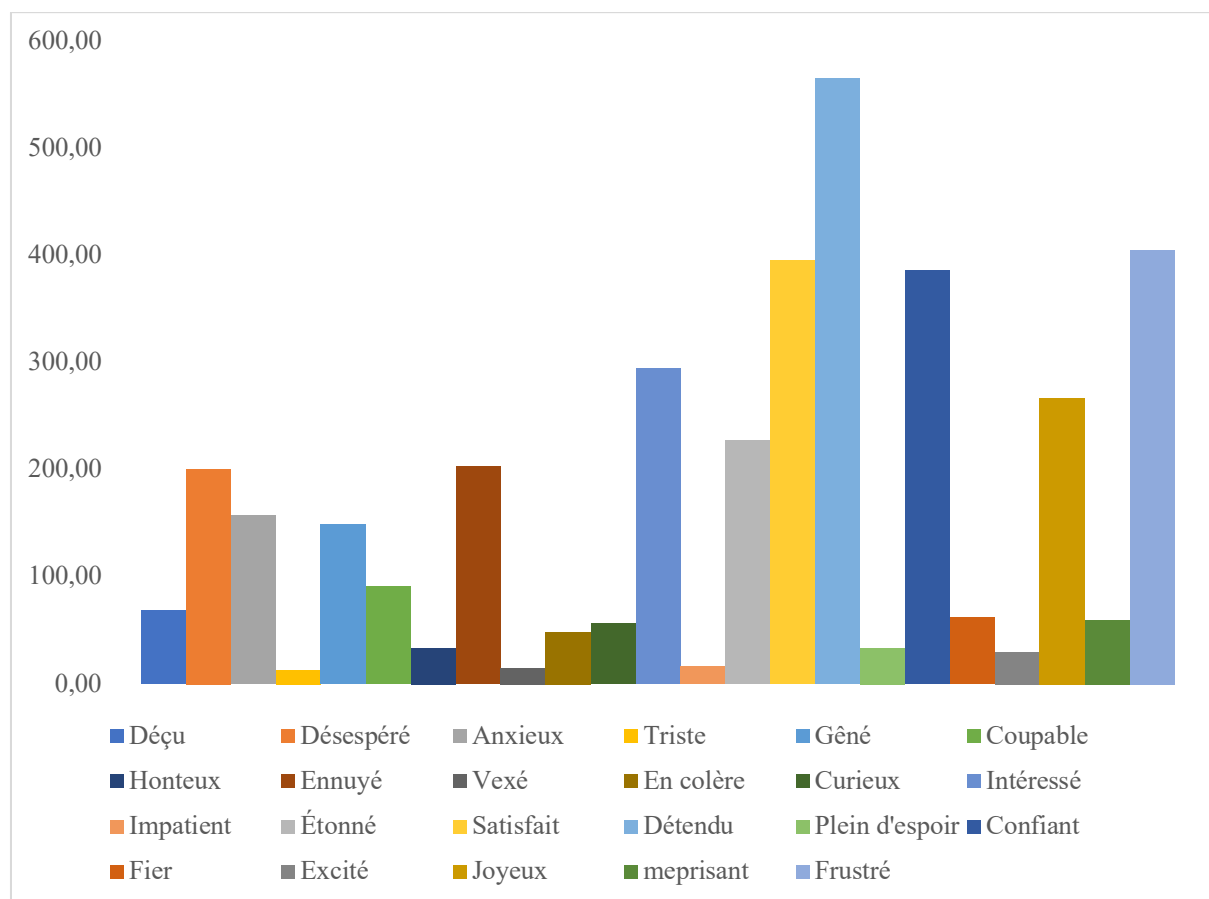
Coupable		Plein d'espoir
Ennuyé		Fier
Vexé		<b>Excité 60</b>
En colère		Joyeux
Anxieux		Etonné
Honteux		Confiant
Déçu		Intéressé
Triste		Satisfait
Gêné		Détendu
Désespéré		Impatient
Frustré		Curieux
Agité		

### **iii. Résultats préliminaires**

La campagne d'annotation est désormais terminée (avril 2023). Nous travaillons actuellement en collaboration avec le LIG afin de construire le corpus de données qui servira à entraîner les modèles de détection d'émotion. Pour l'heure, seules quelques analyses préliminaires ont pu être réalisées sur les annotations dimensionnelles (discrètes et non continues dans le temps) et sur les annotations catégorielles par label. Ces analyses sont reportées ci-dessous. La figure 11 représente le ratio d'occurrence des annotations de label sur l'ensemble des extraits audiovisuels. La figure 12 illustre un modèle graphique Gaussien entre les annotations discrètes de labels et de dimensions de critère d'évaluation sur l'ensemble des extraits audiovisuels.

**Figure 11**

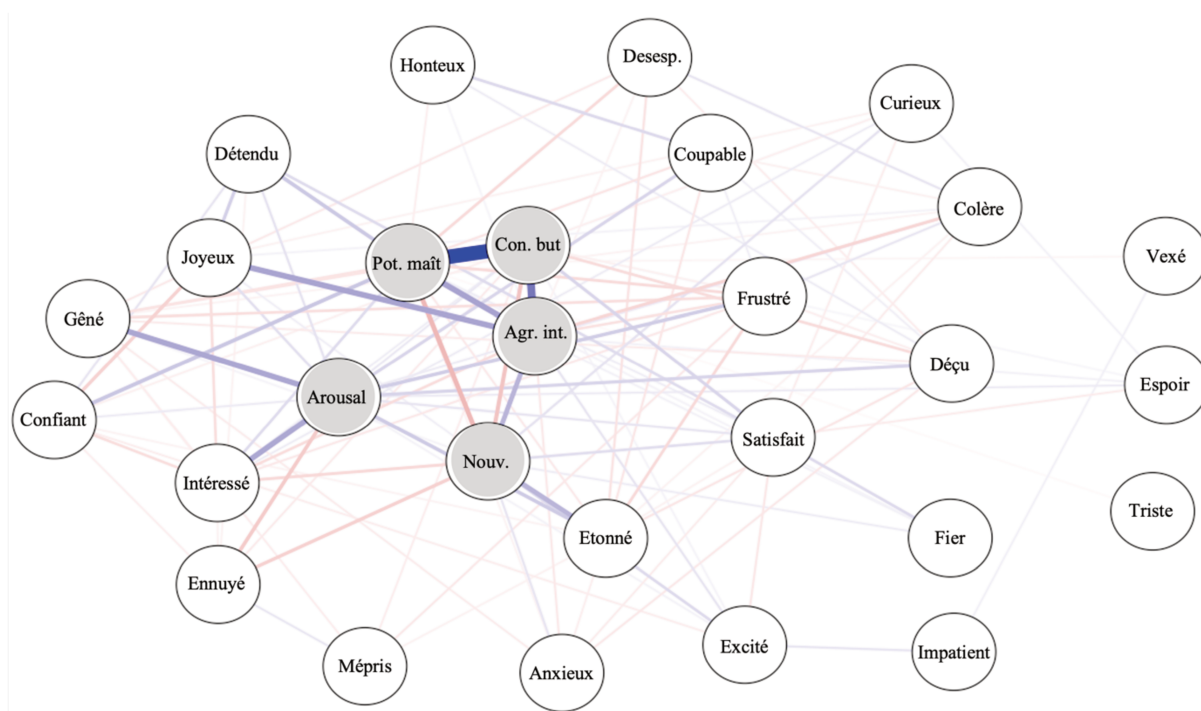
*Représentation graphique de l'occurrence de labels dans l'annotation de la totalité des extraits audiovisuels*



*Note. Chaque occurrence correspond à la présence de l'annotation d'un label quelle que soit sa valeur. Afin de tenir compte de l'accord inter-annotateur chaque occurrence a été divisée par le nombre d'annotateurs*

**Figure 12**

*Modèle Graphique Gaussien des annotations de label et de dimensions discrètes évaluées par les annotateurs*



*Note. Modèle réalisé à l'aide du package « qgraph » version 1.5 (Epskamp et al., 2012). Ce modèle a été créé à l'aide de la méthode lasso et de l'Extended Bayesian Information Criterium comme paramètre d'ajustement selon les recommandations de Epskamp et al. (2012). Les connexions rouges et bleues reflètent les corrélations négatives et positives, respectivement. L'épaisseur de la connexion représente la force de la corrélation. Les critères d'évaluation sont représentés. « Pot. Maît » réfère au potentiel de maîtrise ; « Nouv » réfère à la nouveauté ; « Agr. int. » réfère à l'agrément intrinsèque ; « Con. but » réfère à la conductivité aux buts.*

### **c. État d'avancement actuel du projet et suite**

Comme mentionné précédemment, les différents partenaires du consortium travaillent actuellement sur les données recueillies lors de la phase expérimentale afin de construire une version autonome de l'assistant virtuel. D'autre part, les questionnaires proposés aux participants lors des séances ont déjà donné lieu à la rédaction d'un article actuellement soumis



pour publication (Zsoldos et al., Submitted). Cet article s'intéresse à la manière dont les participants apprécient et pensent utile la présence d'un assistant virtuel lors de l'ECI en fonction de divers facteurs individuels (e.g., âge, fonctionnement psycho-affectif, personnalité, motivations personnelles, compétence cognitive). Cet article a fourni des résultats encourageants pour les objectifs du projet THERADIA, en montrant par exemple que l'ensemble des participants a apprécié et jugé utile d'être accompagnés par un assistant virtuel lors de l'ECI.

Pour la suite du projet, une étude clinique devrait être mise en place dans les prochains mois afin de tester l'impact de l'ECI sur des patients lorsque celui-ci est complété par l'assistant virtuel (autonome cette fois-ci) développé dans le cadre du projet.

## Chapitre 9 - Discussion générale et conclusion

### 1. Discussion générale

La présente thèse s'inscrit dans le cadre des théories de l'appraisal. Ces théories visent à définir les interactions des individus avec leur environnement sur la base d'un processus d'évaluation cognitive de la signification des stimuli selon plusieurs critères (Scherer, 1984, 2009). Ce processus d'évaluation cognitive est défini comme le chef d'orchestre de la régulation interne des sous-systèmes de l'organisme des individus (Grandjean & Scherer, 2014). En d'autres termes, selon la signification de l'environnement émergeant du processus d'évaluation, des activations synchronisées des sous-systèmes de l'organisme seraient déclenchées afin de guider les individus vers les comportements qui faciliteront l'atteinte de leurs buts et besoins. Dans cette thèse, nous nous sommes intéressés à la manière dont le processus d'évaluation cognitive impacte l'attention. Dans la mesure où les ressources attentionnelles sont limitées (Kahneman, 1973), le processus d'évaluation cognitive aurait pour objectif d'allouer ces ressources de manière adaptée. Ce phénomène, défini comme attention émotionnelle (Vuilleumier, 2005), serait réalisé sur la base de l'évaluation du critère de pertinence (Brosch et al., 2008; Sander et al., 2011). La pertinence des stimuli serait établie selon trois sous-critères : leur nouveauté, définie par leur niveau de soudaineté, de familiarité et de prédictibilité ; leur pertinence intrinsèque, définie par leur nature plutôt plaisante ou déplaisante de manière générale, indépendamment de l'état de l'organisme ; et leur pertinence aux buts, définie par leur niveau de facilitation ou d'obstruction pour la réalisation des buts et besoins momentanément importants pour les individus. Si de nombreuses études ont apporté des preuves en faveur d'un impact attentionnel de la nouveauté (Sussman et al., 2003; Yantis & Jonides, 1984), de la pertinence intrinsèque (Brosch & Sharma, 2005 ; Fox et al., 2007 ; LoBue & DeLoache, 2008; Öhman, Flykt, et al., 2001; Öhman, Lundqvist, et al., 2001 ; Pool et al.,

2016) et de la pertinence aux buts (Mazzietti et al., 2014 ; Müller et al., 2016 ; Müller & Wentura, 2023 ; Sun et al., 2017 ; Vogt et al., 2011 ; Vogt et al., 2013 ; Vogt et al., 2017), peu d'entre elles se sont intéressées à leurs impacts combinés sur l'attention. Les expériences menées au cours de cette thèse ont eu pour ambition de répondre à cette question, en explorant les effets combinés de la pertinence intrinsèque et de la pertinence aux buts sur l'attention. Une question complémentaire concernait la manière de mesurer ces effets.

### **a. Effets multiplicatifs des pertinences**

Dans une première étude (reportée dans le Chapitre 2 du manuscrit), nous nous sommes intéressés aux effets combinés des pertinences sur l'attention dans le cas où la pertinence intrinsèque est évaluée comme plaisante et la pertinence aux buts est évaluée comme facilitatrice. Selon l'hypothèse défendue par les théories de l'appraisal, la pertinence des stimuli est considérée comme un marqueur de tendance à l'action. En effet, si un stimulus est évalué comme plaisant et facilitateur pour l'accomplissement d'un but, celui-ci devrait alors générer une tendance à l'approche. Aussi, dans cette étude, nous nous sommes également intéressés à la relation entre la pertinence intrinsèque, la pertinence aux buts et la tendance à l'action. Ces questions ont été étudiées à l'aide d'une tâche de dot-probe dans laquelle des aliments et des boissons étaient mis en compétition après que leur niveau de pertinence intrinsèque, de pertinence aux buts et de tendance à l'approche ait été évalué à l'aide d'un questionnaire.

Au sujet de la relation entre pertinence et tendance à l'action, nos résultats ont mis en évidence que le niveau de tendance à l'approche rapporté au sujet des stimuli était mieux prédit en tenant compte de leur niveau de pertinence intrinsèque et de pertinence aux buts. Ce résultat montre l'importance de distinguer la pertinence intrinsèque de la pertinence aux buts, ce qui n'est pas systématiquement fait dans la littérature (e.g., Feldman Barrett & Russell, 1998 ; Kron et al., 2015). En effet, nombreux sont les auteurs qui résument les concepts d'agrément des stimuli et de leur nature facilitatrice pour l'accomplissement de buts et besoins à la notion

unidimensionnelle de valence. Or, la notion de valence pourrait être rapprochée de celle de la tendance à l'action (voir Lewin, 1938/1951 ; Scherer, 2013, pour une discussion à ce sujet). Le fait qu'une valence positive soit attribuée à un stimulus signifie en effet qu'il est associé à une tendance à l'approche. Or, bien que ce ne soit pas le cas des stimuli que nous avons manipulés dans cette étude, il arrive que certains d'entre eux puissent être à la fois déplaisants et facilitateurs pour un but. On pourrait par exemple citer le cas d'un individu qui consomme des choux de Bruxelles pour satisfaire sa faim tandis qu'il ne les apprécie pas. Dans ce cas de figure, les choux de Bruxelles, en tant que stimuli, seront probablement associés simultanément à une tendance à l'approche par leur caractère facilitateur et à une tendance à l'évitement par leur caractère déplaisant. Aussi, définir leur valence selon un continuum unidimensionnel semble problématique car cela ne permet pas de rendre compte du fait que ces stimuli soient à la fois positifs et négatifs. En réponse à cela, les théoriciens de l'appraisal militent pour définir la valence comme un concept à plusieurs dimensions représenté par les résultats des différents critères du processus d'évaluation cognitive, dont ceux de la pertinence intrinsèque et de la pertinence aux buts (Scherer, 2013; Scherer & Moors, 2019; Shuman et al., 2013). Nos résultats vont dans ce sens en montrant que la tendance à l'approche vers les stimuli serait générée par leurs niveaux de pertinence intrinsèque et de pertinence aux buts.

En ce qui concerne la relation entre pertinence et attention, les résultats de cette étude ont mis en évidence que la pertinence intrinsèque et la pertinence aux buts des stimuli auraient des effets multiplicatifs sur l'attention. Premièrement, ces résultats sont en accord avec les prédictions des théories de l'appraisal selon lesquelles à la fois la pertinence intrinsèque et la pertinence aux buts des stimuli capturent l'attention (Grandjean & Scherer, 2014 ; Scherer, 2009). Deuxièmement, ils amènent à développer une idée intéressante : le niveau de capture attentionnelle par les stimuli est sensible au niveau d'alignement de la pertinence intrinsèque et de la pertinence aux buts. Prenons le cas où un individu perçoit trois stimuli simultanément :

un premier évalué uniquement comme fortement plaisant ; un second évalué comme moyennement plaisant et pouvant moyennement faciliter la réalisation d'un but ; et un troisième évalué uniquement comme pouvant fortement faciliter la réalisation d'un but. Considérer que la pertinence intrinsèque et la pertinence aux buts aient des effets multiplicatifs sur l'attention laisse à penser que le second stimulus capturera plus fortement l'attention, car les résultats de la pertinence intrinsèque et de pertinence aux buts sont alignés (i.e., proches). Dans la mesure où la pertinence intrinsèque et la pertinence aux buts sont des évaluations, cela signifie qu'elles peuvent être sujettes à l'erreur. Dans le cas où les pertinences sont alignées, on peut imaginer que la confiance que le cerveau accorde aux pertinences est plus forte, car plusieurs informations convergent vers le même résultat. A contrario, lorsque le cerveau ne possède qu'une information de pertinence intrinsèque ou de pertinence aux buts, ou alors deux informations de pertinences non-alignées, la qualité de l'évaluation est probablement moins certaine. L'idée que les pertinences aient des effets multiplicatifs sur l'attention afin de favoriser les évaluations les plus certaines est également en accord avec le principe d'énergie libre selon lequel le cerveau viserait à minimiser l'incertitude et la surprise (Friston, 2005, 2010).

### **b. Effets combinés des pertinences dans le cas d'antagoniste et leur temporalité**

Dans une seconde étude (reportée dans le Chapitre 3 du manuscrit), nous avons tenté de comparer l'impact attentionnel des effets combinés de la pertinence intrinsèque et de la pertinence aux buts selon que la combinaison des pertinences était agoniste ou antagoniste. La notion de combinaison agoniste signifie que les pertinences génèrent le même type de tendance à l'action. Dans le cas de notre étude, il s'agissait de l'approche, lorsque la pertinence intrinsèque est évaluée comme plaisante et la pertinence aux buts est évaluée comme facilitatrice pour un but. La notion de combinaison antagoniste signifie que les pertinences génèrent des tendances à l'action opposées. Dans le cas de notre étude, la tendance à l'évitement

était générée par une pertinence intrinsèque déplaisante et la tendance à l'approche était générée par une pertinence aux buts facilitatrice. Pour ce faire, des carrés de différentes couleurs étaient mis en compétition dans une tâche de dot-probe. La pertinence aux buts était manipulée par le fait qu'un carré pouvait permettre de gagner des points dans une tâche d'induction, proposée parallèlement à la tâche de dot-probe. La pertinence intrinsèque était manipulée par un carré qui générait un bruit plaisant pour un groupe de participants et déplaisant pour un second groupe. De plus, cette étude avait pour objectif de tester l'hypothèse selon laquelle la pertinence intrinsèque a un impact plus fort sur l'attention à courte temporalité, alors que la pertinence aux buts a un impact plus fort sur l'attention à temporalité longue. Pour ce faire, deux temps de présentation des stimuli ont été utilisés dans la tâche de dot-probe : un court et un long. Avant, au milieu ainsi qu'à la fin de la tâche, les participants ont évalué les différents carrés selon leur niveau d'agrément, de désagrément, de facilitation au but, de tendance à l'approche et de tendance à l'évitement. Cela nous a également permis d'étendre les résultats de la première étude concernant la relation entre tendance à l'action et pertinence.

Au sujet de la relation entre tendance à l'action et pertinence, les résultats ont montré, d'une part, que la tendance à l'approche rapportée à propos des stimuli était mieux prédite en tenant compte à la fois de leur niveau d'agrément, de désagrément et de facilitation pour l'atteinte du but. Plus précisément, des relations positives ont été trouvées entre les niveaux d'agrément et de tendance à l'approche et entre les niveaux de facilitation au but et de tendance à l'approche. Au contraire, une relation négative a été mise en évidence entre la tendance à l'approche et le niveau de désagrément. D'autre part, les résultats ont montré que la tendance à l'évitement à des stimuli était également mieux prédite en tenant compte de leur niveau d'agrément, de désagrément et de facilitation pour l'atteinte du but. Cette fois-ci, des relations opposées ont été observées, de sorte que la tendance à l'évitement était reliée négativement au niveau d'agrément et de facilitation pour le but alors qu'une relation positive a été trouvée entre la tendance à

l'évitement et le niveau de désagrément. Premièrement, ces résultats soutiennent l'hypothèse selon laquelle la valence doit être considérée comme multidimensionnelle (Scherer, 2013; Scherer & Moors, 2019; Shuman et al., 2013). Cette étude s'inscrit dans la thématique portant sur les stimuli générant à la fois des tendances à l'approche et à l'évitement (Aupperle et al., 2015 ; Bach et al., 2014 ; Ito & Lee, 2016). Comme mentionné précédemment, si les tendances à l'approche et à l'évitement génèrent des valences positives et négatives, respectivement, ces résultats contribuent à rendre douteuse la notion de valence unidimensionnelle. En effet, il n'est pas possible de placer des stimuli générant à la fois des tendances à l'approche et à l'évitement sur un seul axe, dans la mesure où ces stimuli sont à la fois positifs et négatifs. Deuxièmement, ces résultats nous en apprennent plus sur la relation entre tendance à l'action et pertinence en mettant en évidence que la pertinence peut avoir un effet inhibiteur sur la tendance à l'action. De fait, nos résultats montrent que la tendance à l'approche était réduite par la nature désagréable des stimuli, tandis que la tendance à l'évitement était réduite par la nature agréable ou facilitatrice pour le but des stimuli.

Concernant la relation entre pertinence et attention, les résultats de l'étude ont mis en évidence que les combinaisons agoniste et antagoniste des pertinences intrinsèques et des pertinences aux buts avaient des effets combinés sur l'attention, et ce, de manière équivalente. Il est à noter que le design de cette étude n'a pas été conçu pour qualifier la nature des effets combinés des pertinences sur l'attention comme cela a été fait dans la première étude, du fait de la complexité qui en découlait. Dans la mesure où trois informations devaient être prises en compte (niveau d'agrément, de désagrément et de facilitation aux buts), il nous a semblé complexe de calculer des scores de pertinence afin de tester différentes relations mathématiques (e.g., multiplicative). Premièrement, le fait que les pertinences aient des effets combinés sur l'attention dans le cas d'une combinaison agoniste est en accord avec les résultats de la première étude. En effet, une capture attentionnelle plus forte vers les stimuli dont les pertinences étaient

agonistes reflèterait une augmentation de leur niveau de tendance à l'approche dans la mesure où ceux-ci sont plaisants et peuvent faciliter la réalisation d'un but. Dans le cas de combinaison antagoniste, de notre point de vue, une augmentation de la capture attentionnelle vers ces stimuli reflète une grande incertitude gravitant autour d'eux, puisqu'ils doivent probablement être approchés et évités. Aussi, cette capture attentionnelle augmentée reflèterait le fait que ces stimuli nécessitent d'être évalués de manière plus approfondie afin de décider ce qu'il convient de faire à leur sujet. Ces résultats peuvent également être interprétés selon le principe d'énergie libre (Friston, 2005, 2010), dans le sens où le cerveau chercherait à minimiser l'incertitude concernant la signification des stimuli.

Du point de vue de la temporalité, nos résultats n'ont pas permis de mettre en évidence des impacts plus forts de la pertinence intrinsèque et de la pertinence aux buts à temporalité courte et longue, respectivement. Cette hypothèse était portée par la proposition selon laquelle l'évaluation de la pertinence aux buts opérerait dans une fenêtre de temps plus longue que celle de la pertinence intrinsèque (Aue et al., 2007; Grandjean & Scherer, 2008; Lanctôt & Hess, 2007; Scherer, 2013). Aussi, la capture attentionnelle vers les stimuli serait préférentiellement basée sur la pertinence intrinsèque, car le résultat de la pertinence aux buts serait encore incertain à une temporalité courte. Selon nous, l'absence de résultat à ce sujet dans notre étude pourrait s'expliquer par le fait que le temps de présentation défini comme court était trop long pour pouvoir observer ce type d'interaction. Nous avons choisi un temps court fixé à 100ms, pour que la réponse intervienne avant que le traitement de la pertinence aux buts soit terminé. Or, nous n'avons pas pris en compte le délai d'enregistrement du temps de réponse, survenant 100ms plus tard, après l'extinction de la cible à l'écran. Le temps de présentation, défini comme court, représentait en réalité un temps d'évaluation d'environ 200ms, temps trop long pour espérer observer une capture attentionnelle plus forte générée par la pertinence intrinsèque des stimuli. Aussi, pour mesurer un niveau de capture attentionnelle censé refléter la fenêtre



temporelle dans laquelle la pertinence intrinsèque impacte plus l'attention que la pertinence aux buts, le temps de présentation des stimuli aurait dû être plus court. Par exemple, nous aurions dû utiliser un temps de présentation de 30ms, comme cela a été fait dans des études préalables s'intéressant à l'impact des pertinences sur l'attention en fonction de la temporalité (Forrest et al., 2022).

### **c. Annulation des effets de la pertinence intrinsèque par la pertinence aux buts**

La troisième étude (reportée dans le Chapitre 4 du manuscrit) avait pour objectif de tester à nouveau l'impact attentionnel des combinaisons agoniste et antagoniste de la pertinence intrinsèque et de la pertinence aux buts lorsque les stimuli peuvent faciliter un but. De plus, cette étude visait également à tester l'impact attentionnel des combinaisons agoniste et antagoniste des pertinences dans le cas où la pertinence aux buts était modélisée par une obstruction inévitable. Pour ce faire, des stimuli intrinsèquement pertinents (i.e., plaisant : un corps de femme nu ; et déplaisant : une araignée) étaient mis en compétition avec des stimuli neutres (i.e., un champignon et une plante) dans une tâche de dot-probe. En parallèle, les participants réalisaient une tâche d'induction dans laquelle un stimulus était induit à une facilitation d'un but (i.e., gain de points en appuyant sur une touche lors de son apparition) et un autre stimulus était induit à une obstruction inévitable (i.e., perte de points automatique). Les deux autres stimuli présentés dans la tâche d'induction n'avaient aucun impact sur le score. Les participants étaient répartis par groupes dans lesquels les inductions étaient attribuées à différents stimuli. Dans deux groupes, les inductions étaient attribuées aux stimuli intrinsèquement pertinents (i.e., gain au stimulus plaisant et perte au stimulus déplaisant pour un groupe ; gain au stimulus déplaisant et perte au stimulus plaisant pour l'autre groupe). Dans deux autres groupes, les inductions étaient attribuées aux stimuli neutres (i.e., gain et perte attribués au champignon et à la plante pour un groupe ; inversement pour l'autre groupe). Pour

un dernier groupe, les inductions étaient attribuées à un autre set de stimuli, de sorte à pouvoir mesurer le niveau de capture attentionnelle généré par la pertinence intrinsèque des stimuli du set principal, sans induction. À l'issue des tâches, les participants ont évalué le niveau de pertinence intrinsèque des stimuli.

Concernant l'évaluation des stimuli, les résultats ont montré que les stimuli définis comme plaisants et déplaisants étaient effectivement évalués comme plus plaisants et déplaisants que les autres, respectivement. De plus, il a été mis en évidence que les stimuli étaient évalués comme plus plaisants par les participants des groupes dans lesquels ils étaient induits au gain. Au contraire, les stimuli étaient évalués comme plus déplaisants par les participants des groupes dans lesquels ils étaient induits à la perte. Ces résultats sont en accord avec l'hypothèse inspirée du modèle « learning – liking – wanting » (Berridge, 2009; Berridge et al., 2009; Castro & Berridge, 2014), selon laquelle la pertinence aux buts des stimuli dans des expériences passées amène à moduler leur niveau de pertinence intrinsèque. Plus précisément, le fait que les stimuli permettent de faciliter la réalisation d'un but ou d'un besoin renforcerait leur caractère agréable. Inversement, le fait que les stimuli obstruent la réalisation d'un but ou d'un besoin renforcerait leur caractère désagréable.

En ce qui concerne les hypothèses relatives aux impacts des combinaisons de pertinence intrinsèque et de pertinence aux buts sur l'attention, des résultats mitigés ont été observés. Conformément à ce qui était attendu, les stimuli intrinsèquement pertinents captaient plus fortement l'attention que les stimuli neutres lorsqu'aucun d'entre eux n'était sujet à une induction. De plus, nous avons observé que lorsque les stimuli étaient induits à la facilitation au but, leur niveau de capture attentionnelle était augmenté et ce, à la fois pour les stimuli intrinsèquement pertinents et les stimuli neutres. Quant à l'impact attentionnel des stimuli dans le cas de l'induction à une obstruction au but inévitable, des résultats surprenants ont été mis en évidence. En effet, dans une étude précédente (Müller & Wentura, 2023), les auteurs

n'avaient pas réussi à montrer une capture attentionnelle en direction de stimuli attribués à une obstruction inévitable. Aussi, nous ne nous attendions pas à observer de différence attentionnelle selon que les stimuli étaient induits à une obstruction au but inévitable ou n'étaient sujets à aucune induction. Dans notre étude, il est apparu que lorsque les stimuli étaient induits à une obstruction au but inévitable, tous les stimuli captaient l'attention de manière équivalente, de sorte que la capture attentionnelle attendue en direction des stimuli intrinsèquement pertinents était annulée. Cette annulation de la capture attentionnelle était également observée lorsque les stimuli neutres étaient induits à une obstruction au but inévitable. Pour expliquer ces résultats, nous avons supposé que l'obstruction inévitable du but n'était pas assez forte, car la perte de points était égale à la moitié du nombre de points total qui pouvait être gagné par les stimuli induits au gain. Selon notre point de vue, les stimuli induits à l'obstruction inévitable n'étaient finalement pas plus pertinents pour le but que ceux qui n'étaient attribués à aucune induction. Aussi, ces résultats pourraient refléter le fait que tous les stimuli qui étaient considérés comme non pertinents pour le but devenaient donc neutres. En d'autres termes, dans un contexte où gagner des points est très important, il est possible que tous les stimuli qui ne permettent pas de gagner des points soient inhibés du point de vue de l'attention, car les traiter ne permettait pas d'accomplir le but de gagner. Bien que ces résultats ne résonnent pas en faveur de notre hypothèse, il est intéressant de constater qu'une pertinence peut en « écraser » une autre, rendant son impact attentionnel caduc. Ce résultat peut également être interprété dans le cadre du principe d'énergie libre (Friston, 2005, 2010) : si une grande confiance est accordée à une évaluation selon laquelle un stimulus ne nécessite pas d'attention car il n'est pas pertinent pour un but, il est possible que des évaluations (e.g., la pertinence intrinsèque) associées à des niveaux de confiance plus faibles soient ignorées.

#### **d. Considérations méthodologiques de la mesure du biais attentionnel dans le cadre de l'anxiété**

La quatrième étude (reportée dans le Chapitre 5 du manuscrit) s'est intéressée à des aspects méthodologiques de la mesure du biais attentionnel à l'aide de la tâche de dot-probe dans le cadre de l'anxiété. S'il a été largement montré que le biais attentionnel vers la menace était exacerbé pour les individus souffrants d'anxiété (Bar-Haim et al., 2007; Waters et al., 2010), ces résultats n'ont pas été systématiquement observés au cours de la dernière décennie (Kruijt et al., 2019; McNally, 2019). L'objectif de cette étude était d'observer la meilleure manière de décrire le niveau de biais attentionnel des participants en direction de la menace selon : 1) une tâche de dot-probe classique ou une tâche dite « d'eye-dot-probe » (i.e., une tâche développée dans le cadre de cette thèse dont le côté d'apparition de la cible devait être indiqué via une saccade oculaire) ; 2) les différents calculs d'indices attentionnels proposés dans la littérature pour analyser les résultats ; 3) le fait de considérer seulement l'état (impactant la pertinence aux buts des stimuli menaçants) ou le trait (impactant la pertinence intrinsèque des stimuli menaçants) d'anxiété des participants ou un score représentant les effets combinés de l'état et du trait d'anxiété.

Concernant la question des tâches, les résultats ont montré que la tâche d'eye-dot-probe permettait d'identifier de manière plus fine le niveau de biais attentionnel des participants vers la menace que la tâche de dot-probe classique. Ces résultats pourraient s'expliquer par le fait que la mesure du biais attentionnel à travers le mouvement des yeux est plus fine que celle recueillie par pression d'une touche sur un clavier. En effet, les mouvements oculaires sont plus directement liés à la capture de l'attention. Par exemple, il a été montré que l'apparition d'un stimulus pertinent dans le champ visuel génère un déplacement du regard et un maintien de la position des yeux au niveau de sa région d'apparition (Pool et al., 2019).

Concernant la question des indices attentionnels, il a été montré que seul l'Indice de Biais Attentionnel (IBA) proposé par MacLeod et al. (1986) permettait d'expliquer le niveau de biais attentionnel en direction de la menace mesuré dans les deux tâches. En effet, il n'a pas été montré que les indices de Variabilité du Biais Attentionnel (VBA ; Iacoviello et al., 2014 ; Naim et al., 2015) et le "Trial Level-Bias Score" (TL-BS ; Zvielli et al., 2015) permettaient de décrire le niveau de biais attentionnel des participants au sein de ces tâches.

Enfin, concernant la question du lien entre type de mesure d'anxiété et biais attentionnel, il a été montré que le score représentant les effets combinés du trait et de l'état d'anxiété était celui qui permettait le mieux de décrire le biais attentionnel vers la menace des participants dans la tâche de dot-probe. Ce résultat est en accord avec ceux mis en évidence dans les études précédentes, si l'on considère que le niveau de trait d'anxiété impacte la pertinence intrinsèque des stimuli menaçants et que le niveau d'état d'anxiété impacte leur pertinence aux buts. Dans ce contexte, ce résultat s'inscrit dans le cadre des effets combinés de la pertinence intrinsèque et de la pertinence aux buts sur l'attention. Dans la tâche d'eye-dot-probe, le score qui représentait le mieux le niveau de biais attentionnel des participants vers la menace était celui ne tenant compte que du trait d'anxiété des participants. Si ce résultat semble en contradiction avec celui observé dans la tâche de dot-probe, une explication pourrait permettre de réunir ces deux résultats. De fait, répondre par un mouvement oculaire questionnerait une fenêtre temporelle d'attention plus précoce que celle mesurée par la réponse par pression de touche (Bannerman et al., 2010). Cela peut s'illustrer par la différence des temps de réponse mesurés dans la tâche d'eye-dot-probe, environ 100ms plus courts que ceux de la tâche de dot-probe classique. Aussi, il est possible que le biais attentionnel mesuré dans la tâche d'eye-dot-probe corresponde à une fenêtre temporelle dans laquelle la pertinence aux buts n'a pas encore été évaluée complètement. Cela pourrait refléter le fait que le trait d'anxiété décrit le mieux le niveau de biais attentionnel vers la menace dans la tâche d'eye-dot-probe. Cela dit, il est à noter

que l'intérêt de cette étude était d'identifier le score qui décrivait le mieux le biais attentionnel et non de tester si ce score expliquait significativement mieux le biais attentionnel par rapport aux autres. Les résultats concernant la relation entre score d'anxiété et biais attentionnel ne peuvent donc pas être avancés en tant que preuves pour des hypothèses théoriques à ce sujet.

### **e. ITACHI : vers une théorie intégrée de l'appraisal et de l'inférence active**

Enfin, nous avons rédigé un essai théorique (reporté dans le Chapitre 6 du manuscrit) visant à intégrer les postulats des théories de l'appraisal dans ceux de la théorie de l'inférence active au sein d'une théorie appelée ITACHI (i.e., « Integrated Theory of Appraisal under Causal Hierarchical Inference »). L'idée d'ITACHI est de définir le processus d'évaluation cognitive décrit par les théories de l'appraisal (Scherer, 1984, 2009) comme une partie du modèle génératif de l'environnement mis en place par le cerveau proposé par la théorie de l'inférence active (Friston et al., 2017; Hohwy, 2013). Cette idée permet de préciser la manière dont le processus d'évaluation cognitive est conceptualisé. Plus précisément, chaque critère du processus d'évaluation cognitive pourrait être défini comme un niveau de cause hiérarchique de l'implication future des stimuli pour l'homéostasie des individus (Friston, 2012b; Friston et al., 2010; Kirchhoff et al., 2018). Ainsi, les critères seraient évalués par inférence Bayésienne (Colombo & Seriès, 2012; Durante & Dunson, 2018; Lin & Garrido, 2022) en encodant les probabilités de distributions des hypothèses probables relatives à chaque critère afin d'identifier la plus crédible. Du point de vue de l'organisation du processus d'évaluation cognitive, ITACHI amène à penser que les critères d'évaluation cognitive de plus haut niveau permettraient de prédire ceux de plus bas niveaux, tandis que ceux de plus bas niveau viseraient à corriger les prédictions qui leur sont transmises (Friston, 2005, 2018). Ces échanges bidirectionnels de prédictions et d'erreurs de prédiction seraient régis par le principe d'énergie libre (Friston,

2010, 2012) selon lequel la sélection de l'état de chaque critère d'évaluation serait effectuée sur la base de l'hypothèse qui minimise le plus l'énergie libre (i.e., l'erreur de prédiction).

Ces considérations offrent une nouvelle perspective concernant certaines hypothèses des théories de l'appraisal (voir Chapitre 6 pour une discussion plus exhaustive à ce sujet) et sont particulièrement intéressantes pour expliquer, de manière intégrative, les résultats des études menées au cours de cette thèse. Premièrement, ITACHI semble pertinente pour expliquer les effets multiplicatifs de la pertinence intrinsèque et de la pertinence aux buts sur l'attention lorsque ces deux pertinences génèrent une tendance à l'approche. En effet, l'augmentation du niveau de capture attentionnelle en direction des stimuli dont les pertinences sont alignées pourrait refléter une manière de minimiser l'énergie libre. En d'autres termes, la crédibilité de la prédiction sur la pertinence de ces stimuli est très importante car peu d'erreur subsiste lorsque les pertinences sont alignées. Ces stimuli capteraient donc l'attention car il est très probable qu'ils nécessitent d'être approchés. Dans le cas de la capture attentionnelle générée par des stimuli dont les pertinences sont antagonistes, ce résultat pourrait refléter la nécessité de minimiser l'énergie libre. Une grande incertitude subsiste au sujet de ces stimuli dans la mesure où les pertinences prédisent qu'ils doivent probablement être à la fois approchés et évités. La capture attentionnelle augmentée par ces stimuli refléterait la nécessité de les évaluer plus en détail afin de minimiser l'incertitude – ou l'énergie libre. Dans le cas où la capture attentionnelle engendrée par des stimuli intrinsèquement pertinents est annulée par une « non » pertinence aux buts, cela pourrait refléter une différence de poids dans la crédibilité des prédictions des pertinences. En d'autres termes, dans le processus de sélection de l'hypothèse de la pertinence des stimuli, minimiser l'énergie libre pourrait revenir à tenir seulement compte de la prédiction de la pertinence aux buts, selon laquelle les stimuli ne nécessitent pas d'action si celle-ci est très crédible. Enfin, bien que nous n'ayons pas réussi à mettre en évidence un impact attentionnel différent des pertinences selon la temporalité (Aue et al., 2007; Grandjean &

Scherer, 2008; Lanctôt & Hess, 2007; Scherer, 2013), cette hypothèse peut également être entendue dans le cadre d'ITACHI. Cet impact différent pourrait également être un moyen de minimiser l'énergie libre en accordant un poids plus important à la prédiction de la pertinence qui semble être la plus crédible, i.e., la pertinence intrinsèque à courte temporalité et la pertinence aux buts à plus longue temporalité.

## 2. Conclusion

En résumé, cette thèse apporte de nouvelles perspectives théoriques et méthodologiques à la thématique de l'attention émotionnelle et plus généralement à celle de l'émotion. L'interprétation des résultats obtenus dans cette thèse dans le cadre d'ITACHI pourrait amener à considérer un modèle computationnel de l'attention émotionnelle novateur, sous-tendu par les postulats des théories de l'appraisal et de la théorie de l'inférence active. Selon ce modèle, l'attention émotionnelle répondrait aux prédictions résultantes de l'évaluation des stimuli selon le critère de pertinence. La hiérarchie de la pertinence des stimuli serait établie selon un calcul de la quantité d'énergie libre entre les prédictions de leur pertinence intrinsèque et de leur pertinence aux buts. Plusieurs règles régiraient l'établissement de cette hiérarchie :

- 1) Si plusieurs stimuli sont prédits comme étant d'un niveau de pertinence similaire, le stimulus dont les prédictions de la pertinence intrinsèque et de la pertinence aux buts sont les mieux alignées sera priorisé dans la hiérarchie. C'est le cas, par exemple, si la pertinence intrinsèque et la pertinence aux buts du stimulus génèrent un niveau similaire de tendance à l'approche. Ce stimulus est celui qui minimise le plus la quantité d'énergie libre, d'erreur de prédiction ou d'incertitude.
- 2) Si un stimulus est prédit comme nécessitant probablement d'être à la fois approché et évité (e.g., s'il est évalué comme pouvant faciliter la réalisation d'un but et déplaisant), il sera également priorisé dans la hiérarchie, dans le but de l'évaluer de façon plus intensive afin, à terme, de minimiser l'énergie libre.



- 3) Si un stimulus est associé à des prédictions de pertinence intrinsèque et de pertinence aux buts dont la confiance diffère drastiquement (e.g., une prédiction très confiante et prédiction incertaine), il pourra être placé dans la hiérarchie en tenant compte uniquement de la prédiction la plus confiante. Par exemple, si un stimulus est prédit (de manière incertaine) comme nécessitant d'être approché car il est intrinsèquement plaisant mais qu'il est prédit (de manière très confiante) comme ne pouvant pas impacter la réalisation d'un but, il pourra être considéré comme non pertinent. Cela pourrait se modéliser par le fait d'accorder un poids très important à la prédiction confiante au sein du calcul de pertinence totale, afin de minimiser la quantité d'énergie libre.
- 4) À courte temporalité, la prédiction de la pertinence intrinsèque aurait un poids plus important dans l'établissement de la hiérarchie de la pertinence des stimuli car la prédiction de la pertinence aux buts serait encore incertaine, dans la mesure où son inférence nécessiterait plus de temps de traitement. Plus la confiance accordée à la prédiction de la pertinence aux buts augmente à mesure que son processus d'inférence opère, plus le poids de sa prédiction serait important dans l'établissement de la hiérarchie de la pertinence des stimuli. Dans la mesure où la pertinence aux buts tient compte des buts et besoins actuels des individus, sa prédiction devrait être plus précise concernant la signification des stimuli actuels des stimuli pour le bien-être et la survie des individus.

De futures études pourraient être entreprises afin de tester ces prédictions. Elles pourraient nous aider à mieux comprendre le fonctionnement de l'attention émotionnelle.

Enfin, du point de vue méthodologique, cette thèse propose également une version modifiée de la tâche de dot-probe pour mesurer la capture attentionnelle générée par les stimuli. Comme l'ont montré les résultats de notre étude, la tâche d'eye-dot-probe semble permettre une mesure plus fine de la capture attentionnelle par les stimuli, en indiquant le côté d'apparition de la cible à l'aide de mouvements oculaires plutôt que par pression manuelle d'une touche de réponse.

Bien que le design de cette tâche nécessite encore d'être amélioré, elle pourrait s'avérer précieuse dans de prochaines études nécessitant une mesure fine de la capture attentionnelle.

## Bibliographie

- Adams, R. A., Shipp, S., & Friston, K. J. (2013). Predictions not commands : Active inference in the motor system. *Brain Structure and Function*, *218*(3), 611-643.  
<https://doi.org/10.1007/s00429-012-0475-5>
- Anderson, A. K. (2005). Affective Influences on the Attentional Dynamics Supporting Awareness. *Journal of Experimental Psychology: General*, *134*(2), 258-281.  
<https://doi.org/10.1037/0096-3445.134.2.258>
- Anderson, A. K., Christoff, K., Stappen, I., Panitz, D., Ghahremani, D. G., Glover, G., Gabrieli, J. D. E., & Sobel, N. (2003). Dissociated neural representations of intensity and valence in human olfaction. *Nature Neuroscience*, *6*(2), 196-202.  
<https://doi.org/10.1038/nn1001>
- Arnold, M. B. (1969). *Emotion and personality. 1 : Psychological aspects* (3. print). Columbia Univ. Pr.
- Aue, T., Flykt, A., & Scherer, K. R. (2007). First evidence for differential and sequential efferent effects of stimulus relevance and goal conduciveness appraisal. *Biological Psychology*, *74*(3), 347-357. <https://doi.org/10.1016/j.biopsycho.2006.09.001>
- Aue, T., & Scherer, K. R. (2008). Appraisal-driven somatovisceral response patterning : Effects of intrinsic pleasantness and goal conduciveness. *Biological Psychology*, *79*(2), 158-164. <https://doi.org/10.1016/j.biopsycho.2008.04.004>
- Aue, T., & Scherer, K. R. (2011). Effects of intrinsic pleasantness and goal conduciveness appraisals on somatovisceral responding : Somewhat similar, but not identical. *Biological Psychology*, *86*(1), 65-73. <https://doi.org/10.1016/j.biopsycho.2010.10.008>
- Bard, P. (1928). A diencephalic mechanism for the expression of rage with special reference to the sympathetic nervous system. *American Journal of Physiology-Legacy Content*, *84*(3), 490-515. <https://doi.org/10.1152/ajplegacy.1928.84.3.490>

- Bar-Haim, Y., Lamy, D., Pergamin, L., Bakermans-Kranenburg, M. J., & van IJzendoorn, M. H. (2007). Threat-related attentional bias in anxious and nonanxious individuals : A meta-analytic study. *Psychological Bulletin*, *133*(1), 1-24. <https://doi.org/10.1037/0033-2909.133.1.1>
- Barrett, L. F., & Simmons, W. K. (2015). Interoceptive predictions in the brain. *Nature Reviews Neuroscience*, *16*(7), 419-429. <https://doi.org/10.1038/nrn3950>
- Bechara, A., Damasio, H., Tranel, D., & Damasio, A. R. (2005). The Iowa Gambling Task and the somatic marker hypothesis : Some questions and answers. *Trends in Cognitive Sciences*, *9*(4), 159-162. <https://doi.org/10.1016/j.tics.2005.02.002>
- Beck, A. T., Emery, G., & Greenberg, R. L. (2005). *Anxiety disorders and phobias : A cognitive perspective* ([15th anniversary ed.], rev. pbk. ed). Basic Books.
- Berridge, K. C. (2009). 'Liking' and 'wanting' food rewards : Brain substrates and roles in eating disorders. *Physiology & Behavior*, *97*(5), 537-550. <https://doi.org/10.1016/j.physbeh.2009.02.044>
- Berridge, K. C., Robinson, T. E., & Aldridge, J. W. (2009). Dissecting components of reward : 'Liking', 'wanting', and learning. *Current Opinion in Pharmacology*, *9*(1), 65-73. <https://doi.org/10.1016/j.coph.2008.12.014>
- Bottemanne, H., Longuet, Y., & Gauld, C. (2022). L'esprit prédictif : Introduction à la théorie du cerveau bayésien. *L'Encéphale*, *48*(4), 436-444. <https://doi.org/10.1016/j.encep.2021.09.011>
- Bower, G. H. (1981). Mood and memory. *American Psychologist*, *36*(2), 129-148. <https://doi.org/10.1037/0003-066X.36.2.129>
- Bradley, M. M., Codispoti, M., Cuthbert, B. N., & Lang, P. J. (2001). Emotion and motivation I: Defensive and appetitive reactions in picture processing. *Emotion*, *1*(3), 276-298. <https://doi.org/10.1037/1528-3542.1.3.276>

- Brosch, T., Pourtois, G., Sander, D., & Vuilleumier, P. (2011). Additive effects of emotional, endogenous, and exogenous attention : Behavioral and electrophysiological evidence. *Neuropsychologia*, *49*(7), 1779-1787.  
<https://doi.org/10.1016/j.neuropsychologia.2011.02.056>
- Brosch, T., Sander, D., Pourtois, G., & Scherer, K. R. (2008). Beyond Fear : Rapid Spatial Orienting Toward Positive Emotional Stimuli. *Psychological Science*, *19*(4), 362-370.  
<https://doi.org/10.1111/j.1467-9280.2008.02094.x>
- Brosch, T., & Sharma, D. (2005). The Role of Fear-Relevant Stimuli in Visual Search. *Emotion*, *5*(3), 360-364.
- Buckley, C. L., Kim, C. S., McGregor, S., & Seth, A. K. (2017). The free energy principle for action and perception : A mathematical review. *Journal of Mathematical Psychology*, *81*, 55-79. <https://doi.org/10.1016/j.jmp.2017.09.004>
- Cannon, W. B. (1927). The James-Lange Theory of Emotions : A Critical Examination and an Alternative Theory. *The American Journal of Psychology*, *39*(1/4), 106.  
<https://doi.org/10.2307/1415404>
- Castellanos, E. H., Charboneau, E., Dietrich, M. S., Park, S., Bradley, B. P., Mogg, K., & Cowan, R. L. (2009). Obese adults have visual attention bias for food cue images : Evidence for altered reward system function. *International Journal of Obesity*, *33*(9), 1063-1073. <https://doi.org/10.1038/ijo.2009.138>
- Castro, D. C., & Berridge, K. C. (2014). Advances in the neurobiological bases for food 'liking' versus 'wanting'. *Physiology & Behavior*, *136*, 22-30.  
<https://doi.org/10.1016/j.physbeh.2014.05.022>
- Colombo, M., & Seriès, P. (2012). Bayes in the Brain—On Bayesian Modelling in Neuroscience. *The British Journal for the Philosophy of Science*, *63*(3), 697-723.  
<https://doi.org/10.1093/bjps/axr043>

- Craig, A. D. (2002). How do you feel? Interoception: the sense of the physiological condition of the body. *Nature Reviews Neuroscience*, 3(8), 655-666. <https://doi.org/10.1038/nrn894>
- Craig, A. D. (2003). Interoception : The sense of the physiological condition of the body. *Current Opinion in Neurobiology*, 13(4), 500-505. [https://doi.org/10.1016/S0959-4388\(03\)00090-4](https://doi.org/10.1016/S0959-4388(03)00090-4)
- Dan Glaser, E. S., & Scherer, K. R. (2008). Neuronal Processes Involved in Subjective Feeling Emergence : Oscillatory Activity During an Emotional Monitoring Task. *Brain Topography*, 20(4), 224-231. <https://doi.org/10.1007/s10548-008-0048-3>
- Darwin, C. (1998). General Principles of Expression. In *The expression of the emotions in man and animals* (3rd ed, p. 33-44). Oxford University Press. (Original work published 1872).
- Dukes, D., Abrams, K., Adolphs, R., Ahmed, M. E., Beatty, A., Berridge, K. C., Broomhall, S., Brosch, T., Campos, J. J., Clay, Z., Clément, F., Cunningham, W. A., Damasio, A., Damasio, H., D'Arms, J., Davidson, J. W., de Gelder, B., Deonna, J., de Sousa, R., ... Sander, D. (2021). The rise of affectivism. *Nature Human Behaviour*, 5(7), 816-820. <https://doi.org/10.1038/s41562-021-01130-8>
- Durante, D., & Dunson, D. B. (2018). Bayesian Inference and Testing of Group Differences in Brain Networks. *Bayesian Analysis*, 13(1). <https://doi.org/10.1214/16-BA1030>
- Ekman, P. (1992). Are there basic emotions? *Psychological Review*, 99(3), 550-553. <https://doi.org/10.1037/0033-295X.99.3.550>
- Ekman, P., & Friesen, W. V. (1978). Facial Action Coding System [Data set]. In *Environmental Psychology & Nonverbal Behavior*. <https://doi.org/10.1037/t27734-000>
- Ekman, P., Friesen, W. V., & Ellsworth, P. (1972). Introduction. In *Emotion in the human face : Guide-lines for research and an integration of findings* (p. 1-6). Pergamon Press.
- Ellsworth, P. C. (1991). Some Implications of Cognitive Appraisal Theories of Emotion. In

- International review of studies on emotion* (K. T. Strongman, Vol. 1, p. 143-161). Wiley.
- Ellsworth, P. C. (1994). William James and emotion : Is a century of fame worth a century of misunderstanding? *Psychological Review*, *101*(2), 222-229. <https://doi.org/10.1037/0033-295X.101.2.222>
- Ellsworth, P. C., & Dougherty, A. (2016). Appraisals and Reappraisals in the Courtroom. *Emotion Review*, *8*(1), 20-25. <https://doi.org/10.1177/1754073915601227>
- Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance : Attentional control theory. *Emotion*, *7*(2), 336-353. <https://doi.org/10.1037/1528-3542.7.2.336>
- Field, M., & Cox, W. (2008). Attentional bias in addictive behaviors : A review of its development, causes, and consequences. *Drug and Alcohol Dependence*, *97*(1-2), 1-20. <https://doi.org/10.1016/j.drugalcdep.2008.03.030>
- Fodor, J. A. (1983). *The modularity of mind : An essay on faculty psychology*. MIT Press.
- Folkman, & Lazarus, R. S. (1985). If it changes it must be a process : Study of emotion and coping during three stages of a college examination. *Journal of Personality and Social Psychology*, *48*(1), 150-170. <https://doi.org/10.1037/0022-3514.48.1.150>
- Fontaine, J. R. J., Scherer, K. R., Roesch, E. B., & Ellsworth, P. C. (2007). The World of Emotions is not Two-Dimensional. *Psychological Science*, *18*(12), 1050-1057. <https://doi.org/10.1111/j.1467-9280.2007.02024.x>
- Forgas, J. P., & Smith, C. (2007). Affect and emotion. In *The SAGE handbook of social psychology* (M. A. Hogg & J. Cooper, p. 146-175).
- Forrest, J., Vogt, J., McDonald, C., Searle, C., & Sakaki, M. (2022). Blind to threat : The presence of temporary goals prevents attention to imminent threat already at early stages of attention allocation. *Motivation Science*, *8*(3), 239-251. <https://doi.org/10.1037/mot0000261>

- Fournier, H., Lefèvre, E., Guerin, E., & Koenig, O. (Submitted). The eye dot-probe task: an innovative lead to overcome the crisis on attentional bias toward threat.
- Fournier, H., & Koenig, O. (2023). Combined effects of intrinsic and goal relevances on attention and action tendency during the emotional episode. *Emotion*, 23(2), 425-436. <https://doi.org/10.1037/emo0001081>
- Fournier, H. & Koenig, O. (Under review). From emotion to action tendency: Time course of attentional effects of agonistic and antagonistic overlay of intrinsic and goal relevances.
- Fournier, H. & Koenig, O. (Submitted a). Attentional impact of relevance overlay: Effects of goal conduction and unavoidably goal obstruction inductions on pleasant and unpleasant stimuli.
- Fournier, H. & Koenig, O. (Submitted b). Integrated Theory of Appraisal under Causal Hierarchical Inference: an attempt to map appraisal theories to active inference theory.
- Fox, E., Griggs, L., & Mouchlianitis, E. (2007). The detection of fear-relevant stimuli : Are guns noticed as quickly as snakes? *Emotion*, 7(4), 691-696. <https://doi.org/10.1037/1528-3542.7.4.691>
- Frijda, N. H. (1986). *The emotions* (Maison des sciences de l'homme). Cambridge University Press.
- Frijda, N. H. (1987). Emotion, cognitive structure, and action tendency. *Cognition & Emotion*, 1(2), 115-143. <https://doi.org/10.1080/02699938708408043>
- Frijda, N. H. (2009). Emotion Experience and its Varieties. *Emotion Review*, 1(3), 264-271. <https://doi.org/10.1177/1754073909103595>
- Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1456), 815-836. <https://doi.org/10.1098/rstb.2005.1622>
- Friston, K. (2008). Hierarchical Models in the Brain. *PLoS Computational Biology*, 4(11),



- e1000211. <https://doi.org/10.1371/journal.pcbi.1000211>
- Friston, K. (2010). The free-energy principle : A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127-138. <https://doi.org/10.1038/nrn2787>
- Friston, K. (2011). What Is Optimal about Motor Control? *Neuron*, *72*(3), 488-498. <https://doi.org/10.1016/j.neuron.2011.10.018>
- Friston, K. (2012a). The history of the future of the Bayesian brain. *NeuroImage*, *62*(2), 1230-1233. <https://doi.org/10.1016/j.neuroimage.2011.10.004>
- Friston, K. (2012b). A Free Energy Principle for Biological Systems. *Entropy*, *14*(11), 2100-2121. <https://doi.org/10.3390/e14112100>
- Friston, K. (2018). Does predictive coding have a future? *Nature Neuroscience*, *21*(8), 1019-1021. <https://doi.org/10.1038/s41593-018-0200-7>
- Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., & Pezzulo, G. (2017). Active Inference : A Process Theory. *Neural Computation*, *29*(1), 1-49. [https://doi.org/10.1162/NECO\\_a\\_00912](https://doi.org/10.1162/NECO_a_00912)
- Friston, K. J., Daunizeau, J., Kilner, J., & Kiebel, S. J. (2010). Action and behavior : A free-energy formulation. *Biological Cybernetics*, *102*(3), 227-260. <https://doi.org/10.1007/s00422-010-0364-z>
- Friston, K., Schwartenbeck, P., FitzGerald, T., Moutoussis, M., Behrens, T., & Dolan, R. J. (2013). The anatomy of choice : Active inference and agency. *Frontiers in Human Neuroscience*, *7*. <https://doi.org/10.3389/fnhum.2013.00598>
- Fuller, S., & Carrasco, M. (2006). Exogenous attention and color perception : Performance and appearance of saturation and hue. *Vision Research*, *46*(23), 4032-4047. <https://doi.org/10.1016/j.visres.2006.07.014>
- Garg, D., & Verma, G. K. (2020). Emotion Recognition in Valence-Arousal Space from Multi-channel EEG data and Wavelet based Deep Learning Framework. *Procedia*

- Computer Science*, 171, 857-867. <https://doi.org/10.1016/j.procs.2020.04.093>
- Grandjean, D., Sander, D., & Scherer, K. R. (2008). Conscious emotional experience emerges as a function of multilevel, appraisal-driven response synchronization. *Consciousness and Cognition*, 17(2), 484-495. <https://doi.org/10.1016/j.concog.2008.03.019>
- Grandjean, D., & Scherer, K. R. (2008). Unpacking the cognitive architecture of emotion processes. *Emotion*, 8(3), 341-351. <https://doi.org/10.1037/1528-3542.8.3.341>
- Grandjean, D., & Scherer, K. R. (2014). Chapitre 2. Théorie de l'évaluation cognitive et dynamique des processus émotionnels [Chapter 2. Theory of cognitive appraisal and dynamics of emotional processes]. In *Traité de Psychologie des émotions [Treatise on the psychology of emotions]* (In D. Sander, p. 51-87). Dunod. <https://www.cairn.info/traite-de-psychologie-des-emotions--9782100705344-page-51.htm?contenu=resume>
- Gross, J. J. (2015). Emotion Regulation : Current Status and Future Prospects. *Psychological Inquiry*, 26(1), 1-26. <https://doi.org/10.1080/1047840X.2014.940781>
- Grossberg, S. (2009). Cortical and subcortical predictive dynamics and learning during perception, cognition, emotion and action. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1521), 1223-1234. <https://doi.org/10.1098/rstb.2008.0307>
- Hackley, S. A., & Graham, F. K. (1983). Early selective attention effects on cutaneous and acoustic blink reflexes. *Physiological Psychology*, 11(4), 235-242. <https://doi.org/10.3758/BF03326801>
- Harrison, N. A., Voon, V., Cercignani, M., Cooper, E. A., Pessiglione, M., & Critchley, H. D. (2016). A Neurocomputational Account of How Inflammation Enhances Sensitivity to Punishments Versus Rewards. *Biological Psychiatry*, 80(1), 73-81. <https://doi.org/10.1016/j.biopsych.2015.07.018>
- Hesselmann, G., Sadaghiani, S., Friston, K. J., & Kleinschmidt, A. (2010). Predictive Coding or Evidence Accumulation? False Inference and Neuronal Fluctuations. *PLoS ONE*, 5(3),

e9926. <https://doi.org/10.1371/journal.pone.0009926>

- Hoefling, A., Likowski, K. U., Deutsch, R., Häfner, M., Seibt, B., Mühlberger, A., Weyers, P., & Strack, F. (2009). When hunger finds no fault with moldy corn : Food deprivation reduces food-related disgust. *Emotion, 9*(1), 50-58. <https://doi.org/10.1037/a0014449>
- Hohwy, J. (2013). *Into the predictive mind* (Hohwy, J.). Oxford, UK: Oxford University Press.
- Iacoviello, B. M., Wu, G., Abend, R., Murrough, J. W., Feder, A., Fruchter, E., Levinstein, Y., Wald, I., Bailey, C. R., Pine, D. S., Neumeister, A., Bar-Haim, Y., & Charney, D. S. (2014). Attention Bias Variability and Symptoms of Posttraumatic Stress Disorder : Attention Bias Variability in PTSD. *Journal of Traumatic Stress, 27*(2), 232-239. <https://doi.org/10.1002/jts.21899>
- James, W. (1890). *The Principles of Psychology*. Holt, New York.
- Kahneman, D. (1973). *Attention and effort*. Prentice-Hall.
- Kersten, D., Mamassian, P., & Yuille, A. (2004). Object Perception as Bayesian Inference. *Annual Review of Psychology, 55*(1), 271-304. <https://doi.org/10.1146/annurev.psych.55.090902.142005>
- Kirchhoff, M., Parr, T., Palacios, E., Friston, K., & Kiverstein, J. (2018). The Markov blankets of life : Autonomy, active inference and the free energy principle. *Journal of The Royal Society Interface, 15*(138), 20170792. <https://doi.org/10.1098/rsif.2017.0792>
- Kruijt, A.-W., Parsons, S., & Fox, E. (2019). A meta-analysis of bias at baseline in RCTs of attention bias modification : No evidence for dot-probe bias towards threat in clinical anxiety and PTSD. *Journal of Abnormal Psychology, 128*(6), 563-573. <https://doi.org/10.1037/abn0000406>
- Lanctôt, N., & Hess, U. (2007). The timing of appraisals. *Emotion, 7*(1), 207-212. <https://doi.org/10.1037/1528-3542.7.1.207>

- Lange, C. G. (1885). The mechanism of the emotions. In *The classical psychologists* (p. 672-684.).
- Lavazza, A. (2009). Art as a metaphor of the mind : A neo-Jamesian aesthetics embracing phenomenology, neuroscience, and evolution. *Phenomenology and the Cognitive Sciences*, 8(2), 159-182. <https://doi.org/10.1007/s11097-008-9091-5>
- Lazarus, R. S., & Folkman, S. (1984). *Stress, appraisal, and coping* (11. [print.]). Springer.
- LeDoux, J. (2003). The Emotional Brain, Fear, and the Amygdala. *Cellular and Molecular Neurobiology*, 23(4), 727-738. <https://doi.org/10.1023/A:1025048802629>
- Levenson, R. W. (2006). Blood, Sweat, and Fears : The Autonomic Architecture of Emotion. *Annals of the New York Academy of Sciences*, 1000(1), 348-366. <https://doi.org/10.1196/annals.1280.016>
- Lin, C.-H. S., & Garrido, M. I. (2022). Towards a cross-level understanding of Bayesian inference in the brain. *Neuroscience & Biobehavioral Reviews*, 137, 104649. <https://doi.org/10.1016/j.neubiorev.2022.104649>
- LoBue, V., & DeLoache, J. S. (2008). Detecting the Snake in the Grass : Attention to Fear-Relevant Stimuli by Adults and Young Children. *Psychological Science*, 19(3), 284-289. <https://doi.org/10.1111/j.1467-9280.2008.02081.x>
- MacLeod, C., Mathews, A., & Tata, P. (1986). Attentional bias in emotional disorders. *Journal of Abnormal Psychology*, 95(1), 15-20. <https://doi.org/10.1037/0021-843X.95.1.15>
- Massumi, B. (2002). The autonomy of affect. In *Parables for the Virtual, Movement, Affect, Sensation* (p. 25-48). Duke university press.
- Maxwell, J. S., & Davidson, R. J. (2007). Emotion as Motion : Asymmetries in Approach and Avoidant Actions. *Psychological Science*, 18(12), 1113-1119. <https://doi.org/10.1111/j.1467-9280.2007.02033.x>
- Mazzietti, A., Sellem, V., & Koenig, O. (2014). From stimulus-driven to appraisal-driven

- attention : Towards differential effects of goal relevance and goal relatedness on attention?  
*Cognition and Emotion*, 28(8), 1483-1492. <https://doi.org/10.1080/02699931.2014.884488>
- McNally, R. J. (2019). Attentional bias for threat : Crisis or opportunity? *Clinical Psychology Review*, 69, 4-13. <https://doi.org/10.1016/j.cpr.2018.05.005>
- Milne, S., Lomax, C., & Freeston, M. H. (2019). A review of the relationship between intolerance of uncertainty and threat appraisal in anxiety. *The Cognitive Behaviour Therapist*, 12, e38. <https://doi.org/10.1017/S1754470X19000230>
- Mineka, S., & Öhman, A. (2002). Phobias and preparedness : The selective, automatic, and encapsulated nature of fear. *Biological Psychiatry*, 52(10), 927-937.  
[https://doi.org/10.1016/S0006-3223\(02\)01669-4](https://doi.org/10.1016/S0006-3223(02)01669-4)
- Mogg, K., & Bradley, B. P. (2016). Anxiety and attention to threat : Cognitive mechanisms and treatment with attention bias modification. *Behaviour Research and Therapy*, 87, 76-108. <https://doi.org/10.1016/j.brat.2016.08.001>
- Mogg, K., & Bradley, B. P. (2018). Anxiety and Threat-Related Attention : Cognitive-Motivational Framework and Treatment. *Trends in Cognitive Sciences*, 22(3), 225-240.  
<https://doi.org/10.1016/j.tics.2018.01.001>
- Molloy, A., & Anderson, P. L. (2020). Evaluating the reliability of attention bias and attention bias variability measures in the dot-probe task among people with social anxiety disorder. *Psychological Assessment*, 32(9), 883-888. <https://doi.org/10.1037/pas0000912>
- Moors, A. (2014). Flavors of Appraisal Theories of Emotion. *Emotion Review*, 6(4), 303-307.  
<https://doi.org/10.1177/1754073914534477>
- Moors, A., Ellsworth, P. C., Scherer, K. R., & Frijda, N. H. (2013). Appraisal Theories of Emotion : State of the Art and Future Development. *Emotion Review*, 5(2), 119-124.  
<https://doi.org/10.1177/1754073912468165>
- Müller, P., & Wentura, D. (2023). Undeserved reward but not inevitable loss biases attention :

- Personal control moderates evaluative attentional biases in the additional-singleton paradigm. *Quarterly Journal of Experimental Psychology*, 76(4), 888-904.  
<https://doi.org/10.1177/17470218221099125>
- Müller, S., Rothermund, K., & Wentura, D. (2016). Relevance drives attention : Attentional bias for gain- and loss-related stimuli is driven by delayed disengagement. *Quarterly Journal of Experimental Psychology*, 69(4), 752-763.  
<https://doi.org/10.1080/17470218.2015.1049624>
- Myers, D. G. (2002). Practical intuition. In *Intuition : Its powers and perils* (p. 133-227). Yale University Press.
- Naim, R., Abend, R., Wald, I., Eldar, S., Levi, O., Fruchter, E., Ginat, K., Halpern, P., Sipos, M. L., Adler, A. B., Bliese, P. D., Quartana, P. J., Pine, D. S., & Bar-Haim, Y. (2015). Threat-Related Attention Bias Variability and Posttraumatic Stress. *American Journal of Psychiatry*, 172(12), 1242-1250. <https://doi.org/10.1176/appi.ajp.2015.14121579>
- Nijss, I. M. T., Muris, P., Euser, A. S., & Franken, I. H. A. (2010). Differences in attention to food and food intake between overweight/obese and normal-weight females under conditions of hunger and satiety. *Appetite*, 54(2), 243-254.  
<https://doi.org/10.1016/j.appet.2009.11.004>
- Öhman, A. (1999). Distinguish unconscious from conscious emotional processes : Methodological considerations and theoretical implications. In *Handbook of Cognition and Emotion* (T. Dalgleish&M.J. Power, p. 321-352). John Wiley & Sons.
- Öhman, A., Flykt, A., & Esteves, F. (2001). Emotion drives attention : Detecting the snake in the grass. *Journal of Experimental Psychology: General*, 130(3), 466-478.  
<https://doi.org/10.1037/0096-3445.130.3.466>
- Öhman, A., Lundqvist, D., & Esteves, F. (2001). The face in the crowd revisited : A threat advantage with schematic stimuli. *Journal of Personality and Social Psychology*, 80(3),

381-396. <https://doi.org/10.1037/0022-3514.80.3.381>

Öhman, A., & Mineka, S. (2001). Fears, phobias, and preparedness : Toward an evolved module of fear and fear learning. *Psychological Review*, *108*(3), 483-522.

<https://doi.org/10.1037/0033-295X.108.3.483>

Öhman, A., & Wiens, S. (2009). On the automaticity of autonomic responses in emotion : An evolutionary perspective. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Éds.), *Handbook of affective sciences* (p. 256-275). Oxford University Press.

Pecchinenda, A. (2001). The psychophysiology of appraisals. In K. R. Scherer, A. Schorr, & T. Johnstone (Éds.), *Appraisal processes in emotion : Theory, methods, research* (p. 301-315). Oxford University Press.

Pezzulo, G., Rigoli, F., & Friston, K. (2015). Active Inference, homeostatic regulation and adaptive behavioural control. *Progress in Neurobiology*, *134*, 17-35.

<https://doi.org/10.1016/j.pneurobio.2015.09.001>

Pool, E., Brosch, T., Delplanque, S., & Sander, D. (2016). Attentional bias for positive emotional stimuli : A meta-analytic investigation. *Psychological Bulletin*, *142*(1), 79-106.

<https://doi.org/10.1037/bul0000026>

Poppa, T., & Bechara, A. (2018). The somatic marker hypothesis : Revisiting the role of the ‘body-loop’ in decision-making. *Current Opinion in Behavioral Sciences*, *19*, 61-66.

<https://doi.org/10.1016/j.cobeha.2017.10.007>

Posner, M. I. (1980). Orienting of Attention. *Quarterly Journal of Experimental Psychology*, *32*(1), 3-25. <https://doi.org/10.1080/00335558008248231>

Price, R. B., Kuckertz, J. M., Siegle, G. J., Ladouceur, C. D., Silk, J. S., Ryan, N. D., Dahl, R. E., & Amir, N. (2015). Empirical recommendations for improving the stability of the dot-probe task in clinical research. *Psychological Assessment*, *27*(2), 365-376.

<https://doi.org/10.1037/pas0000036>

- Rauss, K., & Pourtois, G. (2013). What is Bottom-Up and What is Top-Down in Predictive Coding? *Frontiers in Psychology*, 4. <https://doi.org/10.3389/fpsyg.2013.00276>
- Reisenzein, R. (1983). The Schachter theory of emotion : Two decades later. *Psychological Bulletin*, 94(2), 239-264. <https://doi.org/10.1037/0033-2909.94.2.239>
- Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6), 1161-1178. <https://doi.org/10.1037/h0077714>
- Rutherford, H. J. V., & Lindell, A. K. (2011). Thriving and Surviving : Approach and Avoidance Motivation and Lateralization. *Emotion Review*, 3(3), 333-343. <https://doi.org/10.1177/1754073911402392>
- Rutkowski, T. M., Abe, M. S., Koculak, M., & Otake-Matsuura, M. (2020). Classifying Mild Cognitive Impairment from Behavioral Responses in Emotional Arousal and Valence Evaluation Task – AI Approach for Early Dementia Biomarker in Aging Societies –. *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 5537-5543. <https://doi.org/10.1109/EMBC44109.2020.9175805>
- Sander, D. (2013). Models of emotion : The affective neuroscience approach. In *The Cambridge handbook of human affective neuroscience* (J. Armony & P. Vuilleumier, p. 5-53).
- Sander, D., Grafman, J., & Zalla, T. (2003). The Human Amygdala : An Evolved System for Relevance Detection. *Reviews in the Neurosciences*, 14(4), 303-316. <https://doi.org/10.1515/REVNEURO.2003.14.4.303>
- Sander, D., Grandjean, D., & Scherer, K. R. (2005). A systems approach to appraisal mechanisms in emotion. *Neural Networks*, 18(4), 317-352. <https://doi.org/10.1016/j.neunet.2005.03.001>
- Sander, D., & Scherer, K. R. (2014). Chapitre 1. La psychologie des émotions : Survol des théories et débats essentiels: In *Traité de psychologie des émotions* (p. 11-50). Dunod.



<https://doi.org/10.3917/dunod.sande.2014.01.0011>

- Schachter, S., & Singer, J. (1962). Cognitive, social, and physiological determinants of emotional state. *Psychological Review*, *69*(5), 379-399. <https://doi.org/10.1037/h0046234>
- Scherer, K. R. (1984). On the nature and function of emotion : A component process approach. In K. R. Scherer & P. Ekman (Éds.), *Approaches to emotion* (p. 292-317). L. Erlbaum Associates.
- Scherer, K. R. (1993). Studying the emotion-antecedent appraisal process : An expert system approach. *Cognition & Emotion*, *7*(3-4), 325-355.  
<https://doi.org/10.1080/02699939308409192>
- Scherer, K. R. (2005a). Unconscious Processes in Emotion : The Bulk of the Iceberg. In L. F. Barrett, P. M. Niedenthal, & P. Winkielman (Éds.), *Emotion and consciousness* (p. 312-334). Guilford Press.
- Scherer, K. R. (2005b). What are emotions? And how can they be measured? *Social Science Information*, *44*(4), 695-729. <https://doi.org/10.1177/0539018405058216>
- Scherer, K. R. (2009). The dynamic architecture of emotion : Evidence for the component process model. *Cognition & Emotion*, *23*(7), 1307-1351.  
<https://doi.org/10.1080/02699930902928969>
- Scherer, K. R. (2013). The Nature and Dynamics of Relevance and Valence Appraisals : Theoretical Advances and Recent Evidence. *Emotion Review*, *5*(2), 150-162.  
<https://doi.org/10.1177/1754073912468166>
- Scherer, K. R., & Ellgring, H. (2007). Are facial expressions of emotion produced by categorical affect programs or dynamically driven by appraisal? *Emotion*, *7*(1), 113-130.  
<https://doi.org/10.1037/1528-3542.7.1.113>
- Scherer, K. R., Ellgring, H., Dieckmann, A., Unfried, M., & Mortillaro, M. (2019). Dynamic Facial Expression of Emotion and Observer Inference. *Frontiers in Psychology*, *10*, 508.

<https://doi.org/10.3389/fpsyg.2019.00508>

- Scherer, K. R., Mortillaro, M., & Mehu, M. (2013). Understanding the Mechanisms Underlying the Production of Facial Expression of Emotion : A Componential Perspective. *Emotion Review*, 5(1), 47-53. <https://doi.org/10.1177/1754073912451504>
- Schimmack, U. (2005). Attentional Interference Effects of Emotional Pictures : Threat, Negativity, or Arousal? *Emotion*, 5(1), 55-66. <https://doi.org/10.1037/1528-3542.5.1.55>
- Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Sciences*, 17(11), 565-573. <https://doi.org/10.1016/j.tics.2013.09.007>
- Seth, A. K. (2015). The Cybernetic Bayesian Brain : From Interoceptive Inference to Sensorimotor Contingencies: From Interoceptive Inference to Sensorimotor Contingencies. *Open MIND*. <https://doi.org/10.15502/9783958570108>
- Seth, A. K., & Friston, K. J. (2016). Active interoceptive inference and the emotional brain. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1708), 20160007. <https://doi.org/10.1098/rstb.2016.0007>
- Smith, C., & Lazarus, R. S. (1990). Emotion and adaptation. In *Handbook of personality : Theory and research* (L. A. Pervin, Vol. 21, p. 609-637).
- Smith, R., Badcock, P., & Friston, K. J. (2021). Recent advances in the application of predictive coding and active inference models within clinical neuroscience. *Psychiatry and Clinical Neurosciences*, 75(1), 3-13. <https://doi.org/10.1111/pcn.13138>
- Smith, R., Parr, T., & Friston, K. J. (2019). Simulating Emotions : An Active Inference Model of Emotional State Inference and Emotion Concept Learning. *Frontiers in Psychology*, 10, 2844. <https://doi.org/10.3389/fpsyg.2019.02844>
- Strack, F., Martin, L. L., & Stepper, S. (1988). Inhibiting and facilitating conditions of the human smile : A nonobtrusive test of the facial feedback hypothesis. *Journal of Personality and Social Psychology*, 54(5), 768-777. <https://doi.org/10.1037/0022-3514.54.5.768>

- Sugawara, A., Terasawa, Y., Katsunuma, R., & Sekiguchi, A. (2020). Effects of interoceptive training on decision making, anxiety, and somatic symptoms. *BioPsychoSocial Medicine*, *14*(1), 7. <https://doi.org/10.1186/s13030-020-00179-7>
- Sun, L., Ding, C., Xu, M., Diao, L., & Yang, D. (2017). Engagement Attentional Bias toward Value-Associated Stimuli. *Current Psychology*, *36*(4), 747-754. <https://doi.org/10.1007/s12144-016-9462-y>
- Uusberg, A., Yih, J., Taxer, J. L., Christ, N. M., Toms, T., Uusberg, H., & Gross, J. (2023). Appraisal shifts during reappraisal. *Emotion*, *Advance online publication*. <https://doi.org/10.1037/emo0001202>
- van Reekum, C., Johnstone, T., Banse, R., Etter, A., Wehrle, T., & Scherer, K. (2004). Psychophysiological responses to appraisal dimensions in a computer game. *Cognition & Emotion*, *18*(5), 663-688. <https://doi.org/10.1080/02699930341000167>
- Vogt, J., De Houwer, J., Crombez, G., & Van Damme, S. (2013). Competing for attentional priority : Temporary goals versus threats. *Emotion*, *13*(3), 587-598. <https://doi.org/10.1037/a0027204>
- Vogt, J., Koster, E. H. W., & De Houwer, J. (2017). Safety first : Instrumentality for reaching safety determines attention allocation under threat. *Emotion*, *17*(3), 528-537. <https://doi.org/10.1037/emo0000251>
- Vogt, J., Lozo, L., Koster, E. H. W., & De Houwer, J. (2011). On the role of goal relevance in emotional attention : Disgust evokes early attention to cleanliness. *Cognition and Emotion*, *25*(3), 466-477. <https://doi.org/10.1080/02699931.2010.532613>
- Vuilleumier, P. (2005). How brains beware : Neural mechanisms of emotional attention. *Trends in Cognitive Sciences*, *9*(12), 585-594. <https://doi.org/10.1016/j.tics.2005.10.011>
- Waters, A. M., Henry, J., Mogg, K., Bradley, B. P., & Pine, D. S. (2010). Attentional bias towards angry faces in childhood anxiety disorders. *Journal of Behavior Therapy and*

- Experimental Psychiatry*, 41(2), 158-164. <https://doi.org/10.1016/j.jbtep.2009.12.001>
- Werthmann, J., Roefs, A., Nederkoorn, C., Mogg, K., Bradley, B. P., & Jansen, A. (2011). Can(not) take my eyes off it : Attention bias for food in overweight participants. *Health Psychology*, 30(5), 561-569. <https://doi.org/10.1037/a0024291>
- Wundt, W. (1897). Outlines of Psychology. *Science*, 5(127), 882-884. <https://doi.org/10.1126/science.5.127.882>
- Yantis, S. (1998). Control of visual attention. In *Attention* (H. Pashler, p. 223-256). Hove, England: Psychology Press.
- Yiend, J. (2010). The effects of emotion on attention : A review of attentional processing of emotional information. *Cognition and Emotion*, 24(1), 3-47. <https://doi.org/10.1080/02699930903205698>
- Zsoldos, I., Tr an, E., Fournier, H., Tarpin-Bernard, F., Fruitet, J., Fouillen, M., Bailly, G., Elisei, F., Bouchot, B., Constant, P., Ringeval, F., Koenig, O., & Chainay, H. (Submitted). The value of a virtual assistant to improve engagement in computerized cognitive training at home: An exploratory study.
- Zvielli, A., Bernstein, A., & Koster, E. H. W. (2015). Temporal Dynamics of Attentional Bias. *Clinical Psychological Science*, 3(5), 772-788. <https://doi.org/10.1177/2167702614551572>