



HAL
open science

Neurocomputational mechanisms of meditative practice : investigating an intensive mindfulness meditation retreat within a predictive processing Bayesian framework

Arnaud Pouban-Couzardot

► To cite this version:

Arnaud Pouban-Couzardot. Neurocomputational mechanisms of meditative practice : investigating an intensive mindfulness meditation retreat within a predictive processing Bayesian framework. Neurosciences. Université de Lyon, 2022. English. NNT : 2022LYSE1110 . tel-04213136

HAL Id: tel-04213136

<https://theses.hal.science/tel-04213136>

Submitted on 21 Sep 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 : 2022LYSE1110

THÈSE de DOCTORAT DE L'UNIVERSITÉ DE LYON

opérée au sein de :

l'Université Claude Bernard Lyon 1

École Doctorale NSCo ED 476

Neurosciences et Cognition

Spécialité de doctorat : Neurosciences cognitives

Soutenue publiquement le 01/07/2022, par :

Arnaud POUBLAN-COUZARDOT

Neurocomputational mechanisms of meditative practice: investigating an intensive mindfulness meditation retreat within a predictive processing Bayesian framework

Devant le jury composé de :

RAFFONE Antonino, Prof., Université de Rome

Président.e

SLAGTER Heleen, Prof., Université de Vrije Amsterdam

Rapporteure

TALMI Deborah, Prof. Assist., Université de Cambridge

Rapporteure

GOMEZ Alice, MC, Université Lyon 1

Examinatrice

LUTZ Antoine, DR, Inserm

Directeur de thèse

PAGNONI Giuseppe, Prof. Assoc., Université de Modène & Regio Emilia

Co-directeur de thèse

MOUNIER Hugues, Prof, Centrale Supélec

Invité

It took years of development and a whole new cosmology for people to accept the insights of Galileo, Newton and others — to really take in the idea that the Earth is not the center of the universe. Similarly, the idea that “I” am not the center of the universe demands an equally challenging — even troubling — paradigm shift.

Gregory Kramer in *Insight Dialog, The Interpersonal Path to Freedom (2007)*.

Il a fallu des années de développement et une cosmologie entièrement nouvelle pour que les idées de Galilée, de Newton et d'autres soient acceptées — pour vraiment assimiler l'idée que la Terre n'est pas le centre de l'univers. De même, l'idée que “je” ne suis pas le centre de l'univers exige un changement de paradigme tout aussi difficile — voire troublant.

Gregory Kramer dans *Le dialogue conscient : le chemin interpersonnel vers la liberté (2007)*, traduction personnelle.

Remerciements

I would first like to thank the members of my thesis jury for accepting to read and comment my work, I hope it will fulfill your scientific curiosity.

Then my deep acknowledgments go to my two PhD directors. I would like to thank Antoine Lutz for accepting me in the “meditation” team and giving me the opportunity to work on a fantastic project. I am not only grateful for all the valuable scientific teaching and feedback, but also for encouraging me to develop my “contemplative side” through reads, retreats and advice. I would then like to express my gratitude to Giuseppe Pagnoni for accepting to take part in this long and complex journey as co-director. Your sharp and critical insights always pushed us in the good direction and you definitely helped me improve my understanding of Active Inference to higher limits.

I would also like to thank the Weissman Foundation who generously funded the last year of my PhD.

Merci à Eric Koun, Yves Rossetti, Alexandre Foncelle de l'équipe Trajectoires et Roberto Vargiolu et Lucas Ouillon de Centrale pour leur aide précieuse dans la conception et l'analyse de la tâche de perception tactile. Merci aussi à Franck Lambertson du Cermep pour ta flexibilité qui a permis les enregistrements IRM dans des conditions intenses.

Merci à Stephane Offort pour avoir accepté d'être notre enseignant pour les retraites de méditation organisées dans le cadre de ce doctorat. Ton apport théorique a été précieux pour affiner de nombreux points scientifiques et à un niveau plus personnel ton enseignement pendant les retraites m'a beaucoup apporté. Je voudrais en profiter pour remercier toutes les personnes qui m'ont inspiré dans ma pratique méditative, les grands maîtres que j'ai eu la chance d'écouter et lire bien-sûr, les compagnons de “coussin”, mais aussi toutes ces personnes rencontrées au quotidien qui ne le savent peut-être pas mais qui m'ont aidé sur mon chemin vers moins de “souffrance”. Les participants à notre retraite “scientifique” de méditation en font partie, j'ai été très touché par la reconnaissance dont vous avez fait part, j'espère que cette expérience aura porté ses fruits pour vous aussi et vous souhaite tout le meilleur pour la suite. Merci aussi à toutes les personnes qui gèrent le magnifique centre Izariat et qui nous ont si bien accueilli pendant les retraites, quelle cuisine... gardez cette belle énergie !

Je tiens à remercier très chaleureusement toute “l'ancienne” équipe Méditation: Enrico, Kristien, Prisca, Jelle, Sucharit et d'autres, pour l'accueil formidable que j'ai reçu de leur part. J'ai été ravi de partager toutes ces discussions scientifiques, moments de méditation et rires et joies avec vous. Un merci tout particulier à Oussama et Stephano qui sont devenus de grands amis. Votre aide aura été inestimable dans mon travail de thèse mais aussi et surtout en dehors, merci pour les moments d'écoute, de joies, de rires, les vacances, l'escalade, la course ou encore les matchs de ping-pong endiablés.

Merci à la “team Longimed”, Kacylia, Thalia, Malo, Igor, Sébastien, sans qui le projet principal de ce

doctorat n'aurait jamais vu le jour, et avec qui nous avons franchi tous les obstacles: une pandémie, trois confinements, la neige et autres inattendus. Mes chers stagiaires, Gansheng, Malo, Thalia j'espère ne pas vous avoir trop mal encadré et que vous aurez tiré quelque chose de positif de ces stages, votre aide m'aura en tout cas été précieuse.

Merci aux autres collègues et ami.es d'Eduwell, Pam et Cophy pour l'aide ponctuelle que vous avez pu m'apporter et les discussions scientifiques. Merci aussi pour l'autre côté de la vie de laboratoire qui est tout aussi important: les matchs de ping-pongs, les discussions de bureau, les repas partagés, les apéros et autres. Merci à Martine pour toute ton aide administrative, je crois bien que je serais ton dernier doctorant... et merci à Gaëlle pour ton aide sur la fin, je crois bien que je serais ton premier doctorant ! Merci Ludo pour ton assiduité sans faille pour les séances de méditation.

Je tiens à remercier très chaleureusement tous et toutes les ami.es dont la joie de vivre et l'écoute m'ont soutenu et motivé pendant ces 4 ans ! Une pensée particulière à mes chers colocataires qui ont eu à supporter mon stress des derniers mois de rédaction, je vous le rendrai mille fois ! Merci aussi à tous ceux avec qui j'ai partagé des moments d'escalade, ce sport ou plutôt cette passion m'aura aussi énormément appris en parallèle de la thèse. Merci aussi à ceux que je n'ai pas cité, vous aurez vos remerciements "personnalisés" au moment opportun.

Un dernier mot pour ma famille, merci pour votre soutien et les moments de détente et de ressourcement qui m'ont aussi beaucoup aidé, j'espère vous rendre fier et vous souhaite beaucoup de bonheur.

Merci à tous et toutes, puissiez vous être heureux et le reste du monde avec vous.

Résumé substantiel

L'étude de la méditation comme outil scientifique pour comprendre « l'inscription corporelle de l'esprit » a été initialement proposée par le neuroscientifique Francisco Varela et ses collaborateurs dans un ouvrage pionnier en 1991. Trente ans plus tard, un nombre croissant de preuves scientifiques confirme que la pratique de la méditation peut avoir un impact bénéfique sur le corps et l'esprit. Pourtant, et malgré des modèles physiologiques et psychologiques prometteurs, une compréhension générale des mécanismes associés aux pratiques méditatives, ancrée dans les neurosciences, fait toujours défaut. En parallèle, la caractérisation de la cognition humaine par la théorie du *traitement prédictif* (*predictive processing*) et son implémentation cérébrale, laisse entrevoir une explication unifiée de processus aussi divers que la perception, l'action, l'attention et l'apprentissage. Basée sur des principes de statistiques Bayésiennes, cette théorie envisage le cerveau comme un « organe d'inférence » qui à la fois prédit et contraint, de manière proactive, les sensations que l'organisme reçoit de son corps et du monde extérieur avec pour impératif de maintenir les constantes vitales propres à chaque être vivant. L'objectif principal de ce doctorat a été d'élucider certains des principes neuronaux et computationnels de différentes pratiques méditatives en utilisant ce cadre théorique. Notre hypothèse générale est que la régulation par la méditation de l'attention et des émotions, est associée à un ajustement des mécanismes prédictifs du cerveau. Entre autres, les degrés de confiance dans la validité des prédictions et des sensations seraient différemment altérés, aboutissant à des a priori cognitifs plus malléables et adaptatifs. Cette transformation de la façon d'appréhender aussi bien l'expérience mentale que le monde extérieur, pourrait expliquer les effets psychothérapeutiques démontrés de la méditation de pleine conscience pour faire face à la rechute en dépression, à l'anxiété, aux douleurs chroniques ou encore aux addictions.

Pour tester cette hypothèse, nous avons mené plusieurs études expérimentales. Dans un premier projet (Etude 1) nous avons combiné l'électroencéphalographie (EEG) avec un paradigme de déviance auditive permettant d'étudier la *négativité de discordance* (*mismatch negativity, MMN*). Ce potentiel électrique généré par le cortex cérébral peut-être observé lors de l'interruption d'une séquence de sons similaires et donc prédictibles, par un son déviant générant alors une *erreur de prédiction*. Contrairement à de précédents résultats, nous avons montré que ce corrélat cérébral du *traitement prédictif* n'était pas affecté par deux styles de méditation différents (l'attention focalisée et la conscience ouverte) ni par leur degré de maîtrise (méditants experts ou novices). Pour comprendre ce résultat négatif, nous avons ensuite commencé par analyser la condition contrôle, non méditative, par une approche de modélisation (Etude 2). Nous avons comparé les principales hypothèses présentes dans la littérature concernant le mécanisme génératif de la MMN, à savoir une *adaptation neuronale* différentielle et un *apprentissage Bayésien* basé sur la théorie du *traitement prédictif*. En confrontant les prédictions issues

de différents modèles simulant ces hypothèses avec nos données EEG, nous avons pu montrer que les deux mécanismes supposés (adaptation et apprentissage Bayésien) sont nécessaires pour expliquer la *négativité de discordance (MMN)*. La modélisation des états de méditation sera poursuivie après le travail doctoral.

En parallèle, nous avons adapté une expérience de recherche existante pour explorer l'effet de la méditation sur les mécanismes prédictifs des dimensions sensorielles et affectives de la douleur. Pour cela nous avons mesuré les variations dans sa perception, en termes d'intensité et de caractère désagréable, selon les attentes induites par des indices visuels et leur interaction avec les à-priori cognitifs individuels sur la douleur. Notre hypothèse de travail était que la sensibilité perceptive et les qualités de non-jugement et non-réaction développées lors de la pratique méditative pouvaient influencer les processus prédictifs automatiques à l'œuvre dans la perception douloureuse. Pour explorer cette hypothèse, nous avons recruté des participants ayant une pratique préalable de méditation et les avons soumis à l'expérience mentionnée ci-dessus (ainsi que de nombreuses autres non développées ici) trois fois: avant, pendant et après une retraite de 10 jours de méditation de pleine conscience. Une nouvelle fois, nous avons analysé les données recueillies en modélisant les mécanismes prédictifs liés à l'expérience douloureuse. Dans un premier manuscrit sur la mesure initiale faite avant la retraite de méditation (Etude 3), nous avons pu reproduire des travaux de modélisation existants sur la composante sensorielle de la douleur, et les étendre pour intégrer sa composante affective. Nous avons également identifié des corrélations intéressantes entre certains paramètres de notre modèle et des questionnaires mesurant le *catastrophisme à la douleur* ou au contraire la *prise de distance cognitive*, ainsi que l'expérience méditative des participants. En appliquant le même modèle aux mesures effectuées pendant et après la retraite (Etude 4), nous avons pu montrer pour la première fois qu'une retraite intensive de méditation peut réduire l'influence d'attentes induites expérimentalement sur la perception sensorielle et affective de la douleur.

Dans l'ensemble, ce travail de thèse fournit certaines des premières découvertes directes de la façon dont la méditation pourrait avoir un impact sur le *traitement prédictif* et, ce faisant offre un éclairage mécanistique nouveau sur la manière dont les pratiques de méditation peuvent contribuer à une régulation bénéfique de formes inadaptées ou pathologiques de perception et de cognition.

Mots clés : méditation, pleine conscience, retraite méditative, cognition incarnée, traitement prédictif, codage prédictif, inférence Bayésienne, inférence perceptive, inférence active, électroencéphalographie, potentiels évoqués, négativité de discordance, MMN, douleur, modélisation Bayésienne, modélisation computationnelle, défusion cognitive, catastrophisme douleur.

Abstract

The idea of meditation as a scientific tool to understand the “embodied mind” was initially suggested by neuroscientist Francisco Varela and his collaborators in a ground-breaking publication in 1991. Thirty years later, a growing amount of scientific evidence confirms that meditation can have beneficial impacts on the body and mind. Despite encouraging physiological and psychological models, a general understanding of the mechanisms at play during meditative practices, grounded in neurosciences, is still lacking. Additionally, the comprehension of human cognition based on the *predictive processing theory*, and its cerebral implementation, could offer a unifying explanation for processes as diverse as perception, attention, learning, and action. Founded on Bayesian statistics, this theory models the brain as an “inference organ” which simultaneously predicts and constrains, proactively, the sensations the organism receives from both its own body and the outer world, with the main purpose of maintaining itself in a viable state. The primary objective of this PhD was to elucidate, within this theoretical framework, some of the neuronal and computational mechanisms of different meditative practices. Our general hypothesis is that the regulation of attention and emotions by meditation is associated with an adjustment of the brain’s predictive processes. The degrees of confidence in the validity of predictions and sensations, among other factors, would be differently altered, leading to more malleable and adaptive cognitive priors. This transformation of the way of approaching mental experience as well as external influences, may explain the proven psychotherapeutic effects of mindfulness meditation to cope with depressive relapse, anxiety, chronic pain or addictions.

Keywords: meditation, mindfulness, meditation retreat, predictive processing, predictive coding, Bayesian inference, perceptual inference, active inference, electroencephalography, event-related potentials, mismatch negativity, pain, Bayesian modeling, computational modeling

Laboratory

INSERM U1028–CNRS UMR5292, Université de Lyon 1

Centre de Recherche en Neurosciences de Lyon (CRNL)

Équipe EDUWELL (Co-leaders Dr Antoine Lutz et Dr Jean-Philippe Lachaux)

Centre Hospitalier Le Vinatier (Bât. 452)

95, Bd Pinel

69500 Bron, France

Contact: Martine Theillère

Tél: +33 (0)4 72 13 89 00

Fax: +33 (0)4 72 13 89 01

Courriel: martine.theillere@inserm.fr

The experimental work presented in this thesis has been funded by a European Research Council Consolidator Grant awarded to Antoine Lutz (project BRAIN and MINDFULNESS, number 617739) and by the LABEX CORTEX of Université de Lyon (ANR-11- LABX-0042), within the program “Investissements d’Avenir” (ANR-11- IDEX-0007).

Contributions of the author

Study 1: PhDs Antoine Lutz and Enrico Fucci designed the study, I and PhD Enrico Fucci preprocessed the EEG data. I analyzed the alpha oscillatory data while PhD Enrico Fucci ran the frequentist analysis on the MMN and PhD Oussama Abdoun the Bayesian one. All authors wrote the article and answered the reviewers.

Study 2: The same data as study 1 was used, as well as previous data collected by PhD Antoine Lutz and preprocessed by PhD Enrico Fucci. We designed the computational models with PhD Oussama Abdoun, based on a template from PhD Françoise Lecaigard for the Bayesian learning model. I ran the modeling and simulation analyses, while PhD Oussama Abdoun performed the model comparisons and subsequent analyses. We wrote a first draft of the article with PhD Oussama Abdoun and all co-authors provided feedback.

Longimed study: In collaboration with PhDs Antoine Lutz, Giuseppe Pagnoni and Oussama Abdoun, as well as valuable advice from meditation teacher Stephane Offort, I designed the Longimed study consisting in a controlled and randomized longitudinal intervention over three time points and two groups of participants. I notably wrote the protocol study for ethical approval, designed the three on-site experiments (with valuable help from MSc Eric Koun for device part of the Force-Matching task), directed research assistant Kacylia Pistoia for the implementation of the psychometric questionnaires on Limesurvey for online use and the recruitment of the 54 participants (with help from Stephane Offort). I organized the logistics and planning for three 10-day intensive meditation retreats with the help of Kacylia Pistoia, Stephane Offort, and master students Thalia Doumengine and Malo Renaud d'Alambra. We collected data over three time points, three experiments and 54 subjects with Kacylia Pistoia, Thalia Doumengine and Malo Renaud d'Alambra and help from PhD Antoine Lutz.

Study 3: I formulated the hypotheses, designed the experimental task and analyzed the data with advice from PhDs Giuseppe Pagnoni and Antoine Lutz. I wrote the first draft of the manuscript which received edits and feedback from PhDs Giuseppe Pagnoni and Antoine Lutz.

Study 4: I formulated the hypotheses, designed the experimental task and analyzed the data with advice from PhDs Giuseppe Pagnoni and Antoine Lutz. I will write a first draft of the manuscript.

Contents

Remerciements	5
Résumé substantiel	7
Abstract	9
Contributions of the author	13
List of abbreviations	19
Preamble	21
I Theoretical background	23
1 Historical and scientific perspectives on meditation	25
1.1 Meditation(s): history and modern transformations	26
1.1.1 A constellation of ancient spiritual and philosophical traditions	27
1.1.2 The mindfulness “boom”	31
1.2 Scientific studies of contemplative practices	35
1.2.1 A recent but rapidly growing research field	36
1.2.2 The need for a scientific classification of meditative practices	38

1.2.3	The effects of attentional training	41
1.2.4	Emotional regulation: the example of pain	43
1.2.5	Hindrances facing contemplative research	52
2	Perception as hierarchical Bayesian inference in the brain	59
2.1	A Bayesian framework for perception, learning, action and much more	60
2.1.1	Bayes' rule, Bayesian inference and the Bayesian brain hypothesis	60
2.1.2	The Free-Energy Principle	63
2.1.3	A neuronal implementation of hierarchical Bayesian Inference : predictive coding	65
2.1.4	Empirical evidence for predictive coding in the brain and criticism	70
2.2	Auditory deviance detection revisited through predictive coding	75
2.2.1	The Mismatch Negativity (MMN): a neural correlate of prediction error ?	76
2.2.2	The effect of attention on the MMN	80
2.3	Does pain perception follow Bayesian principles ?	82
2.3.1	Bayesian accounts of pain perception	82
2.3.2	Effects of expectations and uncertainty	86
3	A predictive processing view of contemplative experience	91
3.1	Explaining the meditative brain (and body) through perceptual, interoceptive and active (Bayesian) inference	92
3.1.1	Preliminary proposals	92
3.1.2	An integrative hypothetical account of the predictive brain under meditation .	95
3.2	From theoretical to active "scientific" inference: predicting and testing the effects of meditative practice on predictive processing	107

3.2.1	Heightened sensory precision and decreased prior formation? Insights from the study of auditory deviance detection under meditation	107
3.2.2	Seeing prior expectations as impermanent: pain processing under meditation	111
3.3	Studying the acute and lasting effects of intensive practice: insights from a 10-day mindfulness meditation retreat	116
3.3.1	Scientific studies of residential meditation retreats	116
3.3.2	The LONGIMED study	120
II	Experimental work	129
4	Predictive coding and meditation in auditory perception	131
4.1	Study 1: No effect of focused attention and open monitoring meditation on EEG auditory mismatch negativity in expert and novice practitioners	132
4.2	Study 2: Time-resolved dynamic computational modeling of human EEG recordings reveals gradients of generative mechanisms for the MMN response	144
5	A Bayesian model of pain perception and its modulation by an intensive 10-day meditation retreat	205
5.1	Study 3: Modulation of the sensory and affective dimensions of pain by expectations and uncertainty: a Bayesian modeling approach	206
5.2	Study 4: Effects of a 10-day mindfulness meditation retreat on pain perception according to a Bayesian framework	252
5.2.1	Rationale	252
5.2.2	Methods	252
5.2.3	Results	255
5.2.4	Discussion	258

III	General discussion	263
6	Discussion and perspectives	265
6.1	Can meditation affect low-level predictive processing ?	265
6.2	Preliminary evidence for the regulation of pain predictive processing by meditation . .	267
6.3	Future directions	272
6.4	Conclusion	273
	Bibliography	276

List of abbreviations

- ACC** Anterior Cingulate Cortex. 22
- AEs** Adverse Effects. 37
- BCE** Before Common Era. 6
- CE** Common Era. 7
- DDS** Drexel Defusion Scale. 31
- dIPFC** dorsolateral PreFrontal Cortex. 27
- DMN** Default Mode Network. 22
- EEG** Electroencephalography. 35
- ERPs** Event Related Potentials. 35
- FA** Focused Attention. 19
- FFMQ** Five Facets Mindfulness Questionnaire. 31
- fMRI** functional Magnetic Resonance Imaging. 26
- IASP** International Association for the Study of Pain. 25
- ITT** intent-to-treat. 35
- MBCT** Mindfulness-Based Cognitive Therapy. 15
- MBI** Mindfulness-Based Intervention. 15
- MBSR** Mindfulness-Based Stress Reduction. 11
- MCC** Mid-Cingulate Cortex. 26

MMN Mismatch Negativity. 35

mPFC medial PreFrontal Cortex. 24

OFC OrbitoFrontal Cortex. 27

OM Open Monitoring. 19

PCC Posterior Cingulate Cortex. 24

PCS Pain Catastrophizing Scale. 31

PFC PreFrontal Cortex. 22

pgACC perigenual Anterior Cingulate Cortex. 27

PPC Posterior Parietal Cortex. 27

rACC rostral Anterior Cingulate Cortex. 27

RCTs Randomized Controlled Trials. 39

SI primary sensory cortex. 26

SII secondary somatosensory cortex. 26

SMA Supplementary Motor Area. 26

vIPFC ventrolateral PreFrontal Cortex. 22

Preamble

The first part of this thesis will present the necessary theoretical background to understand and critically apprehend our experimental work. In Chapter 1, I will first give an historical account of meditation emergence and development (Section 1.1.1) and describe the transformation of mindfulness meditation in the modern world (Section 1.1.2). We will then see how psychology, psychotherapy and cognitive neurosciences got progressively interested into meditation in the last 50 years (1.2.1) and the associated need for a scientific classification of contemplative practices (1.2.2). Primary scientific findings regarding attentional (1.2.3) and emotional (1.2.4) regulation effects, as well as hindrances arising from the scientific study of meditation (1.2.5) will be reviewed.

In Chapter 2, I will present the recently developed *predictive processing* theory of perception and cognition. To do so I will first briefly describe Bayesian concepts (2.1.1) and the *Free Energy Principle* from which it derives (2.1.2) and then get to the heart of the matter by detailing its hypothesized neuronal implementation: predictive coding (2.1.3), as well as the empirical evidence supporting it so far (2.1.4). We will then address the perceptual processes of interest for the current PhD: auditory deviance detection (2.2.1) and pain perception (2.3.1), detailing for both the predictive coding (or processing) mechanisms at play and how they may be modulated by attention for deviance detection (2.2.2) and expectations or uncertainty for pain (2.3.2).

In a final introductory and methodological chapter (Chapter 3) I will try to bind together what we saw previously by reviewing recent theoretical proposals of the interplay between meditation and predictive processing (3.1.1) and drawing mechanistic hypotheses of how meditation may impact our perception and cognition under this framework (3.1.2). In a second time I will cast more specific predictions on the influence of contemplative practice on auditory deviance detection (3.2.1) and pain perception (3.2.2), and present the experimental paradigms we used to test them. Finally, I will describe the primary experimental project of this PhD: the LONGIMED study, a controlled and randomized longitudinal investigation of the effects of an intensive 10-day mindfulness meditation retreat on diverse perceptive and cognitive processes at the phenomenological, behavioral and neuronal levels (3.2.3).

In a second part I will present my experimental work with a first article published in second author on the effect of meditation on auditory deviance detection as tested with the mismatch negativity (MMN, 4.1), and a second article submitted as first author on the computational mechanisms of the MMN (4.2). I will then report the first results of the LONGIMED study through a third (draft) article as first author on Bayesian models of pain perception using the baseline data from this study (4.3), followed by preliminary longitudinal results of the effects of an intensive meditation retreat on pain (4.4).

In a third and last part I will discuss my experimental findings in regard to the existing literature, address their limitations and envision potential perspectives arising from my research work.

Part I

Theoretical background

Chapter 1

Historical and scientific perspectives on meditation

1.1 Meditation(s): history and modern transformations

The point of mindfulness/awareness is not to disengage the mind from the phenomenal world; it is to enable the mind to be fully present in the world. The goal is not to avoid action but to be fully present in one's action, so that one's behavior becomes progressively more responsive and aware.

Francisco Varela, Evan Thompson and Eleanor Rosch in
The Embodied Mind: Cognitive Science and Human Experience (1991)

As Varela, Thompson and Rosch pointed out in their ground-breaking work *The Embodied Mind* [Varela et al., 1992], meditation is not about cutting oneself off from the rest of the world and “do nothing” or stop any thoughts, emotions or sensations. These common misunderstandings still cling to this concept describing millennia-old practices and conceal their spiritual and philosophical richness. At the exact opposite of such misconceptions, and as we shall see throughout this doctoral thesis, meditating is to open widely to the world and be fully present, here and now, to one's body and mind. It is not about blocking all mental activity or resting in a totally passive attitude but rather about actively becoming aware of sensations, thoughts, emotions, fears, joys, desires or any other conscious experience that may arise. In order to have a clearer and fact-based understanding of what is meditation and what it is not, we will engage in this first section in a journey from its origins in ancient India to its latest development in the modern world. First, we will show that there is not one but many different types of meditations, by giving a brief historical account of the numerous spiritual and philosophical traditions which developed mostly in East Asia and used so-called contemplative practices. Most of these traditions perpetuated until today in the Eastern world, from which they finally spread to the West during the last 70 years. The encounter between various meditative practices and Western medicine and science gave rise to a secular technique which thrived to the point of almost becoming a fad: mindfulness meditation. I will try to trace back the historical roots of mindfulness in ancient traditions and describe how it was adapted to occidental clinical settings. As a conclusion to this first introductory part I will discuss the caveats that mindfulness meditation is facing as a side effect of its popularity, emphasizing the need for a careful and scientific inquiry of its effects and potential benefits.

1.1.1 A constellation of ancient spiritual and philosophical traditions

Historically meditation is often thought to be born with Buddhism and his spiritual teacher Siddhārtha Gautama, also known as the historical Buddha, who presumably lived from 563 to 483 Before Common Era (BCE). However the oldest traces of meditation can be found on clay seals crafted by an ancient civilization of the Indus valley, the Harappan Civilization who thrived through 3500 to 1700 BCE. One of those seals represents a deity sitting in a postulated yogic-like meditative posture (Figure 1.1). If the interpretation of these relics is debated [Samuel, 2008], written evidence of the precedence of yogic meditation over Buddhism can be found in the Early Upanishads, one of the principal Vedic (pre-Hinduist) scriptures, approximately dated between the 10th and 6th century BCE [Olivelle, 1998]. Those texts present different yogic techniques such as pranayamas (breath control) or mantra repetition. Besides we know from historians of Siddhārtha Gautama that he was first trained in meditation by two yogis [Armstrong, 2004], indicating that contemplative practices were already cultivated before his teachings.



Figure 1.1: Seal from Mohenjo-Daro representing a deity seated in yogic meditative posture.
Source: J.M. Kenoyer/Harappa.com, Courtesy, Dept. of Archeology and Museums, Govt. of Pakistan.

Yet, he was indeed the driving force in developing a truly formed meditation technique and most importantly in disseminating it to lay people in North India at that time. By teaching meditation not only to monks and nuns but also to lay people, he started the democratization of contemplative practices which were before reserved to few ascetics. I would like to highlight the fact, often forgotten, that if Buddhism can be classified as a religion, it is also and before all a spiritual, experiential and philosophical tradition. It is clear from the transcriptions we have of his discourses, that Gautama never encouraged any devoted cult to either himself or any gods, but instead promoted free, critical

and renewed thinking [Armstrong, 2004].

After his death, divergences in early Buddhist schools concerning philosophical interpretations and acceptance of precepts taught by the Buddha lead to a schism between two main schools: the Sthaviravada and the Mahasanghikas, after the second Buddhist Council held during the 4th century BCE. The Sthaviravada or Theravada in Pāli (“School of the Elders”) expanded during the 3rd century BCE under the reign of the Indian emperor King Ashoka when it was established as the main Buddhist doctrine in the Mauryan empire and spread throughout India and Sri Lanka [Conze, 2007]. The main principles of Theravada meditation practices come from their version of the Tripiṭaka (“Three baskets”) Buddhist sacred scriptures : the Pali Canon. The Tripiṭaka is the first written account of the Buddha teachings, collected and codified during the fourth Buddhist Council around the 1st century BCE, and is the basis of all major Buddhist schools. The particularity of the Theravada tradition is to rely almost exclusively on these early scriptures for guidance of their practice, and to reject any subsequent elaborations. As can be seen from Figure 1.2, it spread over South-East Asia during the last two thousand years, and is now considered one of the three main Buddhist traditions, predominant in Cambodia, Laos, Myanmar, Sri Lanka, and Thailand. The last two centuries saw a revival of Theravada and its extension to the Western world through modern Buddhist movements. For example the Vipassana movement started at the beginning of the 19th century in Myanmar gave rise to the worldwide practice of Vipassana or Insight meditation, whereas Thailand saw the renewed emergence of strict monastic life with the Thai Forest tradition [Conze, 2007].

Following the fourth century BCE schism, the Mahasanghikas school progressively gave rise to the second main school of Buddhism: Mahayana (“The Great Vehicle”), which started to be established under King Kanishka (2nd-1st centuries BCE) whose empire encompassed northern India and large parts of Central Asia. It really became influential only later on, starting around the 5th century Common Era (CE) when it spread to China through the Silk Road, giving rise to Chan Buddhism, which in turn laid the foundations of Thien in Vietnam, Seon in Korea and Zen in Japan. Like Theravadins, Mahayanists relied on the Buddha’s direct teachings presented in the Tripitaka, but they also allowed the developments of new sutras and philosophical treatises inspired by renowned meditation masters. They notably impeded a major transformation of the Buddhist concept of enlightenment. Whereas the Theravada view of meditation was quite personal, with the aim of liberating oneself from suffering, the Mahayana looked for their own liberation, only to help liberate all other sentient beings. Doing so they opened up the previously secluded vision of meditation to include any day-to-day life situations. A very deep and important treatise in the Mahayana doctrine is the Root Stanzas of the Middle Way from Nagajurna, an Indian philosopher who lived from 150 to 250 CE. His work led to the foundation of the Madhyamaka (Middle way) thinking school which was later embraced by Tibetan Buddhism [Kalupahana, 1986]. The third main Buddhist school, Vajrayana (“Diamond Vehicle”), also known as

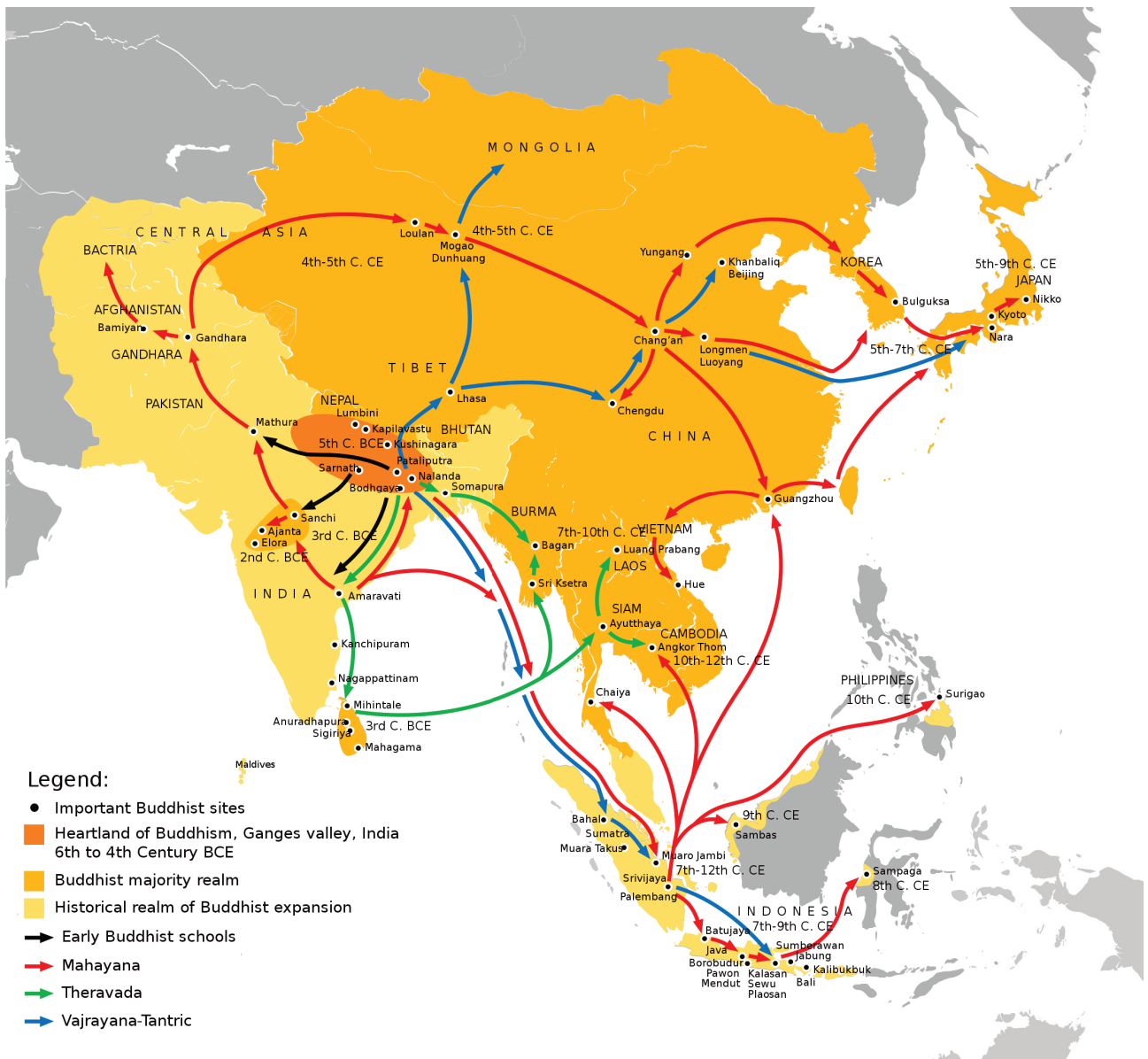


Figure 1.2: Expansion of Buddhism from India in the 5th century BCE to the rest of Asia. *Source: Gunawan Kartapranata, CC-BY-SA-3.0 license.*

Tantric Buddhism, emerged from Mahayana around the 5th century CE. Its adepts base their meditation on Tantras scriptures, using techniques found in Mahayana but also new ones such as mantras or visualizations (e.g. deity yoga), considered to provide a faster way toward enlightenment. This doctrine was imported, preserved and developed in Tibet from the 7th century CE onward, and is now known as Tibetan Buddhism [Conze, 2007]. It is also predominant in neighboring countries such as Mongolia, Nepal, Bhutan and northern Indian States. The Tibetan diaspora following the Chinese invasion in 1959 led to the worldwide propagation of Tibetan Buddhism thanks to charismatic leaders such as the 14th Dalai Lama and other Tibetan teachers.

The three traditions of Theravada, Mahayana and Vajrayana are predominant in contemporary Buddhism. However they are composed of many subschools and countless other Buddhist schools exist,

each with their own particularities depending on historical, philosophical and sociological context. Besides, if Buddhism cultivated the practice of meditation to an art of living, other spiritual and philosophical traditions also used contemplative techniques for personal, religious or philosophical flourishing. The most famous nowadays in the occidental world, although its modern practice diverged greatly from its origin, is indubitably yoga. As we previously evoked, Hindu yoga practices and treatises preceded the development of Buddhism and influenced later traditions such as tantric practices in Vajrayana. In its original form, as described for example in the Yoga Sutras of Patanjali, yoga is at the same time a philosophy, a moral doctrine, and a set of breathing practices, physical postures, and concentration techniques aiming at the cessation of mental fluctuations [Bryant, 2015]. Another tradition perpetuating some form of instructed meditation practices is Jainism, an Indian religion sharing its cosmology with Hinduism and Buddhism.

While India was the cradle of many contemplative traditions, its close neighbor China also saw the development of influential philosophical movements inclined to contemplation such as Confucianism and Taoism, whose founding text the Tao Te Ching was written contemporary to the Buddha's teachings around the 6th century BCE by Laozi. In the same period, called by philosophy historians the Axial Age [Jaspers, 1953] for its striking philosophical and spiritual turmoil throughout the world, the European continent also witnessed the appearance of philosophical contemplation in Greece. Pythagoras is well-known for his mathematical discoveries, yet he was also the founder of an ascetic school of thought which later influenced Plato, Aristotle and other Greek philosophers, whose works set the basis for western philosophy. Actually some scholars advanced that early Buddhism had a direct influence on Greek philosophy through Pyrrho who travelled with Alexander the Great during his Asian conquest up to India around 327-325 BCE [Beckwith, 2015]. Pyrrho was most likely exposed to Indian and Buddhist philosophy there and later founded Pyrrhonism, the earliest school of Greek philosophical skepticism. He notably created the concept of *epoché*, or "suspension of judgement", later embraced by Husserl in the 20th century CE to theorize his philosophical method for the study of conscious experience: phenomenology, which still has tremendous influence on modern philosophy and cognitive science [Husserl, 1931]. Finally for the sake of completeness, we may precise that the three main monotheists religions also have their own contemplative practices, often referred to as "mysticism". We can quote for example "letter permutation" in Judaism, "divine light" contemplation in Christianity and Sufi whirling in Islam.

Putting back meditation into its historical context clearly shows us the inadequacy of trying to give a single and unitary definition of such a varied and distributed concept. Nevertheless contemplative practices, despite forming a broad constellation, also share some commonalities. First, and in opposition to the naive understanding of meditation as a passive task, they all involve some form of mental regulation and training, mainly in ways of redirecting one's attention to specific external or

internal objects such as sensations, emotions or thoughts. To do so a specific posture is often adopted, facilitating stillness (or at least controlled movement) in order to calm the mind and the body, such as sitting, kneeling or laying down. Finally, contemplative practices have the common goal to foster some transformation in the way we apprehend ourselves, others and our world, be it in terms of personal well-being, spiritual flourishing or existential meaning. This general definition is still vast and subject to controversy, therefore when scientifically studying meditation it is important to define which kind of meditation we are talking about. Thereafter I will describe in more detail a recently emerging collection of meditation practices gathered under the umbrella term of “mindfulness”, which was largely used in scientific and clinical research.

1.1.2 The mindfulness “boom”

The coming of Buddhism to the West may well prove to be the most important event in the twentieth century.

Arnold Toynbee, author of *A Study of History*

Modern or secular mindfulness practices have a complex history embedded in the modernization of traditional Buddhist contemplative practices and the growing interest of the West in “Eastern spirituality”. In response to British colonization in Burma (now Myanmar), Theravada Buddhist monks from the late nineteenth century sought to revitalize traditional practices by developing new meditations that avoided explicit religious content and thus were more accessible to lay people. Promoting the direct practice of insight meditation (Vipassana), Ledi Sayadaw, Mahasi Sayadaw and others launched the modern Vipassana movement which had a profound influence in the spread of Asian meditation practices to the Western world [Braun, 2013]. The twentieth century also saw the spread of other traditions to the West, such as Zen and Tibetan Buddhism. This dissemination process was greatly helped by charismatic spiritual teachers such as S.N. Goenka, Thich Nhat Hanh, Shunryu Suzuki or Chögyam Trungpa who recognized early on the Western demand for spirituality, but also the crucial need to accommodate traditional meditation practices to the Westerners’ way of life. Teachings, books, retreat centers opening and the formation of Western teachers led to a most fertile ground for the spread of meditation practices in Europe and America in the second half of the twentieth century.

A turning point was reached in the U.S. in the seventies, when Jon Kabata-Zinn, a medical doctor with a PhD in molecular biology and a personal experience in yoga, Zen and Vipassana, set out to apply the knowledge he gained through these practices to common illnesses. In 1979, Kabat-Zinn developed at the University of Massachusetts Medical School, a 10-week clinical program initially dedicated to

treat the psychological dimensions of chronic pain disorder, through the use of combined meditation and yoga techniques from various traditions (mainly Theravada and Zen Buddhism, as well as contemporary Hatha yoga [Kabat-Zinn, 1982]). Importantly, Kabat-Zinn removed any religious terms or cultural references to their Eastern origins in his instructions, emphasizing his program as secular and scientifically plausible. The program was later shortened to 8-week and renamed Mindfulness-Based Stress Reduction (MBSR) [Kabat-Zinn, 1991]. Subsequently, several studies supported its usefulness in a large range of pathologies from depression to eating disorders, and in synergy with a renewed spiritual interest, led to its global spread through the U.S. and later worldwide. If Kabat-Zinn’s definition of mindfulness as: “the awareness that arises through paying attention, on purpose, in the present moment and non-judgmentally to the unfolding of experience moment by moment” [Kabat-Zinn, 1991], is most known nowadays, semantic ambiguity concerning the historical use of the word and the precise concepts it conveys are still subject to debate among Buddhist scholars [Anālayo, 2019a, Bodhi, 2011, Dreyfus, 2011, Dunne, 2011a, Gethin, 2011]. Indeed, as remarked by some, the Pali term for mindfulness, “sati”, originally meant memory or recollection. Others argue that the Buddha’s teachings transform the use of the word, making it more aligned to Kabat-Zinn’s one, as illustrated by the following extract from the Pali Canon:

And what, monks, is right mindfulness? Here, a monk dwells contemplating the body in the body, ardent, clearly comprehending, mindful, having removed covetousness and displeasure in regard to the world. He dwells contemplating feelings in feelings . . . contemplating mind in mind . . . contemplating phenomena in phenomena, ardent, clearly comprehending, mindful, having removed covetousness and displeasure in regard to the world. This is called right mindfulness. [Bodhi, 2011]

The definition of mindfulness is an important step for its scientific study [Van Dam et al., 2018, Lutz et al., 2015], as the term has been often associated with anything and everything in the scientific vulgarization literature and the news media. Setting aside theoretical disagreement for the moment (but we will come back to it later on, see Section 1.2.5), we can define mindfulness practices as encompassing qualities of attention and awareness, as well as discernment, acceptance and non-judgment, within the present moment ([Van Dam et al., 2018, Bishop et al., 2004] but see [Dreyfus, 2011, Bodhi, 2011, Gethin, 2011] for critical perspectives on the non-judgmental and present centered aspects of mindfulness). Importantly, most Buddhist scholars agree that despite the biases inherent to the use of mindfulness meditation in a secular fashion uprooted from its traditional and cultural milieu, the benefits it can bring to the entire society is very promising as long as it remains faithful to its initial purpose of reducing human suffering [Anālayo, 2019b, Bodhi, 2011]. Doug Oman articulated well the need for Western mindfulness teachers to shift from the traditional spiritual goal of “enlightenment”

and alleviating suffering for all beings to a more proximate goal of developing mindfulness, in order to facilitate the secular needs of individual well-being [Oman, 2021]. More critically, Walsh and Shapiro, pioneers in meditation research, talk about this contemporary transformation of mindfulness as a “re-visioning of the practices within an exclusively Western psychological and philosophical framework” [Walsh and Shapiro, 2006]. Not honoring ancient spiritual traditions from which they emerged, such “re-contextualization” (Ibid.) might prove dangerous for modern mindfulness itself.

Meditation techniques are often taken for granted, forgetting the fact that they were developed and transmitted across centuries, resisting wars, schisms, cultural assimilation, thanks to dedicated practitioners and masters who strove ultimately for liberation from suffering without concern for worldly benefits. Remembering this historical debt and invaluable heritage sets a critical perspective on recent developments of meditation as another profitable commodity. As a matter of fact, the twenty-first century saw the exponential development of mindfulness meditation supports such as books, documentaries, podcasts, magazines, smartphone applications, etc... Some of these can be valuable and have sprung from a genuine intention to help people in their quest for meaning, spirituality or well-being. However, scholars noticed that mindfulness has also been appropriated by the modern capitalist economy [Wilson, 2014, Purser et al., 2016, Titmuss, 2016]. Wholesome initiatives do exist, for example to make mindfulness widely accessible or improve the quality of work life, but purely profit oriented businesses on mindfulness are not that rare anymore. Coaching services promise individual customers to “boost” their attentional abilities and make them smarter. Online platforms sell “workplace mindfulness” programs to large companies as a “mental technology” to increase productivity and save healthcare costs, behind the facade argument of reducing employee stress [Wilson, 2016]. Start-ups trade high-tech products and software supposed to make their owners more aware and present to the moment, while some famous brands qualify goods such as jewels, clothes and gadgets as “mindful”, apparently unfazed by the absurdity of such assertion [Wilson, 2016]. This neoliberal development coined “McMindfulness” or “corporate mindfulness” underwent severe criticism in the last decade [Purser et al., 2016, Purser, 2018]. Although such misuses of mindfulness can seem anecdotal, they do influence the current view of meditation in the general population and may contribute to foster harmful behaviors which meditation actually aims to soothe. As Jeff Wilson wrote, the capitalistic uptake on mindfulness contributes to a “[...] heightened competition around mindfulness: Because money can be made off mindfulness, it invites competition and consumption [...]” [Wilson, 2016], values completely at odds with genuine mindfulness teaching and the traditions it comes from. As was ironically coined by an eminent Buddhist monk and scholar “mindfulness has become a handy buzzword that can be attached at random to virtually any product or skill in order to invest it with a spiritual aura or increase its market appeal: mindful romance, mindful birthing [...] Perhaps, just over the horizon, we will find some entrepreneurs pushing mindful mindfulness.” [Bodhi, 2016] Another risk of the mindfulness hype

is to confuse the means given by meditation practice (attentional training, stress reduction, emotional regulation, etc. . .) for an end in itself. When the fundamental aim to reduce one's own and others' suffering (in its broadest meaning) is forgotten, mindfulness can just become another tool to promote the cultural and moral imperative for self-improvement [Payne, 2016, Bodhi, 2016]. At the opposite, traditional teachings emphasize the impermanence of all things, particularly of our "Self", to which we cling even more when we are drawn into the compulsive business of self-improvement. Meditation should be a remedy, not a catalyst, to the modern attentional overload.

To summarize, we saw how modern mindfulness meditation are rooted in both Eastern philosophical traditions and Western medicine and science. This marriage of convenience planted the seeds for a worldwide dissemination of meditation while thoroughly transforming it in the process, a step deemed necessary to adapt it to Western sensibilities. The encounter with neoliberal economy also raised a new challenge for those two thousand years old practices: how to resist the capitalist incentives, at odds with the ethical and moral values presented in the traditional teachings ? As has always been the case through their history, contemplative traditions need to adapt to a constantly changing world. Entering a discussion with cognitive sciences, their alter ego to the study of the mind in the West, may well be a promising avenue. Scientific investigation could indeed shed a new light on meditative practices and teachings. Deeper comprehension of the mechanisms at stake during meditation might help refine the techniques and adapt it to the fast transformations undergoing our societies. In the process, science may confirm or contradict traditional claims about mind and body function. The Dalai Lama himself said in the columns of the New York Times : "If science proves some belief of Buddhism wrong, then Buddhism will have to change. In my view, science and Buddhism share a search for the truth and for understanding reality. By learning from science about aspects of reality where its understanding may be more advanced, I believe that Buddhism enriches its own worldview." [Gyatso, 2005]. This is exactly what we are going to engage with now, describing the fruitful encounter between science and Eastern spirituality, its main scientific results and the issues still unresolved.

1.2 Scientific studies of contemplative practices

I believe it important to acknowledge that psychology and Buddhist meditation are different knowledge systems with distinct epistemologies and dissimilar final aims.

Nevertheless, they converge on a keen interest in understanding the workings of the mind with a view to alleviate unnecessary suffering.

Bhikkhu Analayo in [Anālayo, 2019b]

Research on meditation is a recent field drawing from domains as diverse as philosophy, psychology, neurosciences and clinical psychology and psychiatry. Investigation of the potential benefits of meditation in clinical settings was one of the starting research lines, as the mindfulness-based stress reduction (MBSR) program created by Pr. Kabat-Zinn in the late seventies had to be validated empirically. It was first tested on common troubles such as chronic pain, depression and anxiety with promising results [Kabat-Zinn et al., 1985, Kabat-Zinn et al., 1986, Kabat-Zinn et al., 1992, Miller et al., 1995]. This line of research is still active today, with the development of new programs such as Mindfulness-Based Cognitive Therapy (MBCT) and others, gathered under the label Mindfulness-based Interventions (Mindfulness-Based Intervention, [Shapiro et al., 2018]). It was also applied to other pathologies and different populations such as children, teenagers and aging people. The important number of studies now published allow researchers to perform meta-analyses and draw conclusions on the usefulness of meditation as a therapeutic tool. For example mindfulness meditation proves to be helpful in coping with chronic pain [Ball et al., 2017, Hilton et al., 2017, Goldberg et al., 2018, Goldberg et al., 2022b]; depression in adolescents [Reangsing et al., 2021], students [González-Valero et al., 2019] and more generally in the adult population [Goyal et al., 2014, Goldberg et al., 2018, Goldberg et al., 2022b]; and anxiety [González-Valero et al., 2019, Goyal et al., 2014, Liu et al., 2021]. Overall meditation seems to improve psychological well-being, as we could expect from a technique aiming at reducing suffering, but methodological flaws were also identified and will be reviewed in the last part of this chapter.

The question that now arises is: *how* do contemplative practices produce their effects ? Indeed a second major line of research is concerned with the mechanisms at work during meditation. How does it train our attention and redirect it closer to our actual experience and away from anxious thoughts ? Does it rewire our habitual psychological and neuronal patterns to be able to deal with strong emotions ? Can it change the way we perceive ourselves, others and our environment ? The study of psychological and neuronal mechanisms underlying such transformations is a central theme in this PhD. Thus, I will shortly present how science, and more specifically cognitive neurosciences, came to

study meditation practitioners, and the current research efforts in this domain. A summary of major and consistent results on the effects of meditation on attentional and emotional regulation will then be given. Finally we will address the hindrances facing this recent but complex research field and how they could be overcome.

1.2.1 A recent but rapidly growing research field

A complete and thorough history of meditation research is yet to be done as the field is very recent, however Doug Oman chapter in *The Oxford Handbook of Meditation* gives a fair summarized account of its first fifty years [Oman, 2021]. Scientific study of contemplative practices started in the late fifties with preliminary experiments on expert Indian yogis [Bagchi and Wenger, 1957], followed by research on Japanese Zen practitioners [Kasamatsu and Hirai, 1966]. They focused primarily on physiological correlates of meditation states such as changes in blood pressure, respiration and brain rhythms. In the seventies then started a series of research works on Transcendental Meditation (TM), a mantra based meditation recently imported in the United States by Indian yogi Maharishi [Wallace, 1970]. Research on TM transformed the field, with the use of new methodologies, longitudinal, randomized and controlled designs. However the proprietary embargo on details of the technique enforced by the TM organization prevented it from having real scientific significance. Indeed if it showed changes in physiological and electrophysiological measures and proved to be of clinical relevance for certain pathologies its mechanisms of action were never studied in detail up to recently and therefore did not bring much to the scientific understanding of processes at stake during contemplative practices. Unfounded claims about some of TM's effects and the close vicinity of researchers as well as research participants to the TM organization added to the skepticism surrounding TM's research. In opposition to the secluded view of meditation research endorsed by the TM organization, other Western scholars defended an open and inclusive investigation of meditation.

Many contemplative practices were actively studied in laboratories throughout the world at that time and empirical results were compiled by Shapiro and Walsh in an influential book [Shapiro and Walsh, 1984]. Similarly to early works, results on physiological correlates such as breathing rate, heart rate, skin conductance, but also electroencephalographic correlates, were presented. As acknowledged by the community, they still lacked consistency and were mostly inconclusive, emphasizing the need for more numerous and thorough studies of meditation. Yet pioneering works were carried out, for example comparing meditation and sleep or other altered states of consciousness, or evaluating potential adverse effects of meditation, a topic taken up by scientists only recently (see Section 1.2.5). Published at the same time as Kabat-Zinn's first studies on MBSR, the volume also presents preliminary results on therapeutic outcomes of meditation practices, while wittily mentioning the methodological flaws facing

such research —which we should come back to later on in this manuscript. Interestingly the editors, as well as renowned psychologist Daniel Goleman and Buddhist scholar Daniel Brown, emphasized the importance for scientific research of meditation to also look in classical texts of contemplative traditions for phenomenological accounts and philosophical analyses of the practices at stake [Shapiro and Walsh, 1984]. Indeed meditation can be seen as an experiential investigation of the body-mind and its thousands years experience on the topic might inform researchers on crucial features of the human *psyché* such as perception, attention, emotions or consciousness.

Chilean neuroscientist Francisco Varela saw this early on. Initially a biologist interested in cybernetics and lifeforms organization [Maturana and Varela, 1980], he was also a major actor in the development of a fully fleshed science of human cognition. Acquainted with Tibetan Buddhist meditation practice and philosophy, he drew with collaborators Evan Thompson and Eleanor Rosch in *The Embodied Mind* [Varela et al., 1992], what ought to be a visionary research program on human experience through the two-sided lens of cognitive science and Buddhist meditative psychology. Parallel to this groundbreaking book, he initiated with others the first Mind and Life Dialogue between the Dalai Lama and Western scientists and scholars in 1987. This meeting aimed to create a concrete conversation between Buddhist philosophy (initially from the angle of Tibetan Buddhism) and cognitive science, both converging on a deep interest to understand the human mind. After the foundation of the Mind and Life Institute in 1991 many similar conferences followed, exploring subjects as diverse as sleep, altruism, emotions, attention and others. Renowned researchers but also monks and scholars from other contemplative traditions played along to explore the functioning of minds and matter and discuss major issues facing mankind at the dawn of a new millennium. It is actually the published account of a Mind and Life Dialogue on destructive emotions [Goleman et al., 2008] which led the author of the present manuscript to engage in meditation research. Beyond creating vocations, those conferences allowed to set promising research lines for meditation science, as well as collaborations between long-term meditation practitioners of different contemplative traditions and research laboratories. Meditation centers being quite recent in the West at the time, such “expert” meditators were not that easy to find and recruit for scientific experiments. Converging efforts to draw research policymakers’ attention on the potential benefits of meditation as a therapeutic tool in clinic, but also a scientific one to investigate the human mind, created a fertile ground for meditation research at the start of the twenty-first century. The two last decades saw an exponential growth of meditation related scientific publications per year with 131 articles (on mindfulness or meditation) in 2000 to 2978 in 2020, totaling in a 19,752 entries literature [Wang et al., 2021]. Summarizing such a large literature is not our intent, we will instead focus on major cognitive and neural results and the conclusions they enable us to formulate on the effects of contemplative practices on the brain.

1.2.2 The need for a scientific classification of meditative practices

[...] meditation can usefully be framed both as a family of mental training techniques, and as a progressive and iterative process of learning and self-inquiry: over time, and with dedicated practice, it is alleged that meditation practitioners can learn to cultivate specific mental faculties conducive to the alleviation of suffering, and to enduring improvements in psychological well-being.

[King et al., 2019]

We saw in the first part of this chapter the large diversity of contemplative practices and therefore the need for detailed descriptions of the meditations at stake in research studies. After multiple attempts [Lutz et al., 2007, Lutz et al., 2008a], researchers have come with an integrative framework classifying most of the known meditations in three principal families based on their underlying cognitive mechanisms [Dahl et al., 2015]. The first one was described as the “attentional” family and regroup practices developing some kind of attentional regulation. Their common characteristic is to train practitioners in stabilizing, sustaining and redirecting their attention. Doing so they are also hypothesized to reinforce meta-awareness (or meta-cognition): the ability to be aware of one’s attentional processes as well as other mental processes which might go on at the same moment such as perceiving, feeling and thinking. Meta-awareness can be seen as the opposite of experiential or cognitive fusion, where one might be aware of the objects of consciousness (sensations, feelings, thoughts) but not of the mental processes sustaining them. The development of meta-awareness in the place of cognitive fusion is thought to mediate some of the positive effects of mindfulness-based practices on well-being. Further distinction can be made inside the attentional family depending on the direction and extent of the practitioner’s attentional stance. A first group of practices entailing the cultivation of single-pointed concentration on an object of experience was previously called FA meditation [Lutz et al., 2008a]. It encompasses practices such as breath counting in Zen Buddhism, awareness of breathing sensations in modern Vipassana, Shamatha or calm abiding with support from the Tibetan tradition, and body awareness scanning common to many traditions. A second group, previously defined as a single category, are Open Monitoring (OM) practices where the meditator also develops its meta-awareness but without focusing on a specific object. OM meditations were further distinguished in the newest classification [Dahl et al., 2015] into “object-oriented” open-monitoring and “subject-oriented” or “awareness-oriented” open-monitoring. In “object-oriented” OM, as practiced in choiceless awareness meditation from Tibetan Buddhism or Western mindfulness-based interventions, attention is directed to any physical or mental object entering the flow of consciousness. Differently “awareness-oriented” open-monitoring aims at attending to the “knowing quality of awareness itself”, like in Shamatha with-

out support practice. In most Buddhist schools focused-attention practices are traditionally taught first to stabilize and calm the mind, before more advanced open-monitoring practices, as well as “constructive” and “deconstructive” practices which we will describe now.

The second family is composed of “constructive” practices which also have a meta-awareness component, however instead of just observing, they actively engage with emotions and thoughts. Doing so they foster the development of wholesome qualities leading to improvement in personal but also universal well-being. They have been further classified by Dahl et al. depending on their constructive orientation towards: relation, values or perception. The first group of “relation orientation” is mostly known for developing harmonious relationships with others by practicing for example loving-kindness meditation. Such practice aims to cultivate compassion, “a feeling of concern for another person’s suffering which is accompanied by the motivation to help” [Singer and Klimecki, 2014], associated with positive emotions such as joy and love. The second group, “values orientation”, consists in changing one habitual perspective to include ethical and moral considerations, for example by reflecting on one’s own mortality or the universal fragility of life and the need for preserving it. Finally the third group of “perception orientation” aims to nurture a shift in mental perspective up to our physical perception of the world and ourselves. A common pattern of those three constructive practices groups is to develop two important cognitive mechanisms: cognitive reappraisal and perspective taking. Through cognitive reappraisal one implicitly learns to alter his natural responses to challenging situations such as fear, anger or sadness, ultimately forming virtuous habits like patience and equanimity. Complementarily the development of perspective taking qualities allows one to project oneself in a difficult context or in someone else’s shoes, contributing to healthy interpersonal relationships and social emotions such as altruism.

The third and last family regroups “deconstructive” practices which have similar objects of contemplation as the previous ones (perception, emotions, thoughts) but instead of just observing them as in the attentional family, or at the opposite altering them through constructive practices, they aim to explore and dissect the objects of conscious experience. Deconstruction is mainly performed through various forms of “self-inquiry” and aims to elicit insight, a new and sudden intuitive understanding about the world (object-oriented insight), oneself (subject-oriented insight), or their relationship (non-dual-oriented insight). “Object-oriented insight” practices can be directed toward any object appearing to our consciousness, be it a visual image, a sound, a sensation, for example analyzing their impermanent nature as in Vipassana Insight meditation or Tibetan analytical meditations. “Subject-oriented insight” meditations focus on mental cognitive and affective processes, reflecting for example on the emergence of thoughts and emotions and how we are usually fused with them. Finally “non-dual-oriented” practices aimed at putting into question habitual distinctions of self/other and subject/object, challenging our natural view of our Self as independent and permanent, and of the mental content it produces

as truth-value descriptions of reality. Mahamudra and Dzogchen practices are examples of non-dual meditation in the Tibetan tradition, in Zen it is known as Shikantaza or “just sitting” meditation. Deconstructive meditations are often described as more advanced as the practitioner first needs to master attentional meditation to stabilize his attention on the targeted object of experience before being able to inquire about it.

Examples from all three families are described in the main table of Dahl et al. (2015) which we reproduce here for the curious reader (Figure 1.3). This classification has proved very helpful to compare different practices and has notably allow to distinguish the brain correlates of the most studied ones (FA, OM, loving-kindness and compassion, see [Raffone et al., 2019] for a review). However, it is important to keep in mind that contemplative practices should not be taken as totally independent from each other, nor from the broader context in which they are embedded. As described previously some practices from the attentional family are often taught in Buddhist traditions as a preliminary training for more advanced constructive and deconstructive practices. However some attentional meditations are practiced by both beginners and experienced practitioners and it is not impossible to go from an attentional practice to a constructive or deconstructive practice from the same tradition within a unique meditation session, for example to calm the mind before starting a meditation on compassion or on the nature of reality. More generally contemplative practices are rarely done in total isolation from social and cultural contexts whose role in framing subjective experience should not be forgotten when scientifically studying meditation.

Attentional family	Constructive family	Deconstructive family
FA: Jhana practice (Theravada); breath counting (Zen); body awareness practices (Zen/Tibetan); Shamatha/calm abiding with support (Tibetan); mantra recitation (various traditions)	Relation orientation: loving-kindness and compassion (Theravada, Tibetan); Bodhichitta/ Bodhisattva Vow (Tibetan/Zen); centering prayer (Christian); CCARE compassion cultivation training (clinical); cognitively-based compassion training-compassion component (clinical)	Object-oriented insight: mindfulness-based cognitive therapy – cognitive component (clinical); First and Second Foundations of Mindfulness (Theravada, Tibetan); Vipassana/insight (Theravada); analytical meditation (Tibetan); Koan practice (Zen)
OM (object-orientation): cultivation of attention (Greco-Roman philosophy); choiceless awareness (Tibetan); mindfulness-based stress reduction (clinical); dialectical behavior therapy-mindfulness component (clinical); mindfulness-based cognitive therapy-mindfulness component (clinical); acceptance and commitment therapy-mindfulness component (clinical)	Values orientation: The Six Recollections (Theravada); The Four Thoughts (Tibetan); contemplations of mortality (Theravada, Tibetan, Zen, Greco-Roman philosophy); well-being therapy (Clinical)	Subject-oriented insight: cognitive behavior therapy (Clinical); Third and Fourth Foundations of Mindfulness (Theravada, Tibetan); Mahamudra Analytical Meditation (Tibetan); Dzogchen Analytical Meditation (Tibetan); Koan practice (Zen)
OM (subject-orientation): Shamatha/calm abiding without support (Tibetan)	Perception orientation: development stage (Tibetan); meditation on foulness (Theravada)	Non-dual-oriented Insight: Muraqaba (Sufi); Mahamudra (Tibetan); Dzogchen (Tibetan); Shikantaza (Zen); Self-inquiry (Advaita Vedanta)

Figure 1.3: Classification table of contemplative practices with examples, reproduced from [Dahl et al., 2015].

Modern mindfulness meditation presented in health care programs or secular contexts uses practices from all three families, as suggested by its mixed origin from multiple traditions. Mindfulness practitioners are first proposed to accustom themselves with their conscious experience and its underlying attentional movements with exercises such as FA on the breath or body sensations. They are then

introduced to a more open practice akin to object-oriented OM where the full spectrum of experiences is attended to in a non-judgmental way. Cognitive processes proper to the constructive family can also be trained in some mindfulness-based interventions, for example by developing self-compassion and loving-kindness for others. Finally elements of the deconstructive family are touched upon without theorizing them too much, for example by noting the constantly changing nature of all sensations, thoughts and emotions and the distance we can voluntarily take from them.

1.2.3 The effects of attentional training

The faculty of voluntarily bringing back a wandering attention, over and over again, is the very root of judgment, character, and will. No one is *compos sui* [master of himself] if he have it not (sic.). An education which should improve this faculty would be the education par excellence. But it is easier to define this ideal than to give practical directions for bringing it about.

William James in *The Principles of Psychology* (1890)

William James, pioneer in the development of modern psychology, understood in advance the tremendous benefits that could carry a formal education of attention. As we saw in the previous part, most contemplative traditions have for preliminary practices mental exercises aimed at controlling and redirecting one's attention. It is not a surprise then that an important part of the scientific research on meditation was directed on such attentional practices and tested if they have lasting effects outside of meditation, both at behavioral and neural levels. Anyone who has ever tried to do a session of focused-attention meditation noticed how his or her mind kept switching between the object of focus (the breath for example), and an uncountable number of physical and mental distractions (pain in the legs, itching on the nose, noises outside the room, remembrance of last night dinner, planning of tonight's dinner, etc). Focused-attention training is all about recognizing that one's mind wandered away, letting go of the distraction and redirecting one's attention to the breath. Researchers identified core cognitive mechanisms and theoretical brain networks supporting the various mental processes at stake during attentional practices [Lutz et al., 2008a, Malinowski, 2013, Raffone et al., 2019]. As can be seen in Figure 1.4, focusing on a chosen object involves sustaining one's attention, a process most likely supported by the fronto-parietal alerting network. Mind-wandering toward distractive sensations, emotions or thoughts is hypothesized to happen through the activity of the Default Mode Network (DMN). Recognition of this wandering thanks to monitoring processes is thought to be endorsed by multiple regions involved in other networks such as the dorsal Anterior Cingulate Cortex (ACC), the

ventrolateral PreFrontal Cortex (vlPFC), and the anterior insula, which together constitute the salience network. Attentional disengagement from the sources of distraction would be mediated by the executive network also composed of the ACC, as well as lateral ventral cortex, PreFrontal Cortex (PFC) and basal ganglia. Finally the redirection of attention to the object of meditation involves the orienting network (superior parietal cortex, temporal parietal junction, frontal eye fields and superior colliculus).

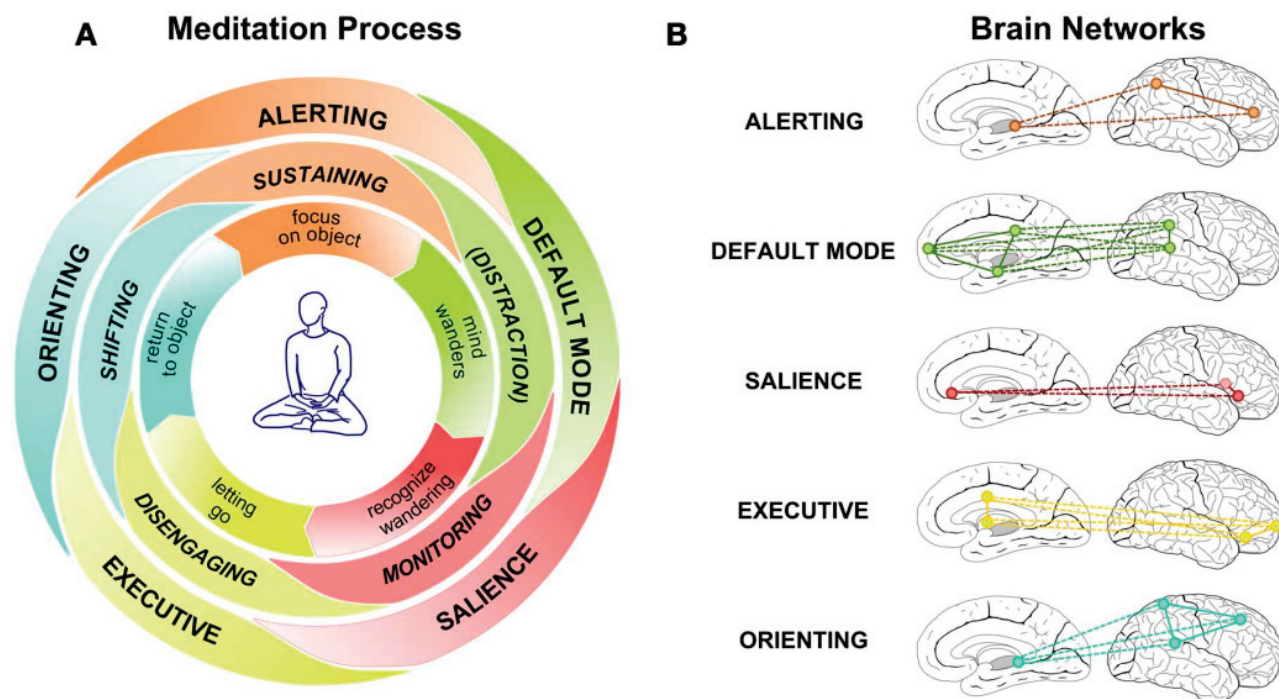


Figure 1.4: Cognitive processes involved in attentional meditative practices (A) and brain networks hypothesized to support them (B). Details of the brain areas involved in each network are given in the main text of the manuscript. *Reproduced from [Malinowski, 2013].*

Those processes and the involvement of brain regions supporting them may not be perfectly delineated in time but instead most likely overlap and interact with each other. Additionally their timing and sequence may vary with the practice at stake as well as meditation practitioners' proficiency. As suggested by Lutz et al. (2008), intensive training in FA should lead to longer periods of attentional sustaining and reduced distractions, as well as less effort for monitoring, disengaging and reorienting the attention to the object of focus. Furthermore, other attention-based practices, mainly object-oriented and subject-oriented OM, may also rely on those networks to become aware of distractions, disengage from them and return to an open attentional stance toward any experience that may arise. Experimental studies of practitioners engaging in single-focus meditation partly confirmed the involvement of hypothesized brain networks [Baron Short et al., 2010, Brefczynski-Lewis et al., 2007, Hasenkamp et al., 2012, Manna et al., 2010]. Of particular interest, meditation experience was found to be correlated with decreased activity in regions involved in monitoring, disengaging and re-orienting attention [Hasenkamp et al., 2012, Manna et al., 2010]. These findings suggest that with accumulated hours of

practice, and so of attentional training, practitioners may need less neural resources to perform those attentional processes, and support the phenomenological descriptions of less effort in FA meditations for experienced meditators. Another study of OM, FA and loving-kindness meditations in both novices and expert practitioners revealed stronger deactivations of two core regions of the DMN (Posterior Cingulate Cortex – PCC and medial Prefrontal Cortex – mPFC) in the experts group [Brewer et al., 2011], emphasizing the importance of increased meta-cognitive abilities in non focused practices.

A more recent randomized study on multiple cohorts of novices participants after 9-month of three different mental training modules, extended these results [Valk et al., 2017]. Of particular interest to us here is the significant increase in cortical thickness in lateral PFC and ACC, regions linked to executive attention, after 9-month training in the “Presence” module which aimed at cultivating present-moment attention and interoception in the same vein as classical mindfulness interventions. The behavioral effects of this meditation module were confirmed, with a seemingly increase in executive control of attention [Trautwein et al., 2020], but no correlation with cortical thickness increase in expected brain regions were found [Valk et al., 2017]. Finally recent systematic reviews on the effect of MBI on cognitive functions revealed strong evidence for a small effect on executive functions — compared to inactive, but not active, comparators— [Cásedas et al., 2020, Whitfield et al., 2021] but no difference in behavioral measures of attention [Whitfield et al., 2021]. The authors emphasized a number of methodological issues in the included studies, notably the poor reliability and lack of consistence of the different measures of attention used. However, a similar meta-analysis focusing only on attentional performance in both experienced meditators practicing FA and OM meditations and novices practitioners following FA/OM based MBI, found that executive and alerting networks of attention were improved by meditation [Sumantry and Stewart, 2021].

We saw that, as suggested by theoretical predictions based on traditional texts [Dahl et al., 2015, Lutz et al., 2008a], contemplative practices have an effect on brain regions involved in attentional processes. Yet the exact neurocomputational mechanisms supporting the attentional effects of meditation are still unknown. The current thesis tried to investigate such mechanisms through the novel understanding of attention brought forth by predictive processing, a theory of brain functioning based on Bayesian inference which we will be the topic of our second chapter. We will now turn toward another important area of research in meditation, namely emotional regulation processes.

1.2.4 Emotional regulation: the example of pain

Mindfulness-based programs have been used in a variety of psycho-pathological conditions as a complement or alternative to traditional medical and psychological treatments. Recent meta-analyses suggest

that they are effective interventions to treat depression, anxiety, pain, schizophrenia and weight/eating-related disorders when evaluated alone ([Goldberg et al., 2018, Goldberg et al., 2022b], Figure 1.5). They also show similar effect as active control interventions for anxiety, pain and weight/eating; and even superiority for depression and addictions (Figure 1.5). When compared to evidence-based treatments (e.g., cognitive behavioral therapy or antidepressant medication), MBI show superior efficacy to prevent major depressive disorder relapse and were equivalent for general depression and anxiety [Goldberg et al., 2018, Goldberg et al., 2022b]. Positive effects of MBI on these pathologies are thought to be mediated, in part, by emotional regulation processes developed in meditative training [Chambers et al., 2009, Wielgosz et al., 2019]. The current PhD project aimed to understand in particular the mechanisms behind meditation-based pain regulation, through the lenses of Bayesian inference and predictive processing. The following section will present some reminders of human pain processing in the brain as well as the view of pain experience from the contemplative side, to finally describe the scientific results obtained from the study of “meditative analgesia” in the last decade. Bayesian models of pain perception will be further presented in the next chapter.

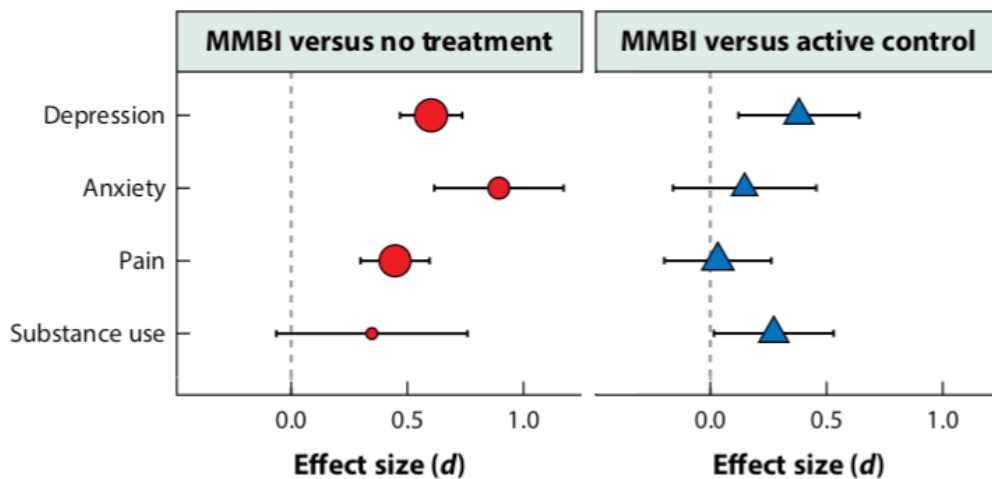


Figure 1.5: Effect sizes for mindfulness meditation-based interventions (MMBI) mediated improvement in different therapeutic conditions compared to no treatment or active control interventions. Comparison to active control treatments led to a reduction in the effect sizes of MMBI, which are still in the range of similar evidence for anxiety and pain, and higher for depression and substance use. *Adapted from [Wielgosz et al., 2019].*

The scientific view of pain processing in the brain

The International Association for the Study of Pain (IASP) recently reviewed its definition of pain as: “an unpleasant sensory and emotional experience associated with, or resembling that associated with, actual or potential tissue damage.” [Raja et al., 2020] and adds as a side note: “that is influenced to varying degrees by biological, psychological, and social factors”. More precisely, pain can be

seen as the interaction between discriminative (where, how much, how long), affective (aversiveness, warning signal) and evaluative (expectations) features [Melzack and Casey, 1968], which indeed can be increased or decreased by one’s biology (pain sensitivity), psychological conditioning (mood, expectation, anticipation) and social context (sport, aggression, therapy, etc. . .). Importantly the IASP specifies that pain can be present without actual tissue damage or even sensation, as it is sufficient to mentally induce a process mirroring such damage to provoke pain, as in the placebo effect [Colloca and Barsky, 2020]. Thus, tissue damage is not necessary, nor sufficient to experience pain as “[...] pain and nociception are different phenomena. Pain cannot be inferred solely from activity in sensory neurons” [Raja et al., 2020]. A characteristic exemplified by rare conditions like congenital analgesia [Nagasako et al., 2003] or pain asymbolia [Gerrans, 2020] where the nociceptive system can be intact yet patients do not *feel* pain.

Lesions and brain imaging studies in animals and humans provided important data to understand how pain experience is formed in the brain. Recent theoretical works converged on the idea of multiple overlapping and flexible networks sustaining painful experience and forming multiple “pain matrices” ([Garcia-Larrea and Peyron, 2013, Tracey and Mantyh, 2007, Garcia-Larrea et al., 2003, Schweinhardt and Bushnell, 2010, Wager et al., 2013], Figure 1.6). A first-order “nociceptive matrix” composed primarily of the posterior insula, secondary somatosensory cortex (SII) and the upper part of the sylvian sulcus —regions hardly distinguishable in functional Magnetic Resonance Imaging (fMRI) and thus usually grouped under the term operculo-insular cortex— as well as Mid-Cingulate Cortex (MCC)¹, primary sensory cortex (SI) and Supplementary Motor Area (SMA), is necessary to produce a sensory pain experience and generate pre-motor and orienting processes. However a fully fleshed conscious pain experience requires the involvement of a second-order “attentional-perceptive matrix” comprising mid and anterior insula, rostral Anterior Cingulate Cortex (rACC), dorsolateral Prefrontal Cortex (dlPFC) and Posterior Parietal Cortex (PPC). These areas are thought to sustain attentional and evaluative pain processes such as anticipation, learning and cognitive control [Garcia-Larrea and Peyron, 2013]. Two particular regions, anterior insula and rACC, are part of a “salience network” involved in the detection of any salient stimuli, not only painful ones [Legrain et al., 2011]. The rACC in particular has been associated more generally to negative affect (including pain), and may be a central node for affective pain modulation due to its close connections to subcortical regions involved in negative stimuli processing (amygdala, nucleus accumbens). The combined activity of nociceptive and perceptual matrices allows pre-conscious nociception to be transformed into a conscious pain experience [Garcia-Larrea and Bastuji, 2018]. All regions of this extended pain matrix receive inputs from the thalamus,

¹The distinction of the MCC as a separate brain region, and not as a dorsal/caudal/posterior part of the ACC, was clearly made only recently [Stevens et al., 2011, Vogt, 2009, Vogt, 2016]), leading to numerous previous articles in the literature talking generally about the ACC, as exemplified in Figure 1.6A. We will try to disambiguate this problematic labelling when needed for the interpretation of neuroimaging results.

an important node in the transmission of nociceptive signals from peripheral to cortical centers [Tracey and Mantyh, 2007].

Finally, a third-order “affective-reappraisive” matrix influences pain experience depending on individual priors about pain (for example the degree of pain catastrophizing [Sullivan et al., 2001]), emotions and expectations [Atlas and Wager, 2012]. This last network is formed of polymodal regions outside of the classical pain matrix: the perigenual Anterior Cingulate Cortex (pgACC), OrbitoFrontal Cortex (OFC), temporal pole, and anterolateral prefrontal areas. They are activated during cognitive modulation of pain such as placebo and nocebo effects, mood related pain modulation and self-regulation of pain [Wiech et al., 2008], likely promoting a “reappraisal” or reinterpretation of painful stimuli. These regions, and particularly the ventromedial prefrontal cortex (composed of pgACC and OFC) are implicated in the top-down modulatory control of pain (through the periaqueductal gray and brain-stem), therefore indirectly influencing primary and secondary matrices through inhibition of ascending nociceptive signals [Bushnell et al., 2013].

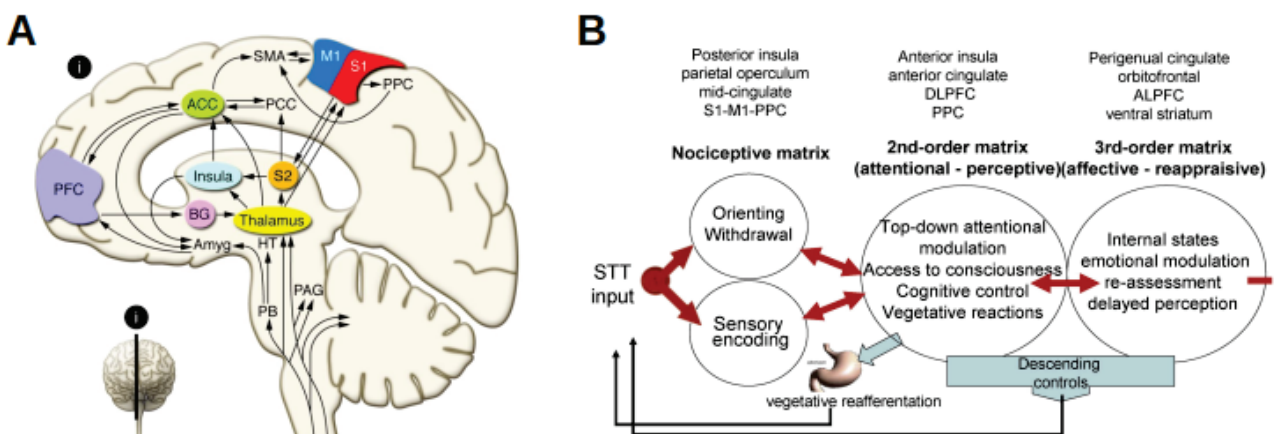


Figure 1.6: **A:** Schematic representation of pain processing brain regions, *adapted from [Schweinhart and Bushnell, 2010]*. **B:** Updated “pain matrices” proposed by [Garcia-Larrea and Peyron, 2013] with first-order nociceptive, second order attentional-perceptive and third order affective-reappraisive networks. S1: primary somatosensory cortex, S2: secondary somatosensory cortex, ACC: anterior cingulate cortex, PFC: prefrontal cortex, M1: primary motor cortex, SMA: supplementary motor area, PCC: posterior cingulate cortex, BG: basal ganglia, HT: hypothalamus, Amyg: amygdala, PB: parabrachial nuclei, PAG: periaqueductal gray, DLPFC: dorsolateral prefrontal cortex, ALPFC: anterolateral prefrontal cortex.

Contemplative view of pain experience

Just as if they were to shoot a man with an arrow and, right afterward, were to shoot him with another one, so that he would feel the pains of two arrows, in the same way, when touched with a feeling of pain, the uninstructed run-of-the-mill person sorrows, grieves, and laments, beats his breast, becomes distraught. So he feels two pains, physical and mental. [...] Now, the well-instructed disciple of the noble ones, when touched with a feeling of pain, does not sorrow, grieve, or lament, does not beat his breast or become distraught. So he feels one pain: physical, but not mental.

Siddhārtha Gautama (Buddha) in *The Arrow, Sallatha Sutta, SN 36* [Bhikkhu, 2003]

This well known discourse attributed to the historical Buddha, emphasizes the psychological view of pain shared by most Buddhist traditions. Akin to scientific theories on pain, Buddhist philosophy makes the distinction between physical and mental pain, and proposes that meditation may be able to modulate the second, mental “arrow”, which can be thought of as combining attentional, affective and cognitive dimensions of scientific models. The Arrow sutra goes on :

“As he [the well-instructed disciple] is touched by that painful feeling, he is not resistant. No resistance-obsession with regard to that painful feeling obsesses him. Why is that? Because the well-instructed disciple of the noble ones discerns an escape from painful feeling aside from sensuality. [...] Sensing a feeling of pleasure, he senses it disjoined from it. Sensing a feeling of pain, he senses it disjoined from it. Sensing a feeling of neither-pleasure-nor-pain, he senses it disjoined from it. This is called a well-instructed disciple of the noble ones disjoined from birth, aging, and death; from sorrows, lamentations, pains, distresses, and despairs. He is disjoined, I tell you, from suffering and stress.” [Bhikkhu, 2003], *SN 36*

The key would be to just observe all sensations, including pain, without trying to resist them mentally or judging them as positive or negative, as we are naturally inclined to do. This seemingly simple technique of mere observation is taught in attentional practices such as OM where pain is seen as just another sensory experience without further elaboration needed. Constructive and deconstructive practices can also bring about further perspective taking about pain, mostly its affective component which we can “disjoin” from by seeing it as a mere creation of our mental. Doing so, only the physical sensation, the first arrow, remains and for the better as it is a necessary life-saving signal selected through evolution. Kabat-Zinn clearly saw the therapeutic potential of such detached, yet aware, attitude toward pain

when he tried to apply his newly created MBSR program — inspired from diverse contemplative traditions — to chronic pain patients resistant to traditional medical care [Kabat-Zinn, 1982]. The results were much convincing, with significant improvement in physical and psychological pain status just after the intervention [Kabat-Zinn et al., 1985] but also up to four years later on more than two hundred patients [Kabat-Zinn et al., 1986]. After over thirty years of clinical research on various mindfulness-based interventions, chronic pain types and patients population, several meta-analyses and reviews allow us to draw preliminary conclusions [Bawa et al., 2015, Hilton et al., 2017, Goldberg et al., 2018, Goldberg et al., 2022b]. In summary, mindfulness interventions are more effective at reducing chronic pain symptoms than passive interventions and show similar efficacy to active controls (traditional evidence based-treatment or education/sport groups). Interestingly overall evidence seems to be stronger for an effect on the affective dimension of pain (depression, mental health-related quality of life) than its physical dimension. Nevertheless all reviews highlighted the need for higher quality studies with rigorous control groups to draw definite conclusions on the efficacy of meditation for chronic pain disorders.

Scientific insights on pain modulation by meditation

As we just saw, the clinical application of mindfulness meditation to relieve pain started in the early 1980s with Kabat-Zinn studies, yet a systemic scientific investigation of the mechanisms involved in “meditative analgesia” is very recent [Grant, 2014]. Two lines of research can be delineated, one interested in the mechanisms behind pain relief in brief mindfulness-based training on naive participants, and another focusing on the state and trait effects of intensive meditation practice, that can be measured in long-term meditation practitioners. Those parallel research programs are complementary and necessary. A precise mechanistic account would more likely be available from the study of expert meditators with a long experience of pain regulation. However generalization of those mechanisms to short meditation training in a health-care context is not straightforward and needs to be investigated. A first pilot study on 42 naive students trained either in mindfulness over six, one hour sessions, or a control condition (guided visual imagery over two, one hour sessions) provided evidence that brief mindfulness training on healthy subjects can increase pain tolerance but this effect was not linked to a similar change in mindfulness skills [Kingston et al., 2007]. Numerous subsequent studies systematically identified a reduction in intensity or unpleasantness pain ratings — associated respectively to sensory and affective pain dimensions — in participants following a short mindfulness intervention compared to relaxation or bare attention to the breath [Zeidan et al., 2010, Zeidan et al., 2011], but also placebo and sham-mindfulness conditions [Zeidan et al., 2015].²

²In this study, sham-mindfulness was an active control mirroring their meditation intervention in all psychosocial factors (group, posture, breathing, eyes closed) except the core mindfulness component: mindful and

Neuroimaging results from some of those studies identified several brain structures mediating mindfulness-based pain regulation. Pain modulation during painful stimulation was associated with significant deactivation in primary sensory cortices (SI) of the stimulated zone as well as in the thalamus, two major regions in the sensory processing of pain. Greater activation of the pgACC, right anterior insula and bilateral OFC were also observed ([Zeidan et al., 2011], see Figure 1.7 middle panel). Modulation of affective and cognitive components of pain by meditation may be sustained by these three regions part of second and third-order pain matrices. Interestingly these brain regions were distinct to the ones implicated in placebo (activation of ACC, dlPFC and deactivation of posterior insula and SII) and sham-mindfulness analgesia (activation of thalamus, putamen, PCC and SI and deactivation of ACC and medial PFC) when tested in the same study [Zeidan et al., 2015]. Finally the same research team showed that even in participants without any meditative training, their degree of “dispositional” or trait mindfulness (assessed through the Freiburg Mindfulness Inventory, FMI or the Five Facets Mindfulness Questionnaire, FFMQ) could also explain lower intensity and unpleasantness ratings and was associated with weaker activation and connectivity between nodes of the DMN (mPFC, PCC, precuneus) ([Harrison et al., 2019, Zeidan et al., 2018], see Figure 1.7 upper panel).

Studies of long-term meditation practitioners revealed coherent but still different mechanisms at stake during pain modulation by extensive training. The first study from Grant and Rainville [Grant and Rainville, 2009] showed that experienced Zen practitioners (> 1000 hours of practice) had higher thermal pain tolerance than age and gender-matched meditation naive controls, a difference most likely mediated by higher scores in observing and non-reacting dimensions of the FFMQ. Long-term practitioners also presented reduced intensity and unpleasantness ratings during concentrative (FA style) and mindfulness (OM style) meditations, compared to controls. Finally these reductions in sensory and affective pain were correlated to meditation experience (only significant for intensity ratings) and breathing rate. Surprisingly these behavioral results, as well as previous ones on brief mindfulness training, are at odds with subsequent studies showing significant differences only in unpleasantness (not intensity) pain ratings between naive and expert subjects [Brown and Jones, 2010, Gard et al., 2012, Lutz et al., 2013, Perlman et al., 2010]. The decoupling between sensory and affective components of pain perception during and as a result from intensive mindfulness meditation recently received further evidence from our research group. Expert practitioners in the Tibetan Buddhist tradition (more than 10,000 hours practice) demonstrated higher reduction in pain unpleasantness but not intensity during both open monitoring practice and a control condition, compared to novice subjects which also exhibited a significant difference between meditation and control conditions [Zorn et al., 2020]. In a subsequent study [Zorn et al., 2021], this sensory-affective uncoupling proved to be better explained by an increase in cognitive defusion, the ability to distance oneself from negative experiences such as non-judgmental attention to the breath, see [Zeidan et al., 2015] for more details.

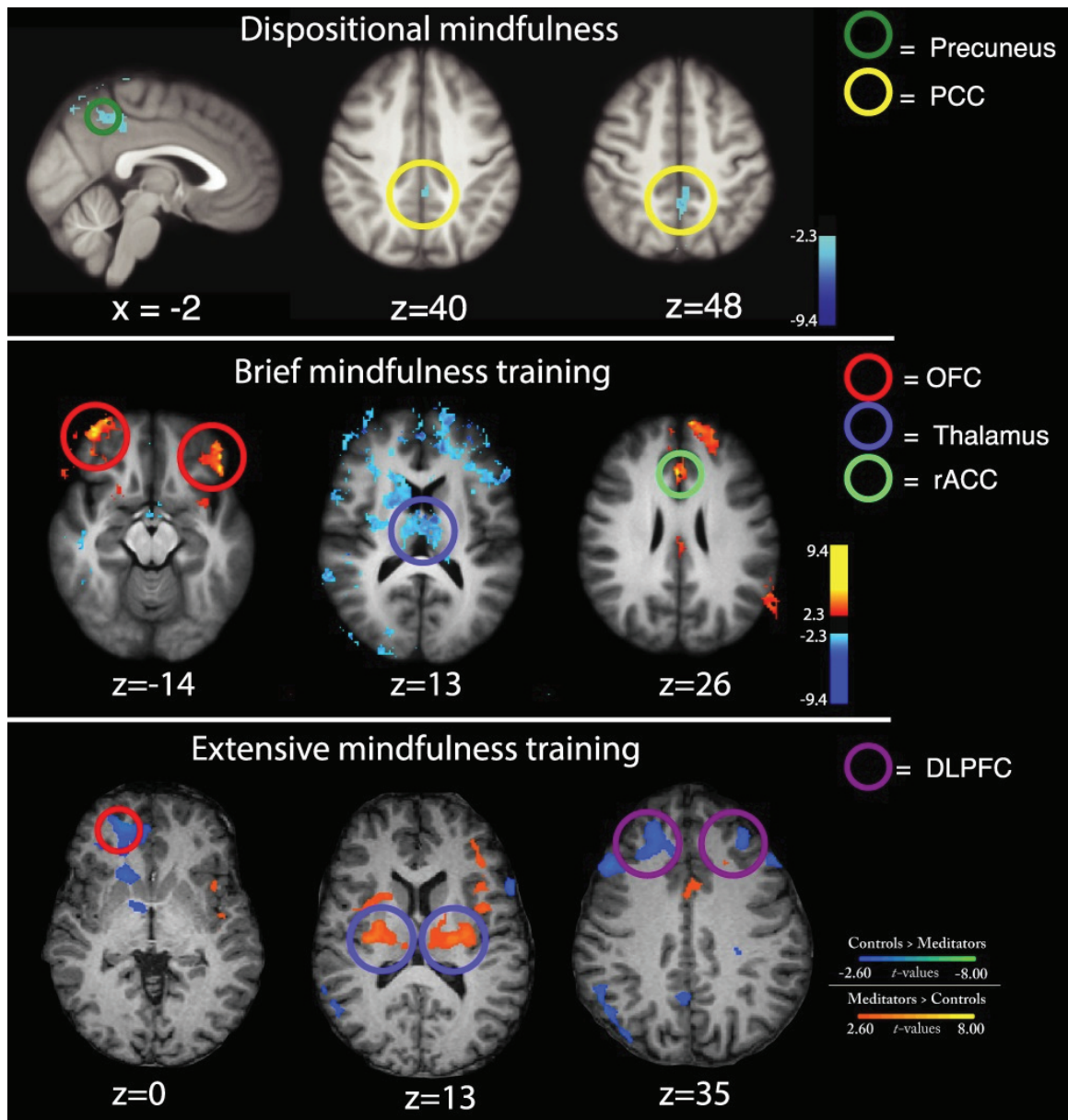


Figure 1.7: Partial summary of fMRI results on pain-related brain activation as a function of dispositional mindfulness (upper panel), brief mindfulness training (middle panel) and extensive mindfulness training (lower panel), adapted from [Zeidan et al., 2019]. **Upper panel:** higher trait mindfulness was associated with greater deactivation of posterior cingulate cortex (PCC) and precuneus, and lower pain. **Middle panel:** pain relief during heat stimulation caused higher activation of orbitofrontal cortex (OFC) and rostral anterior cingulate cortex (rACC), but deactivation of the thalamus, for participants included in a short mindfulness program compared to controls. **Lower panel:** Experienced Zen practitioners presented higher activation of thalamus and deactivation of OFC and dorsolateral prefrontal cortex (dlPFC) compared to age-matched controls.

pain (measured with the Drexel Defusion Scale, DDS [Forman et al., 2012]), rather than a decrease in pain catastrophizing, an exaggerated negative mental attitude expressed during pain or its anticipation (as measured by the Pain Catastrophizing Scale, PCS [Sullivan et al., 1995]). In addition, a mixed-methods study combining the results mentioned above with qualitative data obtained from in-depth structured interviews with expert and novice participants on their relation to pain experience [Poletti et al., 2021], revealed how long-term meditation practice can thoroughly change the experience of pain

(for instance to encompass qualities of acceptance and compassion). Another very recent experimental work on highly-experienced Theravada Buddhist practitioners (mean lifetime practice of 25,000 hours) further advanced our understanding of the mental causation of pain [Nicolardi et al., 2022]. Basing themselves on the Buddhist psychology model of the “two arrows” of pain, the authors could observe (Granger) causal links between identification, aversion and pain experience, with a notable reduction of both pain and its identification during loving-kindness meditation (constructive family of practices). Finally a study of the nociceptive reflex under fear conditioning, combining various meditative styles, provided preliminary evidence that meditation does not block sensory pain processing at the spinal level nor necessary pain related fear learning; but instead impedes its later influence on pain expectations, anticipation and perception [Taylor et al., 2018].

Results from fMRI studies point toward various brain changes associated with pain modulation by long term practice. Experienced Zen meditators were shown to present higher activation of primary nociceptive processing regions (thalamus, insula and dorsal ACC = anterior MCC in the new nomenclature) and lower activation of areas involved in cognitive and affective modulation (dlPFC, OFC) compared to matched controls ([Grant et al., 2011], see Figure 1.7 lower panel). Interestingly the results correlated with their years of meditation and the decoupling between executive and pain processing cortices (decreased dlPFC–ACC connectivity) was linked to higher pain tolerance. The fact that participants did not actually engage in meditation during the fMRI recording supports the claim that meditation-related changes in brain and behavior might extend to day-to-day life. Another study on experienced Vipassana practitioners displayed similar results with increased activation of right posterior insula and SII, and decreased activation of lateral PFC [Gard et al., 2012]. Finally Lutz and collaborators [Lutz et al., 2013] studied highly experienced Tibetan Buddhist practitioners (mean meditation experience of 27,000 hours) during thermal pain but also its preceding anticipatory period. If core regions of the pain salience network (left anterior insula and aMCC) were more activated during pain stimuli in practitioners compared to meditation naive controls; these same regions as well as the amygdala showed reduced activation during anticipation for experts participants only. This deactivation of the left insula was correlated to lifetime meditation practice. Another report using EEG presented a similar decreased activity in MCC during pain anticipation for expert meditators, related to lower unpleasantness ratings and higher lifetime practice [Brown and Jones, 2010].

We saw in this section that the research on meditation and pain is already quite advanced with numerous behavioral and brain imagery studies shedding light on the specific effects of meditation on the affective component of pain and the probable brain regions mediating them. Yet, similarly to the effect of meditation on attentional training, a clear understanding of the mechanisms at play is still awaited for. Once again we suggest that predictive processing and Bayesian accounts of pain

perception (which will be detailed later in chapter 2 and 3) might provide an integrated view of pain modulation by meditation, from behavioral to neuronal levels. In this sense we implemented during this PhD an experimental study testing some core predictions, and that will be presented in chapter 5. For now we will turn to the limitations facing the contemplative research field, some of them are common to all psychological sciences and neurosciences and start to be addressed in a global way (e.g., lack of reproducibility). Others hindrances such as the difficulty to define “mindfulness” and how to deal with adverse effects are proper to meditation research. Identifying these problems allowed us to solve some of them during this PhD.

1.2.5 Hindrances facing contemplative research

We had a sense throughout this chapter, of how deeply the research work presented in this thesis stands upon many shoulders, from early Buddhist practices and philosophy to its modern scholar study, not forgetting recent clinical and fundamental investigation of the effects and mechanisms at stake in contemplative practices. Yet we must also acknowledge that biases and flaws, inherent to the study of a complex phenomenon, are present in the scientific literature on meditation. Indeed contemplative sciences are not immune to a number of issues undermining scientific research in general: false positives [Ioannidis, 2005], p-hacking [Head et al., 2015], HARKing (Hypothesizing After the Results are Known) [Kerr, 1998] and reproducibility [Munafò et al., 2017, Begley and Ioannidis, 2015] to cite only a few. Caveats proper to neurosciences such as double dipping in neuroimaging data analysis [Kriegeskorte et al., 2009, Button, 2019] or low sample size [Button et al., 2013] can also affect results from neuroscientific research on meditation and should be carefully prevented. Furthermore, borrowing numerous methods and theoretical grounds from psychology, experimental works on meditation were affected by recent critics concerning psychological science [Vul et al., 2009, Ferguson and Brannick, 2012, Kühberger et al., 2014, Makel et al., 2012, Simmons et al., 2011]. Yet they could also benefit from the growing incentive for opened, registered and reproducible science lead by researchers in behavioral and brain science [Nosek and Lakens, 2014, Nosek et al., 2018, Open Science Collaboration, 2015]. Hindrances specific to research on meditation exist and have been documented in several reviews and meta-analyses [Davidson and Kaszniak, 2015, Coronado-Montoya et al., 2016, Goldberg et al., 2017, Van Dam et al., 2018].

Reproducibility of meditation-related results and other methodological challenges

A first major hindrance, linked to the replication crisis, is the inconsistencies in many neurobehavioral and clinical results on meditation. For example, as will be presented later in a replication article

to which I participated [Fucci et al., 2021], the effects of meditation on well-studied Event Related Potentials (ERPs) – such as the Mismatch Negativity (MMN) – or brain oscillations (e.g. alpha power from 8 to 12 Hz) derived from Electroencephalography (EEG) recordings are still not clear despite an important number of studies on these measures, and may be biased by previously mentioned caveats such as false positives, lack of replication and low sample sizes. On the clinical side warnings were also raised early on concerning the methodological rigor of the first scientific studies on mindfulness-based interventions from 1980 to 2000's [Baer, 2003, Bishop, 2002], and unfortunately are still relevant today [Davidson and Kaszniak, 2015, Dimidjian and Segal, 2015, Farias et al., 2016, Goldberg et al., 2017]. Goldberg et al (2017) systematically investigated potential improvements of MBI studies in time over the use of active control conditions, larger sample sizes, longer follow-up assessment, treatment fidelity assessment, reporting of instructor training and intent-to-treat (ITT) analyses [Goldberg et al., 2017]. Astonishingly they found no significant evidence for improvements along any of these measures, but changes were at least in the good direction. Exclusion of an early high-quality study [Teasdale et al., 2000] allowed to see significant improvements over time in sample size, treatment fidelity assessment, and reporting of ITT analyses. The authors acknowledged with others that the observed absence of improvement might be due to the development of a large number of mindfulness-based programs in the recent years, which are thus still in testing phase with less rigorous methodology [Dimidjian and Segal, 2015]. Besides, if no improvements is seen overall, an important part of the studies now published make use of an active control for example (40% in [Goldberg et al., 2017] sample), allowing to draw firm conclusions based on meta-analyses restricted to these higher-quality studies [Goldberg et al., 2018, Goldberg et al., 2022b]. However these same studies may also be biased to report positive findings, an issue almost never raised by previous meta-analyses [Coronado-Montoya et al., 2016]. If it does not prevent the field to aim for clinical gold standards, such flaws are not restricted to meditation research but shared among many psychological and neuroscientific disciplines [Ferguson and Brannick, 2012, Luck and Gaspelin, 2017, Open Science Collaboration, 2015, Vul et al., 2009].

How to define “mindfulness” ?

A second important issue, particularly relevant to research on contemporary mindfulness meditation is the difficulty to have a single and unified definition of “mindfulness” [Van Dam et al., 2018, Lutz et al., 2015]. Indeed when looking at it more closely, one sees that “mindfulness” is actually a complex construct, with three main meanings used in its scientific study. Firstly, it can be seen as a mental trait, more or less present in each individual, practicing meditation or not, and which can be measured through validated psychometric questionnaires —for example the Five Facets Mindfulness Questionnaire (FFMQ, [Baer et al., 2006]), the Mindfulness Attention Awareness Scale (MAAS, [Brown and

Ryan, 2003]) or the Freiburg Mindfulness Inventory (FMI, [Walach et al., 2006]) to cite only a few. Secondly it is often described as a “way of life” for cultivating well-being and alleviate suffering in accordance with its Buddhist roots, a usage which had great influence in its initial development and spread in the Western world [Lutz et al., 2015]. A third usage employs it in a more technical way to describe a core cognitive process developed through meditation training and which rejoins Kabat–Zinn initial description as paying attention in the present moment and non-judgmentally. Yet the interpretation of this “mindfulness” capacity is the matter of an animated debate among Buddhist scholars with multiple authors offering critical analyses of the non-judgmental and present–centered features of traditional contemplative practices [Bodhi, 2011, Dreyfus, 2011, Dunne, 2015, Gethin, 2011]. The other meanings of mindfulness cited above also have their own caveats, with trait measures inherently biased by social–desirability responding or demand characteristics for instance, while describing mindfulness as a way of life is not specific enough to allow the scientific study of precise mechanisms [Van Dam et al., 2018, Lutz et al., 2015]. Overall disagreement concerning the definition of mindfulness as well as diverging use of the word led to difficulties in integrating results from different studies, as well as to compare experimental results from mindfulness-based program to more traditional contemplative practices [Van Dam et al., 2018].

To overcome these obstacles, researchers recently proposed to approach mindfulness as “a variety of cognitive processes embedded in a complex postural, aspirational, and motivational context that contribute to states that resemble one another along well-defined phenomenological dimensions” [Lutz et al., 2015]. The problem is then shifted from the definition of mindfulness to its decomposition into multiple dimensions or features of meditative experience, whose dynamic variations could theoretically be mapped to different practice styles as well as degrees of expertise. The initial proposal identified three primary, “functional”, dimensions of formal meditation practice under mindfulness training (object orientation, dereification and meta-awareness), as well as four secondary, “qualitative”, dimensions (aperture, clarity, stability and effort). A detailed description and justification of these dimensions is beyond the scope of the current thesis, yet we can briefly have a look at how these features allow to map different meditative practices and expertise levels in a phenomenological space (Figure 1.8). Building on the distinction made earlier between focused attention (FA) and open monitoring (OM) styles of attentional meditative practices, we can now see concretely how they differ in this phenomenological space. For example FA practice in novice practitioners (the usual starting practice in most contemplative traditions, for example FA on the breath or body sensations) is characterized by a high degree of orientation toward an object and low meta-awareness and dereification, at least at the beginner stage (see FA-nov in Figure 1.8). Further training in the practice may lead to an increase in those last dimensions as well as in clarity of experience and stability, whereas the effort necessary to maintain focus on a given object would diminish (FA-exp in Figure 1.8). Similarly one can describe OM prac-

tices which are from the start less object-oriented than FA and may require less effort. Interestingly the mapping of mental states relevant to psychopathology and which are thought to be alleviated through MBI, reveals their discrepancy to mindfulness experience in this phenomenological space (see mind-wandering, addictive craving and rumination in Figure 1.8).

Following the work from Lutz et al (2015), similar phenomenological approaches for the study of meditation have been derived with varying dimensions and applications [Van Dam et al., 2018, Vago and Zeidan, 2016]. The main advantage of this paradigm shift is to provide a common theoretical framework from which to cast mechanistic hypotheses about cognitive processes at stake during mindfulness meditation. These hypotheses can then be tested by well-defined phenomenological scales given to meditation practitioners (see [Abdoun et al., 2019, Poletti et al., 2021, Nave et al., 2021] for non-exhaustive examples), and ideally correlated to physiological and neurophysiological markers to gain a better understanding of meditation-related changes, rejoining the neurophenomenological research program envisioned by Francisco Varela [Lutz and Thompson, 2003, Varela, 1996].

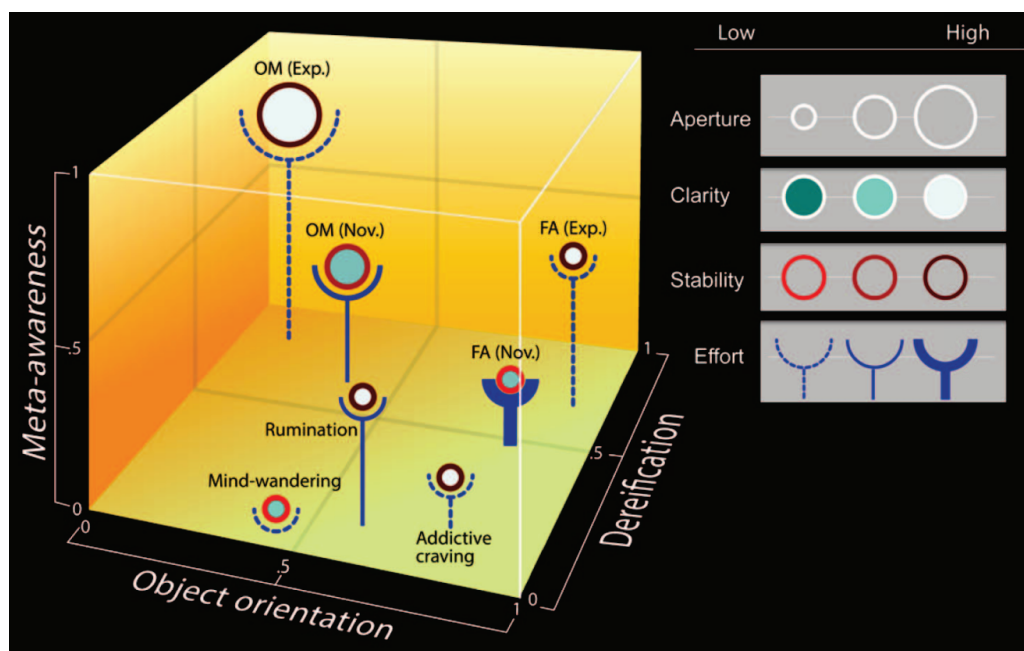


Figure 1.8: The phenomenological matrix of mindfulness presented by Lutz et al. (2015). The primary dimensions (meta-awareness, object-orientation and dereification) are mapped on the orthogonal axes of the matrix while secondary features (aperture, clarity, stability and effort) are described through different symbols. The position of meditative practices (OM, FA) depending on degrees of expertise (Exp.: Experts, Nov.: Novices) within the phenomenological space can be visualized, and compared to typical mental states potentially at stake in some psychopathologies. *Reproduced from [Lutz et al., 2015].*

Adverse effects in meditation practices

A last major hindrance facing contemplative research is adverse effects (AEs) of meditation, or rather their lack of systematic reporting and prevention in MBI research trials. Adverse effects or events

can be any harmful effect resulting from an intervention or treatment, they are qualified of “serious” when their occurrence threatens the life or function of a participant [Ioannidis et al., 2004]. Such effects are rare but still expected in any behavioral intervention, furthermore on clinical populations. Meditation interventions are no exception and adverse effects have actually been reported early on in a seminal review [Shapiro, 1982], a retrospective study on transcendental meditation [Otis, 1984], and in a preliminary prospective study in long-term meditators [Shapiro, 1992]; on top of being well-documented in Buddhist texts [Van Dam et al., 2018]. These unwanted and potentially harming effects are not a problem in themselves if they are known, prevented, anticipated and taken care for. At the opposite if they are silenced and not skillfully investigated, they could put at risk vulnerable patients with ongoing or previous psychiatric conditions. Exclusion criteria guidelines for participating to a MBI (e.g. substance dependence, suicidality, psychosis, severe depression, severe social anxiety, etc) are provided by MBSR and MBCT programs [Kuyken et al., 2012, Santorelli et al., 2017] —the two more prevalent and well-tested MBIs— but without official endorsement by national health agencies they may not be well followed. Medical organizations indeed do not provide clear guidelines concerning the exclusion of at-risk patients from MBI: the American Psychiatric Association identifies only few meditation-induced clinical problems in its *Diagnostic and Statistical Manual of Mental Disorders* [American Psychiatric Association et al., 2013]; and the American National Institutes of Health simply states that “meditation could cause or worsen certain psychiatric problems” and advises patients to check with their doctor.

Fortunately recent research efforts were made to identify and understand adverse effects caused by meditation practice. The extensive review from Baer and collaborators [Baer et al., 2019] provides an integrated view on the topic. A first alarming analysis is that despite the fact that most Randomized Controlled Trials (RCTs) on MBI need ethical approval from human research committees asking for adverse effects monitoring, their actual reporting is very variable between studies, with the largest meta-analyses on respectively 47 and 231 studies showing that only 19 and 16% of them reported data on AEs [Goyal et al., 2014, Wong et al., 2018]. The prevalence of adverse effects when reported was also quite variable, with 0 and 1% of AEs in the largest meta-analyses cited above, up to 10% for a meta-analysis of MBIs for PTSD patients [Banks et al., 2015], with previous psychiatric condition potentially explaining such a high rate. Overall these reported AEs were not qualified as serious and were not more common than in control groups not receiving a meditation intervention [Baer et al., 2019]. Yet these results should be taken cautiously, first because of the poor representativeness of AEs reporting studies in regards to all MBI trials (between 16 and 19%). Second second many factors could impede the veridical report of these AEs, such as not being well-defined for meditation practice or not being reported due to auto-censorship from participant arising from social desirability, combined with a positively biased view of mindfulness. Indeed when prospectively investigated through online

surveys to experienced meditators from diverse traditions, AEs prevalence revealed to be quite higher, ranging from 4 to 63% [Baer et al., 2019].

More recently two online surveys with very high samples (more than 400 and 1000 participants respectively) found prevalence of meditation-related AEs or “unpleasant experiences” in respectively 32 and 25% of regular meditation practitioners [Goldberg et al., 2022a, Schlosser et al., 2019]. Careful scrutiny is again needed to not draw hasty pessimistic conclusions from these results, for example that meditation practice can be “harmful” in one quarter of people and thus should be avoided. As highlighted in previously quoted articles as well as in Baer et al. (2019), adverse or unpleasant experiences are documented in traditional contemplative practices and seen as inherent to the training they propose to ultimately foster well-being, equanimity and relief from suffering. However, it is worth emphasizing that just because an experience is unpleasant does not mean it is necessarily harmful. A simplistic analogy can be drawn with physical exercise where one’s muscles ache yet we still do it because we trust it is good for our health in the long run. But a trade-off between costs and benefits has to be found, exercising too much at once or in a maladaptive way, may indeed cause injury and be harmful. Similarly we can imagine that an appropriate and skill–full “dosing” of meditation could be defined, which optimizes the benefits/costs balance depending on individual factors. It is actually what meditation instructors often do intuitively, adapting and shaping their teachings to their audience. Obviously further scientific understanding of such “dosage” and quality effects is not an easy task, as meditative experiences are embedded in a personal, social and sometimes religious contexts. Multiple theoretical works started to embrace this issue [Baer et al., 2019, Compson, 2018, Lindahl et al., 2019], questioning for instance how to make the difference between relevant psychopathological AEs and expected, sound, contemplative development.

In this respect a recent study offered interesting insights by thoroughly investigating through qualitative interviews meditation-related experiences, and more particularly “challenging” ones, in a sample of 60 Buddhist practitioners equally coming from Theravada, Zen and Tibetan traditions and with varied type of practices [Lindahl et al., 2017]. Of importance 60% of these participants were meditation teachers and 40% of them had more than 10.000 hours of lifetime meditation, stressing their particular experiential knowledge of these experiences. Their database was completed by interviews of 32 “experts” (11 of whom were also interviewed as practitioners) describing their experience as meditation teachers when their students had these kind of experiences and how they managed it. They could identify 59 different meditation-related experiences gathered in seven domains (cognitive, perceptual, affective, somatic, conative, sense of self, and social) as well as their valence, severity, duration, and the practitioners/experts’ responses to it. Overall they found that very few experiences were unanimously described as adverse, suggesting that such adversity is highly dependent on the complex interaction of multiple factors [Lindahl et al., 2017]. These influencing factors were further investigated within the

interviews with practitioners and experts. In summary, this pioneering qualitative study provides an example of the possibility to seize and disentangle adverse effects in meditation practices, and offer some leads for potential remedies. Inspired by their work, we assessed meditation-related uncommon experiences during our investigation of an intensive 10-day mindfulness meditation retreat. The important role of contemplative research in the future would be to acknowledge the potentiality of these unpleasant meditation-related experiences, and to carefully report and study them in order to better know when and how adverse effects can occur, with the ultimate goal of not exposing at-risk individuals, or adapting the practice to them.

Resonating with the justified criticisms developed above on contemplative research we strove to implement some of the proposed solutions in our experimental work. For instance, control groups were systematically used, group assignment was randomized when applicable and sample sizes were raised to high limits in the field. Adverse events, which had a non-null probability to occur in our study of a 10-day meditation retreat, were carefully assessed. We additionally tried to follow the scientific shift for an open and reproducible science by making our data and analysis code available online, moving away from frequentist statistical analysis toward Bayesian statistics and in the case of our second study, preregistering its analysis online. These methodological considerations will be described in more details in our third chapter. Importantly we also based our experimental work and hypotheses on theoretical models such as the one proposed by Lutz et al. (2015) as well as novel understandings of brain functioning within a Bayesian framework. We will now describe this framework and its implementation in the brain as predictive coding which was a keystone of this doctoral project.

Chapter 2

Perception as hierarchical Bayesian inference in the brain

[...] organism and environment enfold into each other and unfold from one another in the fundamental circuitry that is life itself.

Francisco Varela, Evan Thompson and Eleanor Rosch in
The Embodied Mind: Cognitive Science and Human Experience (1991)

2.1 A Bayesian framework for perception, learning, action and much more

The idea that perception by living organisms results from a process of casting inferences about the causes of sensations can be traced back to Helmholtz some 150 years ago [Kahl, 1971]. This avant-garde intuition recently led to the formalization of perception and learning as processes of hierarchical inference in the brain, under the *predictive coding or processing* framework [Friston, 2003, Friston, 2005, Kiebel et al., 2009]. Friston and colleagues further extended the explanatory scope of the framework (termed “active inference” in its latest incarnation) to include action, planning and decision-making [Brown et al., 2013, Friston et al., 2010a, Friston et al., 2011], interoception, emotions and the sense of Self [Allen et al., 2019, Barrett and Simmons, 2015, Limanowski, 2017, Seth et al., 2012, Seth, 2013], up to consciousness [Friston, 2018a, Rudrauf et al., 2017, Williford et al., 2018], to cite only a few domains. In fact, it has been proposed that these various body-brain processes — but also life itself, as well as its evolution — could be usefully examined through the lenses of a single organizing notion, namely the Free-Energy Principle (FEP) [Friston et al., 2006, Friston and Stephan, 2007, Friston, 2009, Friston, 2010, Friston, 2013, Kirchhoff et al., 2018, Ramstead et al., 2018]. In the forthcoming section we will explain in more details the FEP and its hypothesized implementation in the brain. To do so, we will first introduce some necessary concepts such as the Bayes’ rule and its corollary, *Bayesian inference*, which play a central role in upcoming theoretical descriptions. We will also illustrate the Bayesian brain hypothesis [Pouget et al., 2013], which was contemporary to predictive coding formulations, and offers a parallel proposal on how the brain may deal with uncertainty in its environment.

2.1.1 Bayes’ rule, Bayesian inference and the Bayesian brain hypothesis

Bayes’ rule, named from English mathematician Thomas Bayes (1701-1761), is central to modern Bayesian statistics and has had tremendous influence in many scientific fields. It derives from the mathematical expression of the *joint probability* $p(A, B)$ (or $p(A \cap B)$) of observing two events, A and B at the same time :

$$p(A, B) = p(A|B)p(B) \tag{2.1}$$

where $p(A|B)$ denotes the *conditional probability* of observing A knowing that B occurred and $p(B)$ denotes the probability of observing B regardless of whether A occurred or not.

As $p(A, B)$ depends symmetrically on A and B , Equation 2.1 can be rewritten by marginalizing on A :

$$p(A, B) = p(B|A)p(A) \quad (2.2)$$

Combining the two above equations one can obtain the Bayes' rule:

$$p(A|B) = \frac{p(B|A)p(A)}{p(B)} \quad (2.3)$$

Next, we can define Bayesian inference as an “informed guess” about unknown causes from observable consequences. Importantly, Bayes' rule defined on discrete events A and B can be extended to random variables defined by probability densities (e.g. Gaussian densities), and can thus be generally employed to model the process with which sentient beings acquire knowledge about the world (Figure 2.1). If we accept the axiom that we can only build and update our beliefs about the (hidden) states of the world through our sensations, Bayes' inference describes the process by which pre-existing beliefs about the state of the world are updated every time we gather novel sensory evidence [Smith et al., 2021]. Pre-existing beliefs are represented by what is called a *prior* density that, in the Gaussian case, can be completely described by its sufficient statistics of mean and variance (which are then referred to as *prior parameters*). Beliefs about how worldly causes generate observable consequences are represented by a *likelihood* density, which express the probability of gathering the actual observation depending on the various possible worldly causes. Bayes' rule can then be viewed as combining prior and likelihood densities to compute an updated belief about the world (a *posterior density*) in the face of new evidence [Smith et al., 2021]. The posterior belief or knowledge about unknown causes θ given observations y can be described with the Bayes' rule as follows:

$$\underbrace{p(\theta|y)}_{\text{posterior}} = \frac{\overbrace{p(y|\theta)}^{\text{likelihood}} \overbrace{p(\theta)}^{\text{prior}}}{\underbrace{p(y)}_{\text{evidence}}} \quad (2.4)$$

Likelihood and prior together represent the *model* of the world, i.e., the joint probability of causes and evidence. The *model evidence* $p(y)$ is a constant obtained by marginalizing $p(y|\theta)$ over all possible θ values (and is thus also known as *marginal likelihood*), which is necessary to make the posterior a proper density (i.e., integrating to unity). Bayes' rule can also be graphically represented by the likelihood, prior and posterior probability distributions over θ , as in Figure 2.1, which is useful to illustrate the role of uncertainty, or inversely precision, of information. A lack of reliable knowledge about the world is represented by a prior distribution with a large variance (low precision), corresponding to a low

confidence about prior beliefs. Conversely, if prior knowledge is deemed precise and trustworthy, this translates into a “sharp” prior distribution. Similarly, if observations from the world are considered noisy and unreliable, the likelihood distribution would be characterized by a “flatter” shape (large variance). Crucially, the precision of prior and likelihood distributions determine their respective weights during Bayesian inference according to Bayes’ rule, and thus the outcome of the posterior [Smith et al., 2021]. In other words, the posterior can be more or less informed by the prior or the likelihood depending on their respective precision. For example, in the left panel of Figure 2.1, the prior has lower variance (higher precision) than the likelihood and the posterior is consequently “pulled” closer to the prior. The posterior is also “sharper” (more precise) than both the prior and likelihood, indicating an improved knowledge about the world. Conversely, in the right panel of Figure 2.1, the likelihood distribution is more precise than the prior, and the greater confidence afforded to it “attracts” the posterior towards it.

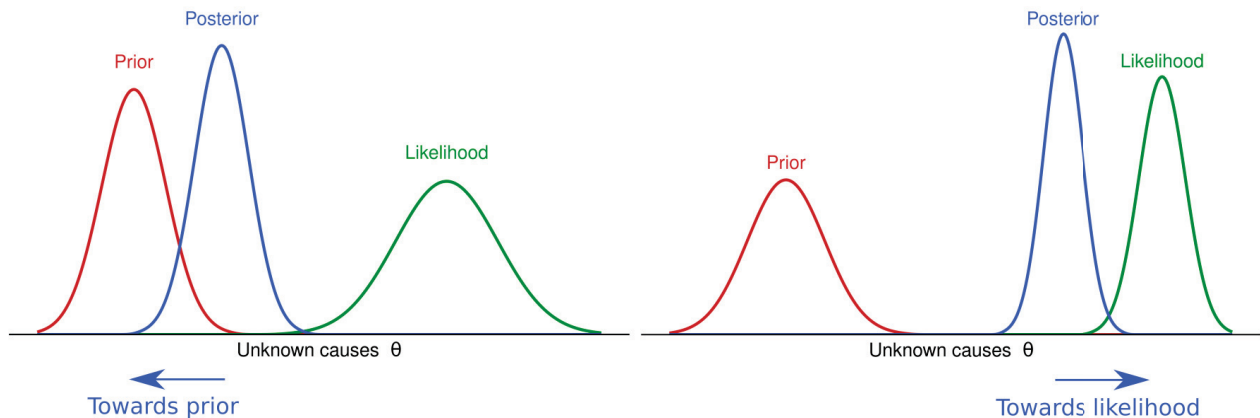


Figure 2.1: Schematic representation of Bayesian inference. Prior, likelihood and posterior are represented as normal distributions over unknown causes θ with given means and variances. (Left) When the prior has a lower variance (or higher precision) than the likelihood, the posterior is pulled towards it. (Right) In the opposite case, when the likelihood has higher precision than the prior, the posterior becomes more strongly informed by the likelihood.

The Bayesian brain hypothesis proposes that our brain actually performs approximated Bayesian inference and learning for a number of cerebral functions, from perception to cognition [Knill and Pouget, 2004]. Indeed our only link to events in the external world (i.e. unknown causes) is through sensory inputs (i.e. observations), and is thus constrained by our body and brain structures. Importantly, to disambiguate between similar or noisy percepts we can rely on previous knowledge accrued through past interactions with the world (i.e. priors), which in turn can bias our perception, as illustrated by perceptual illusions. Such process seems to approximate Bayes’ rule: $p(\text{causes}|\text{sensations}) \approx p(\text{sensations}|\text{causes})p(\text{causes})$. Notably, the need to deal with uncertainty accommodated by the probabilistic nature of the Bayesian brain hypothesis also arises from the presence of internal noise proper to sensory transduction and neuronal communication.

Knill and Pouget developed the notion of the Bayesian brain in an influential review based on numerous previous works [Knill and Pouget, 2004, Lee and Mumford, 2003]. According to their proposal, the brain accounts for different sources of uncertainty by encoding information in the form of probability distributions. Based on computational simulations, they claimed that such encoding could be implemented at the level of neuronal populations, although unequivocal empirical proofs of neuronal probabilistic processing are still lacking [Pouget et al., 2013]. At the cognitive level, multiple lines of evidence seem to validate the hypothesis that human observers behave in a (near) Bayes-optimal manner, not in the sense that their behavior is perfect or ideal — we all intuitively know that this is not the case — but because it conforms to the integration of prior knowledge and current evidence while properly accounting for uncertainty [Knill and Pouget, 2004]. At the perceptual level, for instance, human subjects integrate sensory cues from multiple sources (both within and across sensory modalities) in a Bayesian way depending on their respective uncertainty [Battaglia et al., 2003, Ernst and Banks, 2002, Körding and Wolpert, 2004, Knill and Saunders, 2003], potentially explaining bimodal illusory effects such as ventriloquism [Alais and Burr, 2004]. Moreover, Bayesian inference seems also to take place in higher-order brain functions, such as decision-making or language learning [Behrens et al., 2007, Tenenbaum et al., 2011].

2.1.2 The Free-Energy Principle

The Free-Energy Principle (FEP), developed by Karl Friston and colleagues, intends to describe how self-organizing systems (such as the brain, but also bacteria and social networks) are able to temporarily shield themselves from the tendency to disorder prescribed by the second law of thermodynamics by minimizing an upper bound on statistical surprise called free-energy [Friston et al., 2006, Friston, 2009, Friston, 2010], in agreement with earlier autopoietic accounts of biological organization [Maturana and Varela, 1980]. Statistical surprise, also known as “self-information” or “Shannon surprise”, is a measure of the unexpectedness of an outcome u given a statistical model of the world m . It is defined as the negative log-probability of this outcome: $-\log p(u|m)$ [Shannon, 1948] and is close to 0 when the outcome is expected or not surprising ($p(u|m) \approx 1$), and tends to infinity when the outcome is totally unexpected and thus very surprising ($p(u|m) \approx 0$). As expressed mathematically in [Friston, 2009], to minimize its entropy (disorder), an organism has to suppress surprise over time and so avoid surprising events according to its phenotype (e.g., being out of water is surprising for a fish but fully expected for a terrestrial animal). While tracking probabilities over all potential (hidden) states of the world is not biologically or computationally plausible, living organisms can minimize free-energy which is always greater than surprise and can be computed [Friston, 2009]. Interestingly, Friston hypothesized that free energy minimization can happen through two mechanisms: perception and

action. In most multicellular animals (at the exception of some marine invertebrate such as sponges) these biological processes are governed by a nervous system. Indeed, by sampling statistical regularities of its environment through sensory inputs processed by the brain, an organism can update “its” model of the world. This is actually Bayesian inference and learning supported by sensory perception and one can see how the FEP formulation subsumes the Bayesian brain hypothesis. Importantly and in convergence with enactivist formulations of embodied cognition [Varela et al., 1992], this view departs from classical cognitivist accounts by stating that the brain —and more generally its extension in the body as the nervous system— do not *have* a model of the world *per se*, that represents objects “out there”, but rather *is* itself this model, shaped by a life-long interaction with a co-constructed world, what enactivists named its “milieu” [Varela et al., 1992]. Such co-arising model can even be extended to the past-history of interactions between the ever-changing ecological environment and the adapting organism, namely phylogenetic selection, which was progressively hard-coded in the genome of the organism and which now defines its body —and more particularly brain— organization and the way it perceives the current environment. Free-energy minimization can then be seen as a long-run evolutionary process where accurately predicting and avoiding events of high surprise (e.g. a predator attack or other threats) would lead to a specie survival, and defaulting to do so, its extinction. Making the link with the Bayesian Brain hypothesis, in this scheme priors about expected states of the world —but also physiological states of the organism— can be innate when selected by natural selection, and/or acquired through learning during the organism life-time.

Coming back to free-energy minimization at the timescale of the organism, just knowing which events are surprising is not enough to actually avoid them, thus the FEP also describes how overt action allows to minimize free-energy. This was described as active inference and entails acting upon our body and/or the world to change the sensory signals we receive, so to avoid states of high surprise [Friston et al., 2010a, Friston et al., 2011]. Said otherwise, instead of changing “its” internal model to reduce surprise, a living organism also has the possibility to change its perceptual states (and doing so potentially the states of the world), so its predictions are fulfilled. In practice both perception and action are hypothesized to act in synergy to reduce free-energy.

If the FEP account seems theoretically valid, and could actually be an unifying framework for numerous theories and model of brain functioning [Friston, 2010], it does not tell us exactly *how* free energy minimization is implemented in the brain and we need to appeal to a more physiologically grounded formulation of Bayesian inference.

2.1.3 A neuronal implementation of hierarchical Bayesian Inference : predictive coding

In the 20th century we thought the brain extracted knowledge from sensations. The 21st century witnessed a ‘strange inversion’, in which the brain became an organ of inference, actively constructing explanations for what’s going on ‘out there’, beyond its sensory epithelia.

Karl J. Friston in *Does predictive coding have a future ?* (2018)

Initial proposals

Helmholtz’s insight of seeing “perception as unconscious inference” [Kahl, 1971] had a great influence on connectionist ideas starting from the 1980’s. Pioneering works gave rise to back-propagation learning [Rumelhart et al., 1986] and the so-called Helmholtz’s machine which set the start for the use of hierarchical generative models to explain brain functioning [Dayan et al., 1995]. In a seminal article, Rao and Ballard presented *hierarchical predictive coding* as encompassing two concurrent processes: a feedback stream of predictions flowing from higher to lower levels of the hierarchy, and a reciprocal feed-forward passing of prediction errors—the residual difference between predictions and actual inputs [Rao and Ballard, 1999]. This scheme presents the advantages of removing redundancy in neuronal transmission by discarding the predictable part of input signals, while being highly sensitive to deviations from statistical regularities of the natural world. Training computational networks implementing this algorithm on natural scenes images they could reproduce the hierarchical processing of basic to more complex features observed in ascending areas of the visual cortex, as well as some extra-classical receptive-field effects [Rao and Ballard, 1999]. A theoretical proposal of such reciprocal message passing within cortical areas was already proposed by Mumford based on physiological observations of cortical organization [Mumford, 1992], yet it lacked a concrete computational implementation and simulation which Rao and Ballard adequately provided.

Interestingly, from a mathematical viewpoint, Rao and Ballard algorithm had the advantage to update most of the model’s parameters of interest with gradient descents on a unique energy function which minimized prediction error. According to Karl Friston [Friston, 2018b], it is this simple yet far-reaching mathematical solution which prompted him to the groundbreaking nature of such predictive coding algorithm and lead to the profound paradigm shift he and others impulsed in cognitive neurosciences for the last twenty years. Indeed predictive coding as hierarchical inference in the brain

was further detailed and extended by Friston and colleagues [Friston, 2003, Friston, 2005, Friston, 2008], and integrated within the FEP [Friston and Kiebel, 2009, Friston et al., 2010b] through the translation of Rao and Ballard energy minimization function into variational free energy. An important refinement made by Friston and colleagues was to understand the predictive coding algorithm as implementing approximated Bayesian inference, thus making the link to the Bayesian brain hypothesis. Indeed, descending predictions from higher cortical areas can be intuitively understood as arising from a generative model of the world based on prior knowledge and are sometimes called prior expectations.

The predictive coding circuitry within the cortex

Any theory of brain functioning should be able to describe the neurocomputational mechanisms underlying perception. To do so it must be able to explain empirical findings at the neural level but also the supposed mechanisms should be neurophysiologically testable. In this sense, the mapping between computational implementations of predictive coding and the physiological structure of the brain was an important step for formalizing the FEP as an unifying brain theory [Friston, 2009, Friston, 2010]. Indeed the FEP supposes that to reduce surprise an organism must accurately predict future states of the world (including itself), and to do so maintain —or said differently embody and enact— a generative model of hidden causes in the world. Such model would necessarily be hierarchical or multi-layered to adequately mimics the causal interactions defining the physical world. Friston proposed early on that the brain, and in particular the cerebral cortex, possessed the structural organisation for embodying a hierarchical generative model of the world [Friston, 2003, Friston, 2005, Friston, 2008]. Similarly to Rao and Ballard algorithm, the brain would implement the transmission of backward predictions and forward predictions errors between cortical areas. Starting at the lower level of the hierarchy, in primary sensory cortices, for example A1 for an auditory stimulus, neuronal “error” units would compare sensory inputs from the thalamus to the signal predicted in “representational” units (also called expectation or prediction units) and informed by descending prediction from higher areas (secondary cortices, A2 for example). Only the residual “prediction error” would be transmitted upward to the next cortical level as a new input to be compared against downward predictions from the area above. This process, which is depicted schematically in Figure 2.2A, would be repeated time and again along the full cortical hierarchy with various relays depending on the sensory modality. As described originally by Mumford, such reciprocal passing of sensory information —which goes against the historical view of perception as a feedforward process— is plausible given the hierarchical organization of the cortex and its massive backward connections [Mumford, 1992]. Importantly to implement predictive coding per se, predictions units try to suppress or minimize prediction errors at the level below. They are also supposed to apply precision-weighted modulation over the post-synaptic gain of lower level er-

rors neurons through feedback projections [Feldman and Friston, 2010]. Perception is then thought to arise from the hierarchical minimization of precision-weighted prediction errors at multiple level of the hierarchy. Casting it in Bayesian terms, prediction errors from level $n-1$ (or sensory inputs at sensory cortices) can be seen as the “likelihood” of an observation, predictions from level $n+1$ are “priors” over hidden causes of this observation and their combination as a prediction error would be the “posterior” at level n , becoming the next “likelihood” for the level $n+1$.

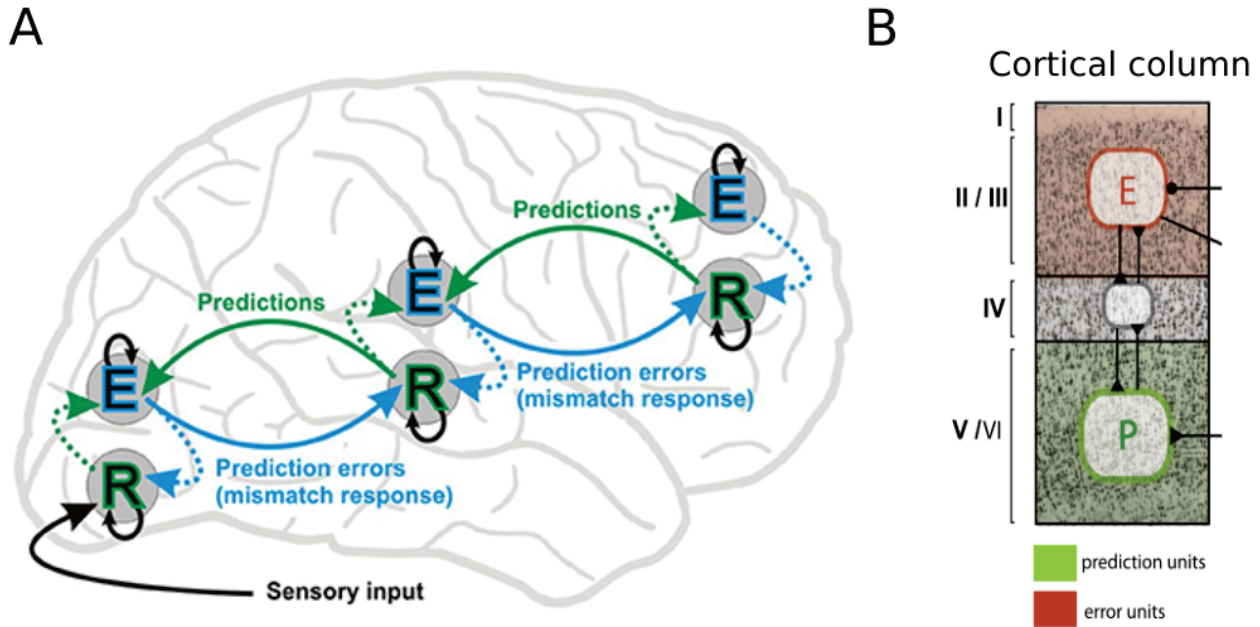


Figure 2.2: Simplified representation of hierarchical predictive coding within and between cortical areas (A) and its laminar organization within a cortical column (B). **A** - At each level of the hierarchy representational units (R, also called “state” or “prediction” units) would receive ascending inputs from the level below (sensory inputs in black for primary cortices or prediction errors in blue for higher areas) as well as lateral connections (dashed arrows) from error units (E) at the same level. Meanwhile error units are hypothesized to receive descending predictions (green) from the level above, as well as lateral predictions (dashed green) from representational units at the same level. The difference between descending predictions and ascending inputs gives rise to a PE which propagates upward, becoming an ascending inputs for the level above. Intrinsic inhibition by inhibitory interneurons is represented by black arrows above E and R units. This scheme is supposed to be repeated along the cortical hierarchy with a varying number of relay depending on the sensory modality (*Reproduced from [Stefanics et al., 2014]*). **B** - Within a cortical column, superficial pyramidal cells in supragranular layers (II/III) are thought to compose error units (E), while prediction or representational units (P) would be constituted of deep pyramidal cells in infragranular layers (V/VI). Inputs would be received in the middle granular layer (IV) of the cortex (*Adapted from [Heilbron and Chait, 2018]*)

Friston also described the potential laminar organization of error and prediction units within a cortical column [Friston, 2008]. The so-called canonical microcircuit for predictive coding was further detailed by Bastos et al. based on a review of cortical organization and projections found in electrophysiological studies [Bastos et al., 2012]. In a nutshell, superficial layers II and III of a given cortical area, also called supragranular layers, send most of their projections to the granular layer IV of higher-level areas, while they receive projections from the deep infragranular layers V/VI of these upward cortices.

Thus superficial pyramidal cells from layers II/III most likely form error units, encoding prediction errors which are sent forward in the cortical hierarchy (Figure 2.2B). In turn, deep pyramidal cells from layers V/VI would be the ideal candidates for representation units sending backward predictions. Intrinsic lateral connections within a cortical column mediate the integration of ascending inputs and descending predictions, which are also modulated by various excitatory and inhibitory interneurons, whose description is not essential to the current manuscript (see [Bastos et al., 2012] for the full depiction of the canonical microcircuit and its link to the FEP and predictive coding equations, and [Kogo and Trengove, 2015] for a critical review).

Interpreting attentional processes within the predictive coding framework

Friston and collaborators also added the critical modulation of prediction errors by their expected uncertainty based on previous literature emphasizing the importance of attentional modulation for proper handling of environmental uncertainty in Bayesian inference [Yu and Dayan, 2005]. Within the Bayesian framework, uncertainty can be understood as the average surprise over all possible causes of sensations — represented by a probability distribution (see section 2.1.1) — and can be measured mathematically by the variance of the distribution, or its inverse: precision. Thus in addition to predictions about the most likely causes of the sensations it receives, the brain would also maintain second-order predictions on the reliability of these sensations. Weighting ascending inputs by their respective expected precision would allow to select only informative inputs to be broadcast upward in the hierarchy as prediction errors. Cognitively this process of reliability attribution of sensory inputs would be mediated by attention, or in the words of Feldman and Friston (2010) : “attention is more concerned with optimizing the uncertainty or precision of probabilistic representations, rather than what is being represented” [Feldman and Friston, 2010]. Inputs (and later in the hierarchy prediction errors) deemed reliable are thus “up-weighted” and transmitted upward with more strength, whereas the influence of uncertain inputs would be down-weighted.

Interestingly, theoretical accounts have proposed that precision weighting for the perception of homeostatic signals, namely interoception, would have been hard-coded by natural selection to a fixed high precision [Millidge et al., 2022, Seth, 2013, Seth and Friston, 2016]. Indeed down-weighting physiological information about pain, hunger or thirst in favor of other non-vital but attentionally attractive signals would seem non-adaptative. Dysfunctioning of precision modulation in the brain may be at the basis of neuropsychiatric disorders such as schizophrenia, where patients are faced with hallucination (false perceptions) and delusions (false beliefs) [Friston et al., 2016], or autism characterized by hypo- or hyper-sensitivity and non-flexible priors [Haker, 2016]. Predictive coding could provide an unifying framework to understand those pathologies as well as other ones linked to cognitive and

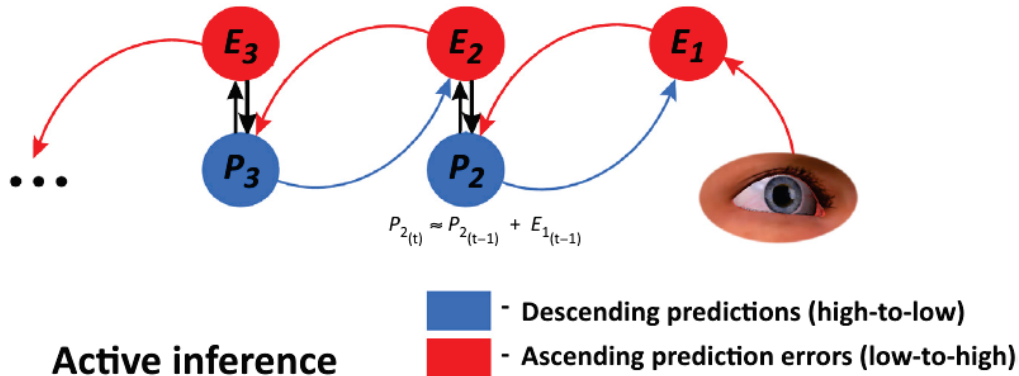
cerebral dysfunctions and thus could foster the design of novel therapeutic solutions.

From perceptual to active inference

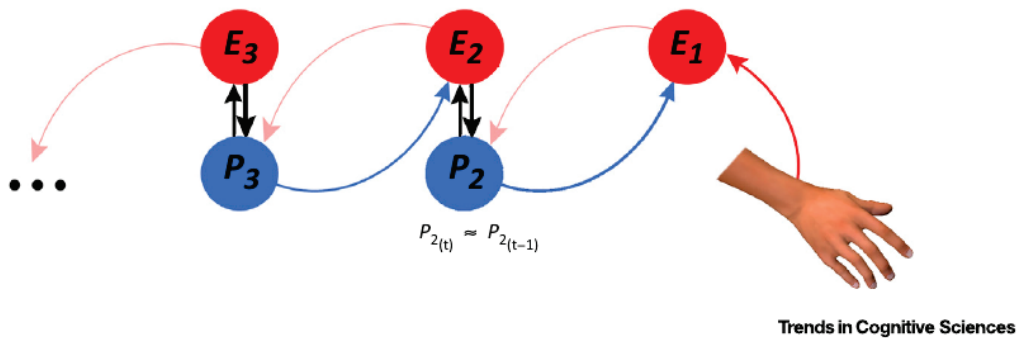
Attention as a precision-weighting mechanism would further mediate the transition from perceptual inference —where we sample information from the environment (perception) and updates our hierarchical model (learning)— to active inference where we directly act upon the world (which includes ourselves) to change the sensory inputs we receive (Figure 2.3). Indeed in active inference, descending projections from cortical motor areas to spinal cord motor neurons are not seen anymore as simple motor “commands” issued by a “central control system”, but as proprioceptive predictions about expected states (position, orientation, forces applied, etc) of the entire musculoskeletal system [Adams et al., 2013, Friston et al., 2010a, Friston et al., 2011, Brown et al., 2011]. Movements would then be the minimization of proprioceptive prediction errors between these predictions and the actual states of our limbs and muscles. However, for action to happen, “counterfactual” predictions errors about the sensory consequences of action need to have higher precision than sensory prediction errors arising from perceptual inference [Limanowski, 2017] (where counterfactual representations refer to beliefs about several alternative potential actions and the states of affairs that these actions would bring about). If not, the actual state of the body is taken as the expected one, no prediction errors are issued and immobility is maintained. Then attention plays a crucial role—in addition to the precision of proprioceptive predictions— by disengaging from sensory inputs when we perform a movement, thereby down-weighting sensory prediction errors [Limanowski, 2017]. Anyone can experimentally test this hypothesis by intensively paying attention to sensations in the legs while walking for example, movements usually become slower and less natural, or by trying to tickle oneself, as tickling sensations mostly come from the unexpectedness of tactile sensation which is abolished when you do (and so predict) it yourself [Blakemore et al., 2000]. More seriously the active inference explanation of motor control can account for sensory attenuation effects observed in laboratory settings during controlled movement [Brown et al., 2013], and its disruption in pathological conditions such as schizophrenia [Shergill et al., 2005, Shergill et al., 2014].

After this introduction of the Bayesian view of perception and its implementation in the brain as predictive coding, one might be thrilled by the potential explanatory power of such framework. Yet, we can wonder whether there is clear empirical evidence to support the view that the brain perceives the world in a Bayesian way and through predictive coding-like mechanisms.

Perceptual inference



Active inference



Trends in Cognitive Sciences

Figure 2.3: Simplified representation of hierarchical perceptual and active (Bayesian) inference in the brain. In perceptual inference (classical predictive coding), the system (our brain) is interested in gaining information from the environment, sensory information is cast forward in the hierarchy by prediction error units (E , red) and serve to update descending prediction signals generated by prediction units (P , blue). Prediction error minimization is achieved at each level of the hierarchy by integrating (prior) predicted information (P_{t-1}) to the ascending error (E_{t-1}) in a Bayesian way (modulated by their respective precision, not shown in the figure) to obtain posterior predictions (P_t). Instead, in the active inference mode, the system is interested in validating its predictions by changing the observed state of the environment (including itself). To do so the precision of ascending error signals is reduced, decreasing their weight in the inferential process at each level and insuring the “stubbornness” of predictions to updating ($P_t = P_{t-1}$), thereby allowing the system to reach the predicted state (movement). *Reproduced from [Yon et al., 2019].*

2.1.4 Empirical evidence for predictive coding in the brain and criticism

Despite the short history of predictive coding, numerous articles —mostly by authors not involved in its initial formulation— have already reviewed the behavioral, electrophysiological and neurophysiological experimental works supporting or disproving it [Heilbron and Chait, 2018, Millidge et al., 2022, Walsh et al., 2020, Clark, 2013]. If some predictions of the theory received strong support from the existing literature, certain assumptions are still a matter of debate because they are more difficult to test experimentally. Regarding the former, accumulative evidence support the claim that predic-

tion errors, mediated by feedforward activity of superficial pyramidal cells, would give rise (in part) to the electrophysiological activity observed with EEG and magnetoencephalography (MEG) [Friston, 2005]. Consistent with this idea, the averaged brain responses time-locked to perceptual stimuli (ERPs) during repetition suppression (RS) have been linked to the suppression of prediction errors upon multiple presentations of a repeating stimulus [Garrido et al., 2009b]. Conversely, the mismatch negativity (MMN), a change in ERP detected whenever a novel, unexpected stimulus is presented, has been linked to the failure of prediction error minimization [Garrido et al., 2009c]. Another important corollary of the predictive coding view is that brain responses should be influenced by the degree of expectations of stimuli [Todorovic et al., 2011, Todorovic and de Lange, 2012]. Such expectations are thought to be embedded in a hierarchical generative model of the world: higher areas bias perceptual processing at lower areas through top-down predictions. This phenomenon has been cleverly revealed by omission paradigms. In these paradigms, the brain responses time-locked to an expected stimuli (e.g. MMN) can be observed even in the absence of actual sensory input [Wacongne et al., 2011a] thereby demonstrating the major influence of prior expectations on perception as well as their top-down nature [Chennu et al., 2013a, Chennu et al., 2016a].

Evidence for the more general assumption of message passing within the cortical hierarchy have come from numerous studies in both humans and animals. An impressive series of studies from Garrido and collaborators used Dynamic Causal Modeling (DCM) on EEG data from various auditory odd-ball paradigms to identify the cortical network underlying deviance detection in humans (as measure through the MMN) and its reciprocal connections [Garrido et al., 2007b, Garrido et al., 2007a, Garrido et al., 2008, Garrido et al., 2009a, Kiebel et al., 2007]. In the auditory modality this network was best modeled by a DCM model composed of brain sources in bilateral primary cortex (A1), bilateral superior temporal gyrus (STG) and right inferior frontal gyrus (IFG), and their extrinsic connections, both forward and backward between each level, as well as intrinsic connections in A1. It was further refined more recently using DCM on EEG and MEG data from auditory local-global and omission paradigms [Chennu et al., 2016b], but also through MEG and electrocorticography (ECoG) [Phillips et al., 2015, Phillips et al., 2016] with an optimized paradigm using multiple deviance features ([Näätänen et al., 2004], duration, frequency, gap, intensity and location). The updated network found in this last MEG study is displayed in Figure 2.4. Compared to Garrido et al. initial model they found an additional involvement of left IFG and its modulation by higher-order predictions [Phillips et al., 2016], a difference potentially arising from the reconstruction of brain sources using MEG which is more precise spatially than EEG or the use of more complex deviance paradigms. Another study indeed found higher evidence for Garrido’s initial DCM model (bilateral A1 and STG but only right IFG) in EEG and MEG on a “classical” mismatch but a model with bilateral IFG and its bilateral modulation by higher-order predictions to win during an omission paradigm in EEG [Chennu et al., 2016b] (they

found no clear omission response in MEG and thus did not apply DCM). In conclusion of these DCM studies of the MMN, a global picture suggests that auditory deviance detection is supported by a bilateral (at least up to STG) hierarchical network with reciprocal extrinsic connections and intrinsic connections in A1, and that the particular involvement of the left IFG and higher-order predictions may depend on the complexity of the auditory deviance. Interestingly, this network was confirmed in two patients implanted with ECoG electrodes which allow a more direct recording of cortical activity than EEG/MEG [Phillips et al., 2016]. Additional evidence for hierarchical message passing of predictions and prediction errors along the auditory cortical pathway recently came from high-density ECoG on macaque and marmoset monkeys [Chao et al., 2018, Jiang et al., 2022]. Nevertheless recent reviews questioned whether these findings are unique to predictive coding and whether the used paradigms are sufficient to discriminate between competing hypotheses about the mechanisms of perception [Heilbron and Chait, 2018, Walsh et al., 2020].

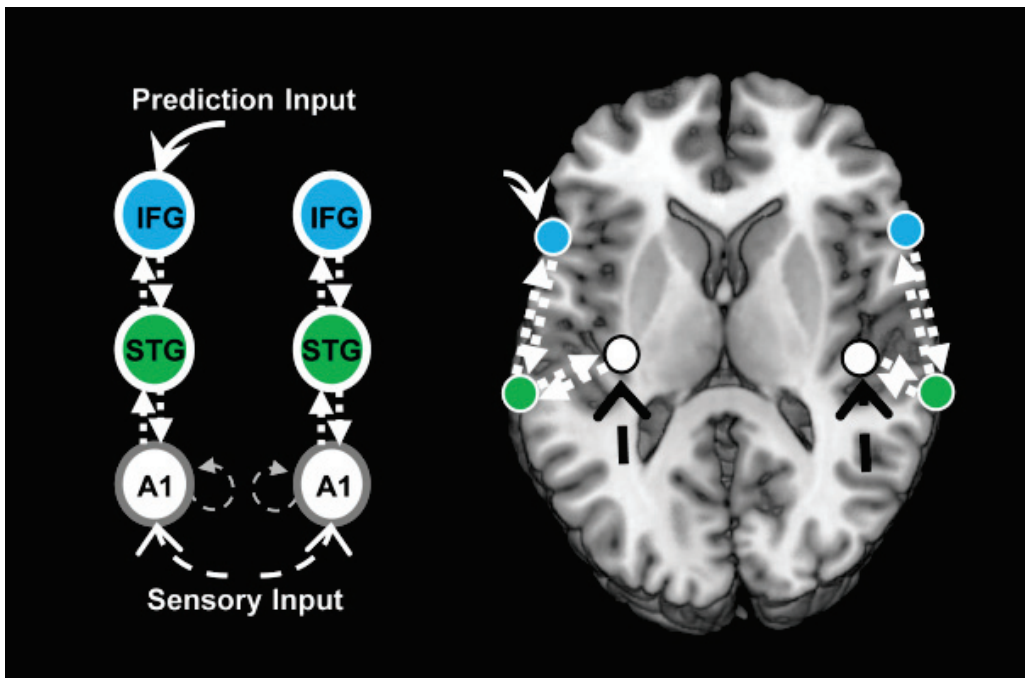


Figure 2.4: Hierarchical cortical network underlying predictive coding processes in the auditory modality, identified through DCM. The interplay between forward and feedback extrinsic connections within the auditory cortical hierarchy, as well as intrinsic connections within A1, is represented schematically on the left and the cortical sources located over an fMRI template of the brain on the right (transversal view). A1: primary auditory cortex, STG: superior temporal gyrus, IFG: inferior frontal gyrus. *Adapted from [Phillips et al., 2016].*

Another assumption, which was (ironically) less attended to, is the role of attention in weighting sensory inputs—and later along the hierarchy, prediction errors—according to their predicted reliability or precision. Basically attending to a certain sensory object of the environment (a sound) or a particular feature within this object (its pitch) would increase its expected reliability and prioritized its processing [Feldman and Friston, 2010]. For example in the visual domain this process of sensory gating is hypothesized to be mediated by cortico-thalamic connections between visual cortices and the

pulvinar, a large nucleus of the thalamus which is involved in visual attention [Kanai et al., 2015]. The hypothesized increase in neural response to attended stimuli is a well observed fact in the neuroscientific literature ([Brown and Friston, 2013, Auksztulewicz and Friston, 2015a, Gordon et al., 2019, Smout et al., 2019a] and see [Feldman and Friston, 2010, Hohwy, 2012] for reviews), yet predictive coding has been criticized for trying to explain any form of attention when it actually may not be able to do so [Ransom et al., 2017]. Besides the particular assumption that predictability and attention should have interactive effect on brain responses—a predicted stimulus evoking a smaller prediction error while attention is boosting it—received contrasted evidence ([Chennu et al., 2013a, Chennu et al., 2016b, Garrido et al., 2018, Kok et al., 2012, Richter and de Lange, 2019] and [Heilbron and Chait, 2018, Walsh et al., 2020] for critical reviews).

Finally a last critical assumption of predictive coding is that prediction and prediction errors are encoded by different neuronal populations, representational (or prediction) units and error units respectively. Two key features implemented in the canonical microcircuit would allow to differentiate them experimentally: their laminar location within the cortex, supragranular layer (II/III) for error units and infragranular layer (V/VI) for prediction units; and the frequency range at which they operate, gamma (> 30 Hz) and alpha/beta (10-30 Hz) respectively. While this latter difference in frequency range between superficial and deep cortical layers is a well-established finding both in animal and human (through ECoG, notably), the laminar segregation of prediction and error units is more difficult to test and therefore only received scarce and sometimes contradictory proof (see [Milledge et al., 2022, Heilbron and Chait, 2018, Walsh et al., 2020, Shipp, 2016] once again for critical reviews). Furthermore predictive coding also received criticism concerning the organisation of the canonical microcircuit at a theoretical level. For example Kogo and Tengrove demonstrated the biological implausibility of some of the algorithmic computation supposedly performed by prediction and error neuronal population [Kogo and Tengrove, 2015]. Actually different neuronal implementation of predictive coding were proposed which may not encounter these problems. For example Spratling, in his predictive coding model based on biased competition in visual attention, advocated early on for a different laminar organisation of predictive coding in the cortex, with error units in granular layer IV and prediction units in both superficial and deep layers and therefore top-down and bottom-up passing of predictions signals (see [Spratling, 2008a, Spratling, 2008b, Spratling, 2017] and Figure 2.5). Critical voices were also raised concerning the “corticocentrist” view of predictive coding (at least in the auditory modality), when numerous studies in animals demonstrated the importance of top-down subcortical connections for perception (see [Asilador and Llano, 2021] for a review).

In summary, we saw that classical views of predictive coding make precise predictions on how algorithmic computations necessary for hierarchical inference should be implemented in the brain. Yet clear and uncontroversial evidence to support these assumptions are still scarce and could also be

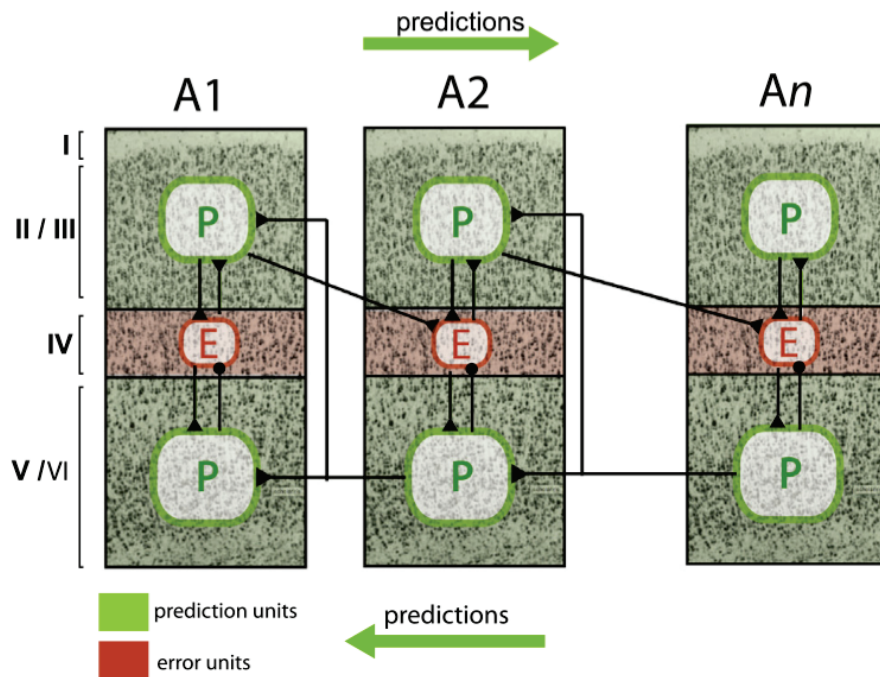


Figure 2.5: Simplified representation of cortical organisation of predictive coding under the biased competition model in the auditory pathway. In this view error units are situated in the granular layer (IV) while prediction or representational units are located in both superficial (II/III) and deep (V/VI) layers. Consequently predictions are thought to be broadcast both backward and forward between cortical areas and errors are suppressed by prediction units within a cortical area through intracolumnar inhibition. A1: primary auditory cortex, A2: secondary auditory cortex, An: higher auditory areas. *Reproduced from [Heilbron and Chait, 2018].*

explained by other theoretical views of perception. Hopefully recent advances in neuroimaging as well as electrophysiological and cellular recordings would most likely allow to decide between competing hypotheses in a near future and potentially refined unclear aspect of this major neuroscientific theory. Indeed predictive coding (and the underlying FEP) is so far the only neuroscientific theory allowing a global understanding of animal and human perception and we will now dive more deeply in its application to two crucial topics of the current thesis: auditory deviance detection and pain experience.

2.2 Auditory deviance detection revisited through predictive coding

Whether we are awake or in an “unconscious” state (sleep, anesthesia or coma), our brain automatically perceives and integrates auditory stimuli that are presented to it, for example simple tones in an experimental protocol. The global electrical activity arising from brain functioning can be measured at the surface of the skull, the scalp, through EEG or MEG for example. To actually measure the specific activity related to the perception of a given sound, one has to separate it from the ambient “noise” of the brain, any activity not related to sound processing. By presenting a stimulus an important number of times (50 for example) and averaging the resulting brain activity time-locked to the stimulus onset across all repetitions, one typically cancels out the noise (whose fluctuations are random) and obtains an event-related potential (ERP): the averaged brain response to an event, in our current example an auditory tone. In the case of human auditory perception this ERP is constituted of a typical negative deflection peaking at fronto-central scalp electrodes around 100 ms after the tone onset and thus called the N100 or N1. The N1 is actually composed of multiple sub-components whose cortical generators were identified as the primary auditory cortex (A1) and the superior temporal gyrus (STG) [Näätänen and Picton, 1987]. Of interest to us is what happens when a deviance is introduced in a sequence of repeating tones at the same frequency (e.g. 1000 Hz), for example by presenting a tone at a larger or lower frequency (e.g. 1100 Hz), a paradigm called an auditory “oddball” (see Figure 2.6 for an example of sequence). Surprisingly the ERP to this rarely appearing “deviant” tone is different to the one of the frequent “standard” tones, notably it presents a larger (more negative) N1, as observed initially by Näätänen and collaborators ([Näätänen et al., 1978], see Figure 2.6 for typical ERP waves of an auditory oddball paradigm). They named the negative wave obtained from the subtraction of deviant and standard ERPs the “Mismatch Negativity” (MMN) and interpreted it as a neural correlate of automatic deviance detection [Näätänen, 1990]. The MMN is better observed at fronto-central electrodes in EEG and usually peaks between 100 and 200 ms after the deviant tone onset with variation depending on the deviance feature (frequency, intensity, duration, gap, spatial location, etc). Its generative mechanisms have been the subject of a long and intense scientific debate for more than 20 years, that we will now review.

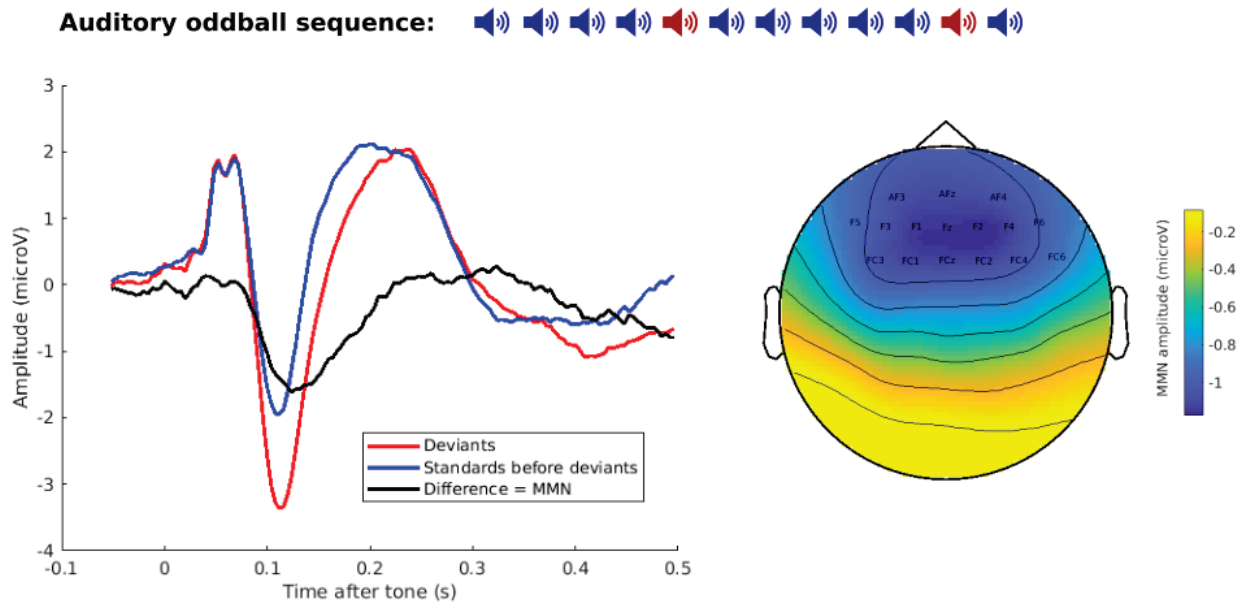


Figure 2.6: Example of auditory oddball sequence and the resulting ERPs to deviants, standards and their difference as the MMN. Oddball sequence are constituted of two tones (or more in the case of a roving oddball, see [Haenschel et al., 2005]) differing in a specific auditory feature (e.g. frequency) and one tone presented more frequently than the other (e.g. 80% of standard tones). Grand average ERPs measured with EEG are presented for standards at 880 Hz (blue), deviants at 988 Hz (red) and their difference: the MMN (black). One can observe the N1 to standards and deviants peaking around 100 ms (0.1s) and the MMN peaking a bit later around 140 ms. The EEG scalp topography averaged over the MMN window (MMN significantly different from 0, here from 84 to 192ms) is presented on the right with a negative cluster at fronto-central electrodes. *Based on data from [Fucci et al., 2021].*

2.2.1 The Mismatch Negativity (MMN): a neural correlate of prediction error ?

Memory-based and model-adjustment accounts

Early accounts of the MMN cast it as a marker of sensory memory processes [Näätänen, 1990]. The rationale was that auditory change detection requires a memory trace of previously encountered sounds to compare to the current input. However, while the so-called *memory-based account* can explain some properties of the MMN, it is unlikely that the detection of abstract deviations [Paavilainen, 2013] can be based solely on short-lived sensory memory traces. To address this limitation, early proponents of the memory-based hypothesis developed a refined account of the MMN under the *model-adjustment hypothesis* [Winkler et al., 1996, Winkler, 2007].

In this view the main mechanism behind the MMN is not only the detection of changes in some physical features of a stimulus, but more generally the detection, representation and updating of regularities in the auditory stream. Being more flexible, the model adjustment account can explain most of the MMN properties, from omission responses to violations of abstract rules. Besides, dynamic model

updating addresses the challenge of adapting to the frequent contextual changes present in natural environments [Sussman and Winkler, 2001]. Inasmuch as it sees the MMN as a marker of error detection and correction with respect to a perceptual model of statistical regularities in the world, the model-adjustment hypothesis is also compatible with the predictive coding account of perception presented below.

Adaptation hypothesis

Early on, defenders of the sensory-memory account acknowledged the possibility of the MMN being an epiphenomenon of differential neuronal adaptation between frequent standard and rare deviant tones. Indeed, due to stimulus-specific adaptation (SSA), neurons tuned to frequently repeated tones would adapt faster than neurons tuned to rare deviant tones, causing their N1 to be attenuated compared to the N1 response to novel fresh stimuli, hence the presence of a negative deflection in the subtraction between the ERP to deviant and the ERP to standard tones (Figure 2.7). This alternative interpretation has long been discarded in light of the evidence available at that time [Näätänen, 1990], namely that (i) MMN and N1 latencies do not match, (ii) they have different cortical generator locations and (iii) adaptation cannot account for mismatch to sound omissions. More recently, these arguments against the adaptation hypothesis have been undermined by simulation [May and Tiitinen, 2004, May and Tiitinen, 2010] and experimental [Jääskeläinen et al., 2004] findings suggesting that (i) the delayed latency of the MMN compared to the N1 potential can be explained by lateral cortical inhibitions between feature-specific neuronal populations [May and Tiitinen, 2004]; (ii) the seemingly shifted location of the MMN can result from differential adaptation of anterior and posterior auditory cortex N1 sources [May and Tiitinen, 2004, Jääskeläinen et al., 2004]; (iii) cortical neurons responsible for the N1 response settle in a steady-state oscillation that does not stop immediately after the end of the stimulus train [May and Tiitinen, 2001], thus providing a plausible mechanism for the presence of a mismatch negativity in response to omitted sounds. At the same time, the adaptation hypothesis gained plausibility with the demonstration of SSA processes in the auditory cortex in response to repetitive sounds, both in humans [Jääskeläinen et al., 2004] and animals [Ulanovsky et al., 2003, Ulanovsky et al., 2004].

As May et al. [May and Tiitinen, 2010] highlighted, memory-based and adaptation accounts fall in two different categories: psychological and physiological, respectively, which are not mutually exclusive explanations of the MMN. On the other hand, both accounts stumble over some properties of the MMN such as its sensitivity to abstract rules, justifying further efforts in the development of a more encompassing theory of the MMN.

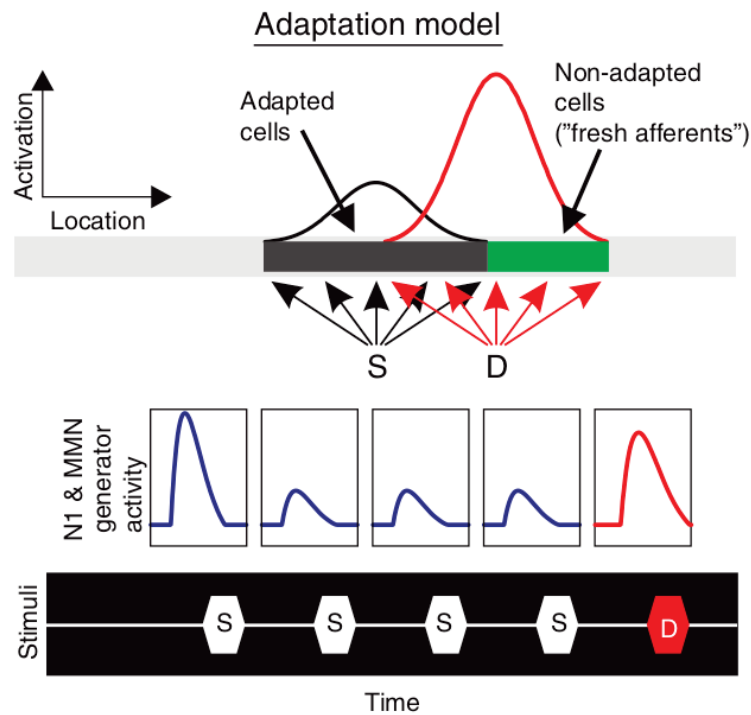


Figure 2.7: Schematic representation of the adaptation hypothesis of the MMN. [Upper panel] In the adaptation account the N1 and the MMN are thought to arise from the same cortical generator. Thus, overlapping neuronal populations answer to standard and deviant tones depending on their favorite frequency. After many standard tone repetitions, neurons tuned to the standard frequency become adapted, meanwhile neurons tuned to the deviant frequency adapt less, this is the SSA phenomenon. [Lower panel] At a global scale (EEG or MEG recordings for example), the brain response to a first standard is high but then adapts quickly. Upon presentation of a rare deviant stimulus, its response is higher than the preceding standard due to non-adapted “fresh” neurons, the MMN is then interpreted as an “enhanced N1 response”. *Reproduced from [May and Tiiinen, 2010].*

Predicting coding: a unifying framework?

Under the predictive coding theory, the MMN is viewed as an electrophysiological marker of the discrepancy between top-down predictions based on repeated stimuli (i.e. standards) and bottom-up unexpected sensory inputs (i.e. deviants): a prediction error [Friston, 2005]. Propagated through forward connections, prediction errors would trigger the updating and optimization of an internal model of sensory causes at higher cortical areas, in agreement with the model-adjustment hypothesis. Adjusted predictions would then be sent backward to explain away or suppress prediction errors at lower cortical areas. In this framework, suppression of the MMN when deviant tones are repeated and thus become the new standards, as in the roving paradigm [Haenschel et al., 2005], is interpreted as a minimization of prediction errors, the internal model having learnt a new statistical regularity [Garrido et al., 2009c]. Additionally, hierarchical minimization of prediction errors would also be weighted by the relative precision of predictions and sensory inputs, depending, as seen previously, on their expected reliability. Physiologically, changes in local, intrinsic connectivity of prediction error units would mediate this weighting process in auditory cortices via modifications of synaptic gains,

reminiscent of the adaptation account [Garrido et al., 2009c].

Using Dynamic Causal Modelling (DCM), Garrido and colleagues showed that model-adjustment or adaptation hypotheses alone are not sufficient to explain changes in extrinsic and intrinsic connectivity specific to the MMN in both the classical and roving auditory oddball paradigms [Garrido et al., 2007c, Garrido et al., 2007a, Kiebel et al., 2007, Garrido et al., 2009a, Garrido et al., 2008]. In an attempt to unify, [Garrido et al., 2009c] they proposed that both hypotheses are necessary to fully account for N1 and MMN brain responses, and are reconcilable under predictive coding. More recently, two extensive and complementary reviews further emphasized this viewpoint. The first one from Fitzgerald and Todd (2020) relied on their studies of first-impression or primacy bias of the MMN through the multiple-timescale paradigm in humans [Fitzgerald and Todd, 2020]. This paradigm implements two kinds of regularity: a local regularity associated with tone probabilities and a long-term one associated with block length. Surprisingly, they showed that the MMN to deviant sound A was modulated differently by long-term regularities (short vs. long blocks or stable vs. unstable sequences), depending on the initial local regularity pattern (sound A initially presented as deviant or standard), and that this effect was present for different kinds of deviance (duration and frequency) [Fitzgerald and Todd, 2018, Todd et al., 2014]. The authors came to the conclusion that such order-driven modulations of the MMN cannot be explained by adaptation alone but need predictive coding to account “for the influence of tightly held, top-down representations of sounds on future sound processing” [Fitzgerald and Todd, 2020].

A second review focused on single-cell studies of the SSA phenomenon in animals, the hypothesized neural correlate of the MMN according to the adaptation hypothesis [Carbajal and Malmierca, 2018]. In a nutshell, Malmierca and collaborators could dissect the neuronal response traditionally associated to deviance detection into a repetition suppression component reflecting adaptation processes, and a prediction error component expected under predictive coding. To do so, they made use of newly created “no-repetition” control sequences for the oddball paradigm, namely the “many-standards” where tones of varying frequencies, whose the oddball “standard” frequency, randomly alternate; and the ascending and descending “cascades” sequence where a tone of the same frequency as the deviant is heard within a sequence of ascending or descending frequency tones (see [Parras et al., 2017] for a detailed description). Their results prompt toward a segregation of repetition suppression and prediction errors mechanisms across the auditory pathway, with the former more prevalent in subcortical regions (midbrain and thalamus) and prediction errors-like neuronal activity gaining strength in auditory cortex [Parras et al., 2017] until being predominant in prefrontal cortex [Casado-Román et al., 2020]. Moreover they could show that this distinction also held within each regions of the hierarchy, with stronger prediction error-like activity in neurons of the non-lemniscal (or secondary) auditory pathway compared to the lemniscal (or primary) one. In summary, they interpreted their findings in light of the supposed

hierarchical organization of predictive coding and, as Garrido et al. [Garrido et al., 2009c] before, proposed a unification of previously opposing views of SSA and MMN under an integrated “physiological generative mechanism of hierarchical inference founded on plastic changes in synaptic connectivity”. Besides they also emphasized that their results supported the growing view of predictive coding has not being restricted to the cortex [Carbajal and Malmierca, 2018].

Evidence for Bayesian views of the MMN were also found in multiple computational modeling studies [Ostwald et al., 2012, Lieder et al., 2013, Wacongne et al., 2011b, Lecaigard et al., 2021], yet without decisive conclusion against the adaptation hypothesis. Using two EEG datasets (final sample size of 82 subjects) from similar auditory oddball paradigms, we compared again computational models inspired from both theories using trial-wise as well as time-resolved Bayesian modeling. Interestingly and in agreement with previously mentioned animal studies we found a temporal distinction in generative mechanisms of the MMN, with adaptation models better explaining the early part of the MMN and Bayesian learning models predominantly explained the later latencies of the MMN. These results constitute a submitted article presented in the current thesis as Study 2 (Chapter 4).

2.2.2 The effect of attention on the MMN

The interplay between the MMN and attentional processes has also been a debated topic since the first reviews on the mismatch response [Näätänen, 1990, Näätänen, 1992] where it was qualified as “pre-attentive” as it could be elicited without any attentional engagement to the auditory stream. If, indeed, attention is not necessary for the MMN generation (e.g. MMN found in unconscious patients [Morlet and Fischer, 2014]) it was still shown to increase the MMN response [Woldorff et al., 1991, Woldorff et al., 1998, Näätänen et al., 1993]. However, these earlier results have been questioned as they included highly focused attention condition and for some of them were most likely conflated by the N2b ERP component, a negative-going wave elicited by target detection in attend conditions at the same latencies as the MMN [Sussman, 2007]. Indeed experimental protocols which manage to dissociate N2b and MMN components found no difference in the MMN between ignore and attend conditions [Sussman, 2007]. These observations led Sussman (2007) to propose that there is no general effects of attention on the MMN, but rather that these effects are specific to the attentional task performed by the subject. Specifically, Sussman et al. (2002) showed that attention might influence the standard formation phase of the MMN but not deviance detection per se, as attending to a global pattern of tones led to the disappearance of the MMN to local deviant tones [Sussman et al., 2002]. They still advocated against the qualification of the MMN as “pre-attentive”, but emphasized that the MMN is highly context-dependent and that this context can be shaped or not by attention [Sussman, 2007], a view further corroborated by recent first-impression bias results on the MMN [Fitzgerald and

Todd, 2018, Todd et al., 2021]. In turn, the MMN itself is sometimes described as a signal for the reallocation of attentional resources to informative sensory streams [Damaso et al., 2015, Näätänen et al., 2011]. This explanation accommodates well finding of poorer performance in attentional tasks following deviants from an irrelevant oddball sequence for example, but again seems to depend on the context as highly demanding tasks do not show this effect but instead a reduced MMN [Damaso et al., 2015]. The proposal suggests a varying threshold for attentional shifting which can allow redirection of attention to highly informative (surprising, high MMN) signals, while “protecting” attentional resources when the primary task is important and attentionally demanding [Damaso et al., 2015, Näätänen et al., 2011].

We described previously how the MMN is interpreted as a neural correlate of prediction error signaling within the predictive coding framework [Garrido et al., 2009c]. We also saw how attention was integrated within this framework as a modulation or weighting of expected precision associated to ascending sensory and prediction error inputs [Feldman and Friston, 2010]. Heightened attention cast onto sensory signal would lead to higher expected reliability and therefore stronger prediction errors. This view is not opposite to the one of Sussman (2007), as depending on the context attention may enhance the precision afforded to local deviations or instead to global regularity patterns. Empirical evidence seems to support an attentional enhancement of prediction error signals in both auditory [Aukstulewicz and Friston, 2015b, Chennu et al., 2013b, Garrido et al., 2018] and visual domains [Smout et al., 2019b, Gordon et al., 2019].

Given the effects of meditative practices on attention (section 1.2.3), we could expect that they might influence the MMN. Hypotheses concerning the direction of a potential effect and its variation depending on the kind of practice and the expertise of practitioners would be discussed in Chapter 3.

2.3 Does pain perception follow Bayesian principles ?

As we saw in a previous section (1.2.4), pain experience has — most often but not always — a sensory dimension which makes it a perceptual information. Pain perception is based on the nociceptive system which provides an immediate information about a dangerous event and the necessity for action. As such it is an ideal candidate for a mechanistic explanation based on Bayesian inference and the minimization of bodily threatening surprise. In the coming pages we will review Bayesian or predictive coding accounts of pain and how they can provide parsimonious explanations to an extensive body of empirical work on the influence of priors and uncertainty on pain, as well as dysfunctioning of pain perception in clinical cases.

2.3.1 Bayesian accounts of pain perception

A first theoretical perspective on pain perception from a Bayesian view point was given only recently by Büchel and colleagues, with an emphasis on placebo hypoalgesia [Büchel et al., 2014]. Placebo effects can be defined as the clinical benefits obtained from a medical intervention where the key active treatment (e.g. a drug, medical device, etc) has been replaced by a passive one without any physiological effects. The remaining active effects are the psycho-social ones, which include the participant's beliefs in the benefices of the treatment and the medical context [Tracey, 2010, Finniss et al., 2010]. Placebo interventions have been mostly used in clinical trials as a control to assess the added benefits from new interventions. However, and contrarily to common misunderstandings, they do have positive outcomes that go beyond a mere report bias. When evaluated with adequate designs, on diseases amenable to psychological factors and controlling for confounding factors, placebo treatments do have an effect of their own on both subjective and objective measures, which can sometimes approach the effect size of real treatments [Wampold et al., 2005]. The scientific literature gave particular interest to placebo treatments for pain, an effect referred to as placebo hypoalgesia. While placebo hypoalgesia has psycho-sociological causes, mainly positive outcomes expectations, it is mediated by at least one specific neurophysiological pathway: the activation of the opioidergic descending pain control system [Wager and Atlas, 2015, Tracey, 2010, Petrovic et al., 2002]. This pathway in turn inhibits the processing of nociceptive inputs at low-levels of the nervous system, therefore decreasing the activation of pain specific brain regions and reducing painful experience. The added value of Büchel et al. (2014) theoretical work has been to explain how expectation on painful stimuli might be combined to nociceptive inputs in a Bayesian way to give rise to hypoalgesic effects. Using the vocabulary introduced in section 2.1.1, expectations of lower pain due to placebo (for example from an inactive drug described by the physician as a pain killer) constitutes a prior on pain, most likely issued at higher brain levels

of the “descending” pain system [Büchel et al., 2014]. Priors would be combined at different levels of the pain processing hierarchy with nociceptive signals –constituting the “likelihood”– sent by afferent nociceptors through numerous “ascending” pain pathways. Similarly as in other sensory modalities, pain perception would thus arise from Bayesian inference over prior and likelihood to give rise to a posterior, the actual pain experience. In mathematical notation:

$$\overbrace{p(\text{pain experience}|\text{sensory signals})}^{\text{posterior}} \approx \underbrace{p(\text{sensory signals}|\text{pain expectations})}_{\text{likelihood}} \overbrace{p(\text{pain expectations})}^{\text{prior}} \quad (2.5)$$

Büchel et al. (2014) also provided an hypothetical neural implementation of predictive coding mechanisms in pain related brain regions. In the same vein as the convolution of multiple “pain matrices” proposed by Garcia-Larrea and Peyron [Garcia-Larrea and Peyron, 2013] and in adequacy with the hypothesized hierarchical implementation of predictive coding in the brain [Friston, 2005], they suggested to reformulate the dichotomy between independent pain ascending and descending pathways into a single hierarchical system constituted of connected recurrent networks along the pain processing hierarchy. Their proposal rests on a growing body of electrophysiological evidence for reciprocal connections within the pain “descending” modulatory system, as well as observations of pain modulatory signals through fMRI at all levels of the cerebral hierarchy [Büchel et al., 2014]. They also evoked the possibility that in the final segment of the modulatory pain system –constituted of Periaqueductal gray (PAG), rostral ventromedial medulla (RVM) and the dorsal horn of the spinal cord–opioids may not only have an analgesic effect, but also regulate the influence of top-down predictions on pain perception.

Such Bayesian understanding of pain has been further extended to pain perception in general [Wiech, 2016] as well as placebo effects ([Tabor et al., 2017], how negative expectations about the painfulness surrounding an actually inert experimental treatment or stimulus lead to increased pain). Those converging accounts agree on the idea that the brain must maintain a model of expected pain depending on previous painful experience. In relation to the pain matrix, prior expectations might be modulated by affective and evaluative factors, for example a negative mood might set a higher than usual prior for incoming painful experiences. Predictions from this pain model are then compared to nociceptive inputs to emit a prediction error or “surprise” signal which, depending on its strength, may lead to the update of the model, allowing the learning of cue-pain contingencies for instance (Figure 2.8).

Another important aspect of predictive coding formulations of perception is the influence of preci-

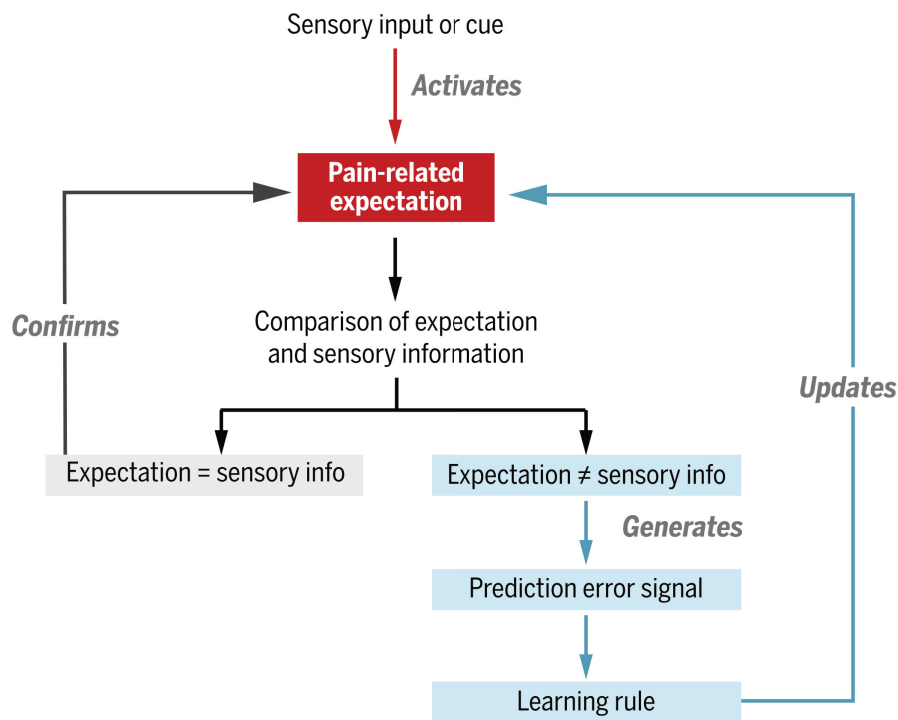


Figure 2.8: Schematic representation of pain processing from a predictive coding perspective. Illustration of the comparison between pain expectations (predictions arising from prior information about pain) and sensory input. The process can give rise to an update of the brain model (learning) about incoming pain if signaled by high prediction errors, or not if predictions accurately reflected sensory information. *Reproduced from [Wiech, 2016]*

sion or uncertainty on the Bayesian inference process. In the same way as for other kind of perceptual processes, the precisions of prior and sensory information about pain weight their respective effect on the final pain experience. Büchel et al. described well how a more or less accurate description of a placebo treatment can lead to increased or diminished placebo effect ([Büchel et al., 2014], illustrated in Figure 2.9). In the opposite nocebo effect, a physician providing a very confident prognosis on the highly painful evolution of a disease could explain a potentially amplified pain experience [Tabor et al., 2017].

Related to nocebo, chronic pain is the transformation of an initial acute pain event or disease into a painful experience persisting after the commonly observed healing phase [Apkarian et al., 2009]. Despite being one of the primary cause of long-term disability worldwide, the physiological and brain mechanisms causing and maintaining it are still poorly understood and very few effective treatments exist [Bushnell et al., 2013]. The change of acute pain into a chronic state has been associated to peripheral and central sensitization processes [Cagnie et al., 2014, Kuner, 2010, Gold and Gebhart, 2010] most likely mediated by a dysfunctioning of glial cells [Ji et al., 2013], but also more specifically to abnormal structure and activity of the dorsolateral Prefrontal Cortex (dlPFC) [Seminowicz and Moayedi, 2017]. This cortical region has been identified as a core component of many different functional networks, it notably seems to play a crucial role in pain detection and modulation. Under

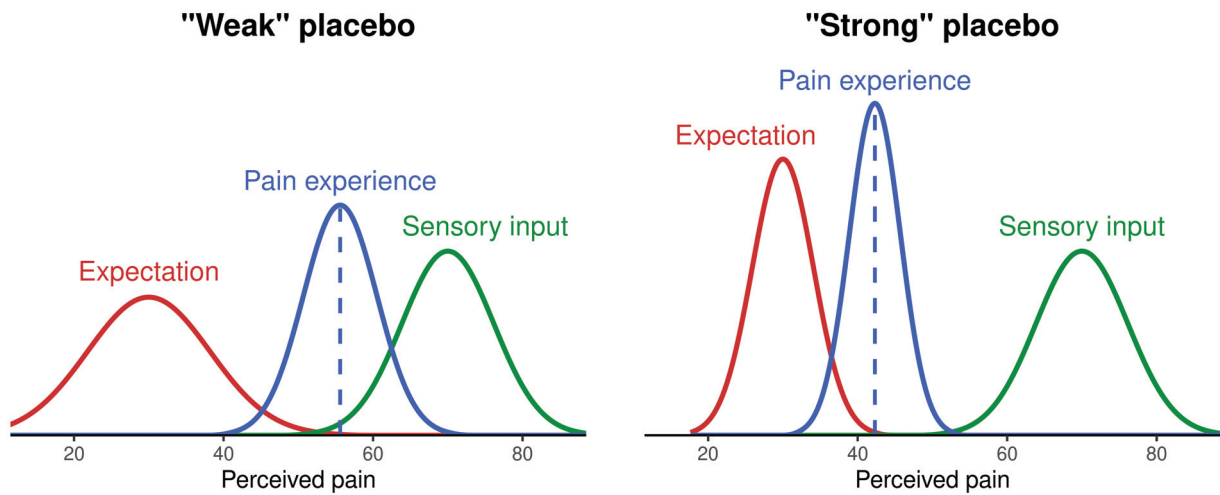


Figure 2.9: Bayesian interpretation of precision or uncertainty effect on placebo. In both graphics the means of prior (expectation) and likelihood (sensory input) are fixed, and only the variance of the prior information is changed. According to [Büchel et al., 2014], if a placebo intervention is given without much precision on its efficacy (“weak” placebo), it gives rise to a broadly defined expectation for lower pain and thus a final posterior perceived pain (pain experience) close to the initially high painful sensation (sensory input). Instead if physicians give precise suggestions for the positive effect of the intervention (“strong” placebo), the resulting expectation for lower pain will have stronger influence on the Bayesian inference process and leads to lower perceived pain.

chronic pain, not only its activity and connectivity are altered (higher activation and mixed connectivity pattern to other regions), but also its structure with an observed reduction of gray matter volume [Seminowicz and Moayedi, 2017]. Of clinical importance, non-invasive stimulation of the dlPFC has been shown to reduce acute and chronic pain [Moisset et al., 2016], making of this brain region an interesting therapeutic target. Chronic pain conditions have also been associated to other nodes of pain processing and modulation, with notable structural decrease in gray and white matter of ACC and insula as well as reduced opioid receptors binding [Bushnell et al., 2013].

Recently, Bayesian accounts of this leading health problem have also been proposed. A short perspective on the topic was first offered by Moseley and Vlaeyen in 2015, where they developed the “imprecision hypothesis” [Moseley and Vlaeyen, 2015]. In their view, chronic pain develops from an initial low-precision encoding of the acute pain event (for example in low back pain, bending forward) which may lead to the generalization of a high pain expectations for usually not painful events involving similar movements. This novel idea was further articulated in predictive coding terms [Hechler et al., 2016, Tabor and Burr, 2019, Jones and Brown, 2018] and two separate processes were suggested to cause chronic pain. Firstly an aberrant heightened precision given to predictions of body threat, and wrongly generalized to non-threatening situations through fear learning [Zaman et al., 2015], a view akin to Moseley and Vlaeyen’s “imprecision hypothesis”. Secondly a disproportionate weight given to sensory information, either painful or not. If the link to underlying physiological causes still have to be tested, one can hypothesize that aberrant pain expectations and associations might be

due to dlPFC dysfunction [Seminowicz and Moayedi, 2017], while excessive sensory precision could be related to peripheral and central sensitization processes observed in chronic pain patients [Cagnie et al., 2014, Kuner, 2010, Gold and Gebhart, 2010].

Once again we saw that predictive coding could be used as an encompassing theoretical framework to understand the mechanisms at stake during maladaptive pain perception, and further investigation in this sense is needed. Meanwhile experimental testing of Bayesian views on “normal” pain perception already lead to interesting results that we will now describe.

2.3.2 Effects of expectations and uncertainty

Is pain perception (and its neural correlates) modulated by expectations in an experimental context ?

The modulation of pain by positive or negative expectations in clinical settings is a well know phenomenon which can cause hypoalgesia (placebo) or hyperalgesia (nocebo) [Tracey, 2010]. Yet, such expectations are most often about the applied treatment, have an extended temporal span and can imply numerous cognitive, social and motivational factors which are difficult to disentangle. To gain more understandings on the cognitive and neuronal mechanisms supporting the effects of expectations on pain perception, researchers used experimental paradigms designed to set cue-based stimulus expectancies. Ploghaus et al. (2003) offered a first review of this field of research which suggests that even short-term expectations varying on a trial-to-trial basis, can have a strong influence on participants behavioral ratings of pain, but also on brain activation [Ploghaus et al., 2003]. They identified the rostral anterior cingulate cortex (rACC) and the insula, core brain regions of the hypothesized second-order “attentional-perceptive matrix” of pain ([Garcia-Larrea and Peyron, 2013], see section 1.2.4), as well as the posterior cerebellum, to be activated by *certain* expectations about impending painful stimuli no matter if a lower or higher pain was expected. A more recent review from Atlas and Wager (2012) confirmed these results but also emphasized the importance of the thalamus in setting these expectancies, as well as more frontal regions associated with emotional appraisal such as the dlPFC and orbitofrontal cortex (OFC) which would mediate the anticipatory response to cues [Atlas and Wager, 2012].

The conclusion from these reviews, as well as more recent results on cue-based expectancies [Hoskin et al., 2019, Lobanov et al., 2014, Peerdeman et al., 2021, Zaman et al., 2022], fit well with the predictive coding view of pain, both at behavioral and neuronal levels. Indeed prior expectations of the painful or innocuous nature of upcoming stimuli would set strong predictions for respectively high or low pain, which will be integrated to the actual nociceptive input in a Bayesian way, driving pain

perception towards expectations if they are precise enough (as in the case of a “strong” placebo evoked earlier for example) and updating expectation in case of a high discrepancy between predictions and inputs (strong prediction errors). Retrospective interpretations of previous results would point to rACC and anterior insula as central hubs of the predictive processing of pain in the cortical hierarchy. Geuter et al. 2017 directly tested this hypothesis by submitting participants to warm or hot thermal stimuli following cues of varying validity, leading to a combination of stimulus type (“warm” or “hot”), expectation (“more likely warm”, “more likely hot” or “both as likely”) and therefore prediction errors [Geuter et al., 2017]. They found that both skin conductance and pupil size were better explained by a predictive coding-like model (where physiological response are modeled as the sum of differently weighted predictions and predictions errors) than a stimulus only model and a stimulus plus expectation model. More interestingly, the activation of the anterior insula recorded with fMRI was also best explained by the predictive coding model (as well as the right amygdala) whereas the activity in other regions of the pain matrix (whose the ACC and posterior part of the insula) was best explained by a stimulus only model [Geuter et al., 2017]. The involvement of the posterior insula only in stimulus encoding was expected as it has been associated to the first-order nociceptive matrix of pain [Garcia-Larrea and Peyron, 2013], yet surprisingly the ACC did not reflect expectations contrary to previous results [Atlas and Wager, 2012, Ploghaus et al., 2003]. This discrepancy could have arisen again from a broad definition of the ACC as Broadman areas 24 and 32 which does not distinguish the rACC from its historically more posterior or caudal part now called mid-cingulate cortex (MCC) and which is known to be more involved in primary pain processing [Garcia-Larrea and Peyron, 2013]. Fazeli and Buchel (2018) replicated and further expanded their previous results while controlling for the aversiveness of their stimuli which could have driven expectations response in the anterior insula [Fazeli and Büchel, 2018]. They showed again an involvement of the anterior insula in coding stimulus intensity and expectation which could not be explained by aversiveness and interestingly they further characterized the ventral part of anterior insula as specifically coding prediction errors.

We described how the behavioral and neural correlates associated with *certain* pain expectations seem to agree with a predictive processing view of pain. However in this framework a consequent importance is also afforded to the precision of expectations. The effect of *uncertain* expectations on pain perception was also studied scientifically and we will now review some results on the topic.

The uncertain effect of uncertainty

As hinted to in section 2.3.1, Büchel et al. (2014) suggested that according to a Bayesian view of pain, the precision with which a placebo intervention is given, and therefore the trust patients put on it, may play a role in its efficacy (Figure 2.9). They notably base their hypothesis on a series of studies from

Colloca and colleagues (2006, 2010) on placebo conditioning. In a nutshell, the authors found that placebo-induced analgesia was stronger and lasted longer (up to 7 days) when the effectiveness of the “fake” treatment was demonstrated empirically to the participants, instead of giving them only verbal suggestions [Colloca and Benedetti, 2006]. A second study showed that participants who received four conditioning trials showed again a stronger and longer-lasting (within a session) placebo-induced analgesia (but also nocebo-induced hyperalgesia) effects than participants who only received one [Colloca et al., 2010]. Colloca et al. concluded that the placebo and nocebo effects rely on the learning of expectation-stimulus contingencies that are reliable enough. Similarly in a predictive processing framework, to form strong and long-lasting expectations about pain, these expectations need to be precise.

Büchel’s team empirically tested their hypothesis in a recent study by artificially conditioning participants to experience lower heat pain during a transcutaneous electrical nerve stimulation (TENS) condition [Grahl et al., 2018]. Unbeknownst to the participants, there was actually no TENS application but stimulus intensity was reduced during “Tens on” trials (Plac) of the conditioning phase to either a constant 30% of pain tolerance for the high precision treatment group (HPT) or to a varying intensity around 30% for the low precision treatment group (LPT, see Figure 2.10A). This manipulation allowed them to create a high precision placebo (HPT group) and a low precision placebo (LPT group) who indeed showed more variable pain ratings during the conditioning (Figure 2.10B). The subsequent test phase presented participants with the same medium level stimulation in both “TENS on” (Plac) and “TENS off” (Ctrl) trials, yet participants rated the pain as less painful during the “TENS on”, showing a clear placebo effects (Figure 2.10B). Interestingly this effect was stronger in the participants who received a high precision placebo, validating Büchel et al. (2014) hypothesis.

The placebo conditioning results we just reviewed globally show that the learning and subsequent long-lasting effect of expectations depend on their associated precision or inverse uncertainty. However these experiments did not test the direct effect of the uncertainty of cue-based trial-to-trial uncertainty. Relying on preliminary results of these kind, Ploghaus et al. (2003) argued that painful stimulations that are expected, but uncertain about their exact intensity, would cause hyperalgesia through anxiety related mechanisms mediated by the ventro-medial prefrontal cortex and mid-cingulate cortex [Ploghaus et al., 2003]. They hypothesized that knowing with certainty the painful nature of impending stimuli would activate the fear response system, leading to a decrease in pain sensitivity [Rhudy and Meagher, 2000], whereas uncertainty would activate anticipatory and anxiety mechanisms leading to an increase in pain sensitivity. A subsequent empirical work controlling for the mean value of cue-based expectations while changing their variance (inverse precision) showed indeed higher pain ratings under high uncertainty conditions irrespective of the direction of the induced expectations [Yoshida

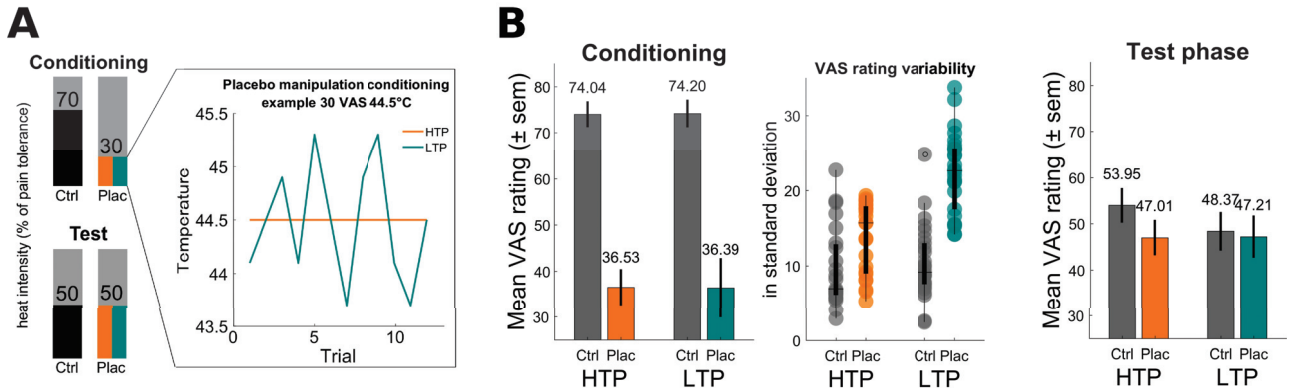


Figure 2.10: Methods and behavioral results from placebo conditioning with variable precision in [Grahl et al., 2018] (A) All participants received a placebo conditioning where stimulus intensity was reduced to 30% of pain tolerance during “Plac” trials and was set to 70% during control (Ctrl) trials. For participants in the high precision treatment (HTP) group this simulated analgesia was constant, whereas it presented some variation and therefore was less precise for the low treatment precision (LTP) group. (B) Behavioral ratings of pain intensity during the conditioning phase that placebo simulation was effective in both groups, yet LTP participants had higher variation of pain ratings in the simulated placebo trials, as expected. During the test phase, both groups showed a placebo effect (significant main effect of condition with Ctrl > Plac), yet when controlling for difference in response to the Ctrl condition, the placebo was stronger for the HTP group (significant interaction). Adapted from [Grahl et al., 2018].

et al., 2013]. However as emphasized in a recent review by Zaman et al. (2021), within the Bayesian view of pain the effect of uncertainty (understood as the amount of variation of trial-to-trial cue-based expectations) should be to decrease the overall influence of short-term expectations on the pain inferential process. This is most likely not an all-or-none phenomenon and uncertain expectations may have varying effects depending on experimental and individual factors but it has no reason to lead forcefully to hyperalgesia [Zaman et al., 2021]. Trying to replicate the findings from Yoshida et al. (2013) on a larger sample of subjects (40 vs 13), they instead found that in case of high uncertainty of prior information pain ratings were biased toward the average expectation [Zaman et al., 2017], as suggested by other results [Brown et al., 2008, Lorenz et al., 2005] and in agreement with a Bayesian understanding of pain. Yet, as they emphasized, if short-term, cue-based, uncertainty may not have an effect in experimental settings, it is most likely an aggravating factor in long-term chronic pain conditions for instance [Zaman et al., 2021]. Actually meditation-induced psychological relief observed in chronic pain patients following an MBSR program [Kabat-Zinn, 1982, Kabat-Zinn et al., 1985] could be related, if not to a decrease, at least acceptance and equanimity towards the life-threatening uncertainty of their pathology, a prospect we will develop later.

We saw in this section how scientific results on pain expectations and uncertainty lend some support for the predictive processing view of pain. Another way to test predictions from this newly emerging framework of pain perception is to use computational models implementing the hypothesized Bayesian-like processes. Computational modeling allows to predict at once both expectations

and their uncertainty while taking into account trial and subject-wise variations, offering a more integrative approach than classical analyses of trial and subject averaged ratings. We implemented one such approach in our study of pain perception and its potential regulation by intensive meditation practice (chapter 5). Importantly, we got inspired from recent modeling works from different research teams which overall provided considerable evidence for the superiority of predictive processing inspired models to explain pain perception in humans in terms of behavior [Anchisi and Zanon, 2015, Grahl et al., 2018, Hoskin et al., 2019, Nickel et al., 2022], physiological responses [Geuter et al., 2017, Fazeli and Büchel, 2018, Nickel et al., 2022], fMRI brain activations [Geuter et al., 2017, Fazeli and Büchel, 2018, Grahl et al., 2018] and EEG oscillations [Strube et al., 2021, Nickel et al., 2022] as well as in freely-moving animals (rats) with invasive electrophysiological recordings [Song et al., 2019, Song et al., 2021].

Chapter 3

A predictive processing view of contemplative experience

In this last introductory chapter we will try to link what was discovered in the last 50 years of contemplative research — but also what has been already theorized by contemplative traditions for centuries — to the recent explanations of the body-brain system as a predictive engine. In this effort we will aim at integrating several preliminary theoretical works on this topic into a broader (but most likely provisional) account of how meditative practices may alter predictive mechanisms. To do so, we will develop three main working hypotheses on which we based our subsequent empirical work. Finally, we will present the exact study design and experimental paradigms that we implemented during this PhD, notably to investigate the acute effect of meditation during a 10-day residential retreat.

As a side note, following the terminology used in most scientific works in the field and to gain clarity, we will interchangeably talk about predictive processing or Bayesian inference to refer to both predictive coding (which was historically restricted to perception) and active inference (at first, limited to motor behavior). Yet, it should be kept in mind that recent developments of active inference are construed as the most general formulation of the Free Energy Principle applied to the brain and, as such, have been proposed to explain not only action and perception but also interoception, mental actions and extended cognition. To bridge this most recent developments of the theory with previous formulations of mostly perceptual processes, such as the auditory deviance detection or pain perception, in terms of “predictive coding” we chose to stick most of the time to the more popular term of “predictive processing”, reserving the usage of “active inference” to the necessary cases (e.g., to account for mental phenomena). We hope that the very recent release of the “Active Inference” volume by Parr, Pezzulo and Friston (2022) will foster a broader use of the term and harmonize the field [Parr et al., 2022].

3.1 Explaining the meditative brain (and body) through perceptual, interoceptive and active (Bayesian) inference

3.1.1 Preliminary proposals

A first proposal to link contemplative practice and predictive coding is to be found within a theoretical article from Khachouf, Poletti and Pagnoni (2013) whose main focus was to revive the neuropsychological enterprise initiated by Varela [Varela et al., 1992, Varela, 1996] with Kantian ideas on transcendentalism [Khachouf et al., 2013]. Briefly stated, they proposed that concurrent free-energy (or prediction errors) minimization under low sensory stimulation and heightened attention during meditative practice could allow the refinement of internal models of both the world and the self, progressively revealing to the practitioner its own cognitive structure.

Similar ideas were developed in parallel by Farb et al. (2015) who focused more specifically on interoception and its modulation by contemplative practices [Farb et al., 2015]. Indeed the perception of internally generated sensory signals (heart-rate, breathing, body temperature and any other homeostatically-regulated physiological variables) could also be explained in terms of predictive processing, with a central role of the insular cortex in the integration of interoceptive signals ([Seth et al., 2012, Seth, 2013, Seth and Friston, 2016, Allen et al., 2019, Barrett and Simmons, 2015]; see also Figure 3.1 for a simplified mechanistic account). Farb et al. (2015) claimed that meditation would partly exert its transformative effects on maladaptive habits and behaviors through enhanced sensitivity and non-reactivity to bodily interoceptive signals but also increased regulation of automatic emotional reactions [Farb et al., 2015]. Such modulation would be mediated in a first place by attentional precision weighting (i.e., amplification) of ascending interoceptive sensations proper to the body-oriented (e.g., breathing) meditative focus, whose sensory acuity have been found to be enhanced as a result of meditation [Daubenmier et al., 2013a, Fox et al., 2012, Kerr et al., 2008, Mirams et al., 2013]; with the caveat that this mechanism may not apply to interoceptive signals in general — for instance, experimental assessments of increased heartbeat perception, which is not a common meditation object, have given mixed results [Bornemann and Singer, 2017, Khalsa et al., 2008, Melloni et al., 2013]. Farb et al. (2015) also hypothesized that meditative training would induce an overall shift of cognition from active Bayesian inference, where error minimization is performed by automatically acting upon the world, to “passive” perceptual inference where prior knowledge is revised and down-weighted in the face of current evidence [Farb et al., 2015]. This process of passive observation and regulation of one’s own cognitive mechanisms was linked to the emotional and cognitive attitude of “decentering” that could

help foster non-reactivity toward adverse experiences faced during the course of illness or emotionally difficult events.

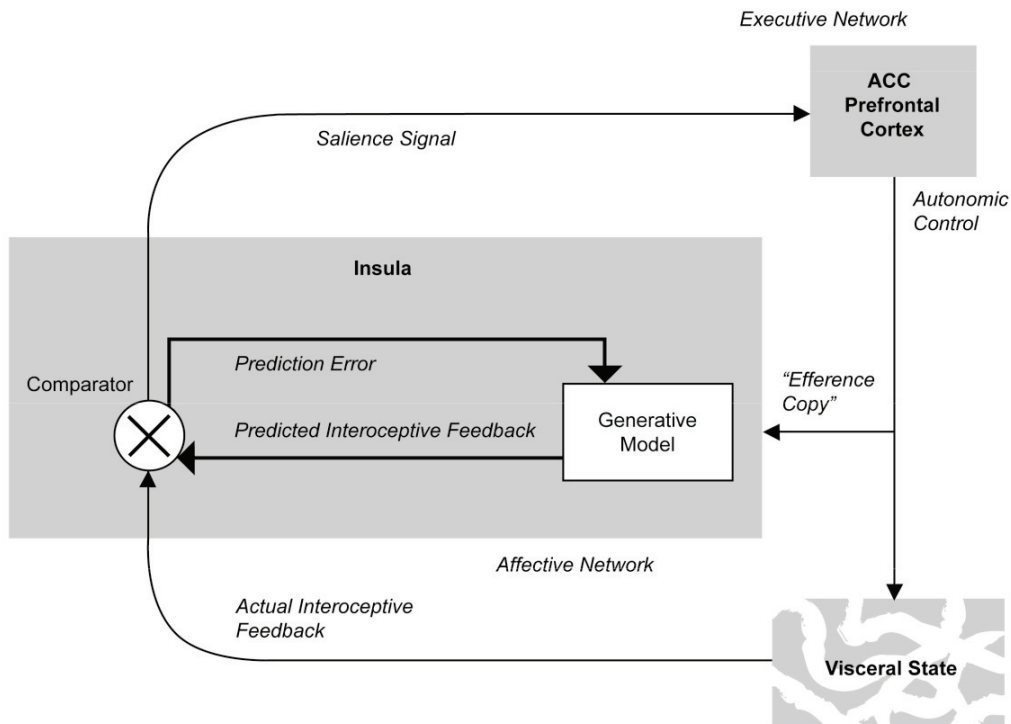


Figure 3.1: Simplified mechanistic representation of “interoceptive” inference. Similarly to classical perceptual inference on exteroceptive inputs, interoceptive information from visceral states of the body would be compared to interoceptive predictions in the insular cortex. Resulting prediction errors would serve to update a generative model of our homeostatic states, but could also take part in the salience signaling function of the insula. Detection of a surprising physiological sensation, e.g., higher-than-expected body temperature, would thus lead to recruitment of the executive network and other relays (not represented in the figure) to regulate through appropriate autonomic/hormonal/behavioral action the homeostatic variable back to a viable range (for vital signals, most likely defined by genetically-encoded, non-flexible priors). *Reproduced from [Jamieson, 2016].*

The perspective of interoceptive predictive processing was also adopted by Jamieson (2016) in the context of hypnotic states and meditative practices, with an emphasis on the central role of the insula, whose activity has been shown to be modulated by both hypnosis and meditation [Jamieson, 2016]. Finally, Pagnoni and Guareschi (2017) approached the question through the original format of a dialogue between a neuroscientist and a Zen Soto abbot [Pagnoni and Guareschi, 2017]. Exploring topics as diverse as mind-body interaction, the importance of attention and posture in meditation, time perception and dreaming, they discussed how parallels could be drawn between the Free Energy Principle explanation of cognition, perception and action, and doctrinal and —perhaps most importantly— empirical knowledge coming from the Buddhist Zen Soto tradition.

If these preliminary discussions were crucial in developing the bases for a predictive coding understanding of processes at stake during meditation, they nevertheless lacked an articulated organization and general viewpoint and did not cast precise hypotheses and predictions on how meditative practices

should influence predictive coding processes. This complex but potentially groundbreaking venture was only initiated recently and theoretical articles are still sparse: a seminal work focused on the open presence practice of *shikantaza* in Zen Buddhism ([Pagnoni, 2019], further developed in [Pagnoni and Guareschi, 2021]), while a more general account including focused attention meditation was developed at the same time [Lutz et al., 2019]; and finally extended to encompass focused attention, open-monitoring and non-dual practices common to most contemplative traditions [Laukkonen and Slagter, 2021], as well as modern mindfulness practice under the MBCT therapeutical intervention [Manjaly and Iglesias, 2020]. Relying on these independent proposals we will now try to delineate an integrated, but most likely provisional, view of meditative experience and effects within the scope of predictive processing.

3.1.2 An integrative hypothetical account of the predictive brain under meditation

[The free energy principle] ties together in a coherent theoretical scaffold the core meditative notions of attention (top–down deployment of precision weighting), the conditioning power of habitual self-related patterns of thought and behavior (priors), and the embodied nature of cognition and emotion (interoceptive inference)

Lutz, Mattout and Pagnoni in *The epistemic and pragmatic value of non-action: a predictive coding perspective on meditation* (2019)

Calming the monkey mind through attentional modulation and sustained posture: a precision-weighting story ?

The term “monkey mind” is often used in Buddhist teachings to refer to the restless and incessant jumping from thoughts to thoughts that naturally arises when we do nothing special, or distract us when we are engaged in a task [Rinpoche and Swanson, 2010a, Rinpoche and Swanson, 2010b]. This cognitive process, now also widely studied in cognitive science as mind-wandering or “day-dreaming”, has been investigated with brain imaging techniques and associated with the activity of the default mode network (DMN, [Christoff et al., 2009]). It is important to note that mind-wandering has most likely a significant adaptive value in supporting future planning, creativity and problem solving, yet both contemplative traditions and empirical research agree that it is also often associated with confusion, rumination, worry and overall negative mood [Killingsworth and Gilbert, 2010, Ruby et al., 2013, Smallwood and Schooler, 2015].

The first and foremost method employed by most meditative practices to “loosen the grip” of this monkey mind and attain calm and clarity is attentional training, which we described at length in the first chapter (see section 1.2.3). Indeed, meditative practices have been shown to behaviorally increase the recognition of mind-wandering episodes through enhanced meta-awareness, often leading to less overall distraction and negative thoughts in daily-life [Kok and Singer, 2017, Jazaieri et al., 2016] or while performing a cognitive task [Zanesco et al., 2016, Mrazek et al., 2013]. In neuroimaging studies, meditation has been in fact associated with reduced activity of the DMN both at rest and during tasks [Brewer et al., 2011, Garrison et al., 2015, Pagnoni et al., 2008], and with a recruitment of the salience and executive networks, a finding interpreted to signal the role of these circuits in disengaging from mind-wandering [Hasenkamp et al., 2012].

A number of theoretical works have proposed various mechanistic explanations for attention-related

effects of meditation (see, e.g., [Lutz et al., 2008b, Malinowski, 2013, Tang et al., 2015, Vago and Zeidan, 2016], non-exhaustive list). We will focus here on recent developments of a simple yet far-reaching predictive processing explanation: meditative attention would differentially modulate the precision-weighting of high-order cognitive contents and low-level sensory information.

This is most prominently exemplified in focused attention (FA) practices where meditators focus on a single target, often breathing or bodily sensations — but it can also be on visual or auditory stimulus, or even thought and emotions — while monitoring potential distractors and disengaging from them whenever distracted. It most often involves to hold a steady sitting posture but lying down or walking FA meditations also exist. In terms of predictive processing, this focus should increase the precision afforded to attended sensory streams and their associated prediction errors, for example interoceptive inputs in the case of breathing, proprioceptive ones for postural sensations and exteroceptive for visual images and sounds. Most predictive processing accounts of meditation [Laukkonen and Slagter, 2021, Lutz et al., 2019, Pagnoni, 2019, Manjaly and Iglesias, 2020] agree that such precision up-weighting of low-level sensory content should mechanically down-weight the relative precision of abstract cognitive content associated to mind-wandering (e.g. thoughts, memories, future planning, etc), as the latter rely on “deep temporal models” of the world, established in higher-order areas (most likely prefrontal) over longer timescales [Laukkonen and Slagter, 2021] ¹.

Predictive processing can also account for the constant cycle between focus and mind-wandering observed in FA practices. The usually calm and undisturbed environment prescribed for meditation leads to a situation of relatively low surprise. After some time without much informative prediction errors, the estimated “newsworthiness” of sensory streams declines and their precision is naturally down-weighted. This increases the probability that the automatically-arising, habit-driven functioning of the brain based on higher-order autobiographical priors (mind-wandering or deep temporal models), becomes preponderant again. Crucially, though, when one is engaged in meditation such shift is likely to be more promptly noticed, as meditation instructions install highly-precise predictions regarding proprioceptive (body stillness) and interoceptive (regular breathing) signals. After few seconds of mind-wandering, loss of attention to the posture may lead, e.g., to the head and shoulders naturally stooping forward, the body leaning sideways, the gaze losing focus or the breath rhythm changing ². These bodily changes would trigger interoceptive and proprioceptive prediction errors, due to

¹Abstract cognition has been linked in the predictive processing framework to a refined version of active inference involving “mental actions” and entailing hierarchical generative models whose inference concerns extended periods of time and are thus called temporally “deep” or “thick” models [Friston, 2018b, Limanowski and Friston, 2018]. This more elaborated kind of inference is necessary for selecting the best sequence of future actions (policy) to minimize *expected* surprise, based on beliefs about desired states of the organism (or sensations), its environment and the consequences of its own actions. To do so accurately an organism needs to entertain a certain counterfactual richness of future states it may fall in, a process therefore described as “counterfactual inference” [Friston, 2018b, Corcoran et al., 2020].

²Slow and deep breathing is encouraged in some traditions and has been hypothesized to occur as a side effect of meditation, even when not explicitly instructed [Gerritsen and Band, 2018]. Slower respiratory rate was

the high precision concerning sensory signals and posture set by the meditative policy, that would accumulate and eventually cause the involuntary attentional drift to be disclosed to consciousness (most likely through the insula and salience network). Executive brain networks would then initiate the disengagement from distracting mental content, thereby stopping abstract deep temporal processing, and shift back attention to the meditative target.

The above perspective underwrites the importance of an active posture in meditative practices. Pagnoni and Guareschi (2017, 2019, 2021) emphasized this point in the case of Zen *shikantaza* practice, where the posture is integral to the meditative process, functioning almost as a “biofeedback device” with numerous channels (upright posture, half-closed eyes and steady gaze, specific hands position [Pagnoni and Guareschi, 2017, Pagnoni, 2019, Pagnoni and Guareschi, 2021]). A great care in setting and monitoring bodily posture would prescribe a very high precision for interoceptive and proprioceptive inputs and lead to easier detection and disengagement of ongoing mind-wandering during meditation, a notion which was further developed and construed as a specific and disciplined form of “mental action” in Lutz et al. 2019. The importance of interoception — and of its main cortical actor, the insula — in mediating the “positive” effects of contemplative practices on bodily and mental health were reviewed at length in [Farb et al., 2015], and the specific role of respiration detailed in [Gerritsen and Band, 2018].

The example of *shikantaza* meditation is of particular relevance as it largely belongs to the family of open-monitoring (OM) practices³, and we can thus ask ourselves whether the precision-weighting process described above for FA meditation also fits OM practices, where attention is monitored but not directed to a single sensory object.

Laukkonen and Slagter (2021) hypothesized that under a distributed attentional stance towards anything that may arise in consciousness, “any content of experience is assigned equal precision, and consequently low precision in relative terms” [Laukkonen and Slagter, 2021]. Thus, the consequences of OM and FA practices on mind-wandering — i.e., an overall reduction of this process — would be quite similar, as (spontaneous) mental content that is not afforded a relatively greater precision compared to other content could not be sustained for very long. However, while one could predict shorter and shorter mind-wandering episodes as a practitioner gain experience in the practice, the goal is not to stop the natural flow of thoughts. In fact, a quality of non-judgment and non-reactivity (equanimity) towards the content of experience is often instructed in OM meditations, which prominently applies to any past- or future-oriented thoughts. A non-judgmental attitude could be construed in

indeed observed empirically in both novices (with also increased respiration volume when attending specifically to their breath [Farb et al., 2013]) and experts meditation practitioners at rest [Wielgosz et al., 2016], compared to control participants. Experts were additionally shown to have somewhat higher interoceptive accuracy of their breathing compared to controls [Daubenmier et al., 2013b].

³But it could also be argued that *shikantaza* is closer to an advanced form of OM referred to as open presence, a practice sometimes placed in the non-dual family of contemplative practices [Dunne, 2011b].

predictive processing as a flattening of prior beliefs. No hypothesis, cause or explanation would be greatly favored over another at higher cognitive levels. This would lead higher-order (temporally-deep) predictions based on autobiographical ingrained priors and habits to have low precision and so impede their triggering of self-reinforcing confirmative actions (habits are formed and maintained via their repetition), ultimately causing them to degenerate (in the sense of having a shorter lifetime). According to Laukkonen and Slagter (2021) this non-judgmental aspect of meditative experience should reach its climax in highly advanced “non-dual” practices supposed to halt even the distinction between self and objects, reaching a state of awareness “that is hence devoid of concepts (self, objects), intentionality or the experience of time and space, i.e., a state in which even the most basic constructs of cognition allegedly no longer persist” [Laukkonen and Slagter, 2021]. The realization of a non-dual state would theoretically lessen the selection of any deep temporal processing or in the authors’ words, it would “prune the counterfactual tree”, hypothetically interrupting any cognitive or sensory prior formation for the duration of the practice (see Figure 3.2).

Pagnoni (2021) and Lutz (2019) further argued that the regulated physical and mental poise instantiated during most meditative practices would suspend the habitual (and unaware) confirmatory nature of our actions *vis-à-vis* our engrained priors, thus reducing the natural tendency to judge and act compulsively [Pagnoni and Guareschi, 2021, Lutz et al., 2019]. In other words, meditative practice would involve nudging Bayesian inference in the direction of “perception” (i.e. “changing the model to better match the world”) rather than “action” (“changing the world to better match the model”), as Farb et al. (2015) also emphasized, from which its epistemic value “of gaining an intimate knowledge about the constructive and dynamical nature of mental function (e.g. recognizing the impermanence of a thought)” [Lutz et al., 2019] that we are going to detail now.

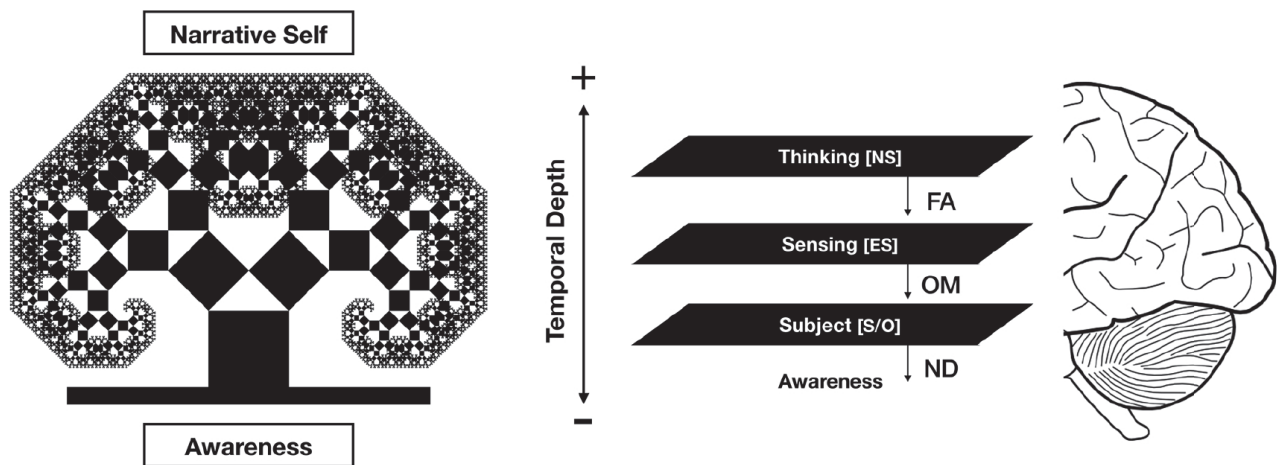


Figure 3.2: Schematic representation of the reduction in deep temporal processing supposed to happen during meditative practices by Laukkonen and Slagter (2021) The left panel represents the increase in temporal depth of processing, when ascending in the brain hierarchy, by the complexity of a “counterfactual tree”. Different levels of self-experience are supposed to arise at multiple levels of the hierarchy. At the top we find a narrative self [NS] resulting from deep temporal models (over long timescales) on autobiographical priors, which are supposed to predict future states and guide behavior to minimize *expected* surprise. At an intermediary level is found an experiencing self [ES] based on less temporally thick models embodying the causal structure of the world and minimizing actual surprise based on current inputs. At the lowest level a primitive form of self-experience arising from the subject/object distinction would still be present. The right panel represents the effect of meditative practices in reducing the temporal depth or thickness within this scheme, or as metaphorically coined by the authors how they “prune the counterfactual tree”. Focused Attention (FA) meditation would disrupt the narrative self (cause of mind-wandering processes) by passing from a thinking to sensing mode of experience. Open monitoring (OM) practices are hypothesized to further break down temporal depth to a state where any experience is appraised equally but the subject/object duality still remains. Finally non-dual (ND) practices would remove even this duality to reveal the latent background of all experience, awareness itself. A more precise description of such “pure awareness” states are provided in [Laukkonen and Slagter, 2021], and notably the paradoxical assertion of being able to report on these states despite the absence of “subject” to experience it is treated. The authors also acknowledged that the three layered cognitive hierarchy depicted here is a simplification to map the three meditative practices developed above, and that the real hierarchy is certainly composed of many more layers. *Reproduced from [Laukkonen and Slagter, 2021].*

From mental transparency to opacity: meditation as a tool to reveal the cognitive structure of the mind

The hypothesis that meditative practice, through the modulation of predictive processing, may foster insights into the deeply constructed nature of cognition was present in the earliest work on the topic [Khachouf et al., 2013]. Farb et al. (2015) then developed how enhanced meta-awareness of interoceptive processes may lead to a clearer view of the complex relationships between bodily reactions, emotions and thoughts, informing the practitioner on both its mental and emotional cognition [Farb et al., 2015]. A crucial link between meditation epistemic goal and predictive processing account of mental activity was introduced by Lutz, Mattout and Pagnoni [Lutz et al., 2019], using the philosophical notion of phenomenal transparency [Metzinger, 2003], which we will briefly illustrate in the

following.

When we perceive, think or feel we are usually aware of their *content*, but not of the underlying cognitive *processes* allowing these experiences, which are therefore “transparent” to us. In other words, we have only access to the final product of our cognition (like when seeing a bird passing by our window) but not of the medium by which cognition is possible (the transparent glass allowing us to see outside). Conversely, a mental process is said to be “opaque” when we are clearly aware of the underlying cognitive and contextual factors leading to a given experience, for instance the logical reasoning needed to obtain a mathematical result. Within the predictive processing framework, conscious access to a given mental process or object through “introspective attention” (“opacifying” it) has been linked to the increased precision afforded to its corresponding representation⁴ which crucially can be mediated by top-down deploying precision-weighting at a specific level of the hierarchy of the generative model, a “mental action” associated with attentional control [Limanowski and Friston, 2018].

Interestingly, phenomenal transparency seems to be our default cognitive mode [Lutz et al., 2019, Pagnoni, 2019], and this cannot be changed by a simple conceptual understanding of this notion. Lutz (2019) and Pagnoni (2019) suggested that meditative practice may enable a progressive opacification of habitually transparent cognitive processes such as perception, emotions, up to the unfolding of thoughts themselves. Carefully “looking inward” at the perceptual and mental phenomena accompanying an emotional reaction (e.g. being angry) for instance, would progressively disclose its building blocks and dynamics (e.g., the tension in the body, the sensation of a looming threat, etc.), thus revealing its constructed and provisional nature. This specific meta-cognitive quality developed through meditation training has been variously described as “phenomenological reduction” [Varela, 1996, Husserl, 1931], “decentering”, “defusion” or “dereification” in the literature [Farb et al., 2015, Lutz et al., 2015] and generally entails the ability to experience thoughts, emotions and percepts as transient mental events, rather than taking them for granted as accurate depictions of reality or fixed ontological entities, and therefore realize their dependent and impermanent nature.

However, as Limanowski and Friston (2018) observed, a disruption of the natural phenomenal transparency at higher levels of cognition, up to our minimal sense of self-hood, may also cause experiences of derealization and depersonalization, which appear in psychopathological conditions such as schizophrenia [Friston et al., 2014], under psychoactive substances [Millière et al., 2018], but also during meditation [Lindahl et al., 2017]. If the involuntary disruption of the sense of self has been characterized as an unpleasant “side effect” of meditation practice [Lindahl et al., 2017, Schlosser et al., 2019]

⁴Representation is here not used in the classic cognitivist sense of representations *about* objects which would only exist “out there”, and would need a *central controller* to manipulate them —a view now fairly criticized, see [Varela et al., 1992] for a detailed critic and the initial *enactivist* proposal. Differently in predictive processing / active inference we can talk of patterns of brain activity behaving “as if” they were representing external objects (after all, representation is in the eye of the beholder) and there is no need for a central controller as the system is a self-organizing one.

⁵, such experience is also intentionally brought forth in the so-called non-dual contemplative practices described above [Laukkonen and Slagter, 2021, Dunne, 2011b]. More conventional meditative practices such as FA would already change the transparency or “realness” of higher-order, temporally-thick cognitive priors through the deployment of greater precision to the lower layers of the generative model closer to the sensory surfaces [Laukkonen and Slagter, 2021]. OM practices are hypothesized to increase dereification even more through the non-judgmental attitude they promote, and in doing so should “reveal that the embodied and narrative self are just *processes* rather than concrete entities” ([Laukkonen and Slagter, 2021], e.g. [Teasdale and Chaskalson (Kulananda), 2011]) .

Manjaly and Iglesias (2020) provided another interesting mechanistic explanation for how mindfulness meditation may foster dereification through changes in predictive processing, focusing specifically on the “being mode” and decentering qualities promoted within the MBCT (Mindfulness-Based Cognitive Therapy) program [Manjaly and Iglesias, 2020]. In their view, the heightened attention, and therefore precision, afforded to varying incoming sensory signals during meditation would induce a constant update of beliefs (or priors) about states of the environment (world and body), as the prediction errors associated to these signals are expected to carry meaningful information. According to the self-efficacy hypothesis of bodily states control [Stephan et al., 2016], the brain would entertain a specific model of internal physiological states, constantly monitoring vital signals of the body (allostasis). In addition a secondary level in this system would monitor (make predictions on) the “controllability” of these body states, a process that could be construed as meta-allostasis. Higher-order metacognitive beliefs about physiological control would (in part) create our sense of agency: “I (the system) am in control of my bodily states”. Meditation, by triggering a constant updates of beliefs about bodily states through the heightened attention to interoceptive signals, would underscore the continuous variability of such signals, which in turn would alter allostasis predictability [Manjaly and Iglesias, 2020]. As a consequence, the metacognitive belief of being in control would be weakened, loosening our felt sense of agency (see Figure 3.3). Said otherwise and perhaps more intuitively, paying attention to the body informs the practitioner of the constant change of states of the world (of which its own body states), what is often called in Buddhist teachings the “law of impermanence”, and reveals the illusory nature of believing that one is totally in control, even of its own body.⁶

⁵While experientially clarifying the impermanent and constructed nature of the self is deemed a necessary step to alleviate one’s own suffering in Buddhist philosophy [Varela et al., 1992], this does not guarantee future “positive” consequences in the meditator, which may speak to the important role of cultural/religious framing of these practices and of a close teacher (see [Lindahl et al., 2019, Schlosser et al., 2019] for insightful discussions on this complex topic.)

⁶One could wonder if such disruption of the feeling of being in control may not be aversive and trigger for example a panic attack during meditation, a problem not tackled by [Manjaly and Iglesias, 2020] and which could be further enhanced by pre-existing individual conditions. Adverse events of this kind are not impossible in meditation, yet quite rare and known to meditation teachers who are usually trained to prevent and manage them (see [Lindahl et al., 2017, Lindahl et al., 2019] and section 1.2.5 where we covered meditation-related adverse events in details). More specifically, as envisaged in the self-efficacy hypothesis of allostatic control: “While perceived low self-efficacy likely represents an inevitable consequence of persistent dyshomeostasis, various

A critical commentary of Manjaly and Iglesias (2020) proposal was made by Verdonk and Trousselard (2021). Briefly, they suggest that the increase precision in likelihood (sensory inputs) resulting from mindfulness practice (on which they agree), is not compatible with Manjaly and Iglesias’s secondary proposal of reduced prior precision [Verdonk and Trousselard, 2021]. Increased likelihood precision at time t would indeed lead to increased precision of the posterior belief about causes of a given sensory signal, a posterior which becomes the new prior at time $t+1$ according to Bayes’ rule. Instead they propose that mindfulness would enhance the context-updating of prior beliefs in response to changes in the environment so they are optimally adjusted to the present moment. Actually both proposal may be reconcilable in the sense that to allow for an optimal updating of beliefs, priors need to be flexible enough (not afforded overwhelming precision compared to sensory likelihood), what meditation would permit.

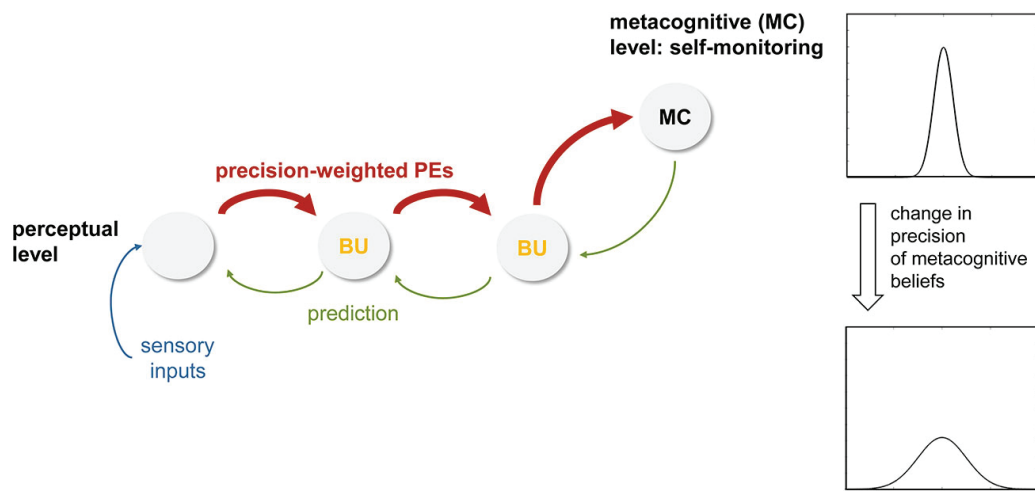


Figure 3.3: Hypothesized mechanism for meditation induced decentering or dereification proposed in [Manjaly and Iglesias, 2020]. The high-precision afforded to sensory prediction errors (PE) would trigger a constant belief updating (BU) at perceptual levels modeling the inferred sensory states, crucially including bodily ones. As a consequence, higher-order metacognitive beliefs about body states control would see their relative precision reduced, leading to a loosening of the sense of control and agency, a first step for seeing thoughts, emotions and percepts as transient and dependent mental processes (dereification). *Reproduced from [Manjaly and Iglesias, 2020].*

protective factors may prevent its spread to other cognitive domains and block the generalization to hopelessness. For example, intellectual abilities or social support may maintain a sense of mastery that shields against an all-encompassing feeling of loss of control.” [Stephan et al., 2016] We can hypothesize that in the case of meditative practice, group support and instructor teachings provide these protective factors, emphasizing the importance of guidance from qualified instructors and the needed carefulness regarding individuals with predispositions to “hopelessness”.

Predictive model optimization, the *praxis* of contemplative stillness ?

Another way meditation could impact predictive processing in the brain is through the optimization of the internal generative model [Pagnoni and Guareschi, 2017, Pagnoni, 2019, Pagnoni and Guareschi, 2021]. This notion is inspired by the predictive processing view of sleep and dreaming proposed by [Hobson and Friston, 2012]. In a nutshell, Hobson and Friston proposed that rapid eye movement (REM) sleep in mammals and birds serves a crucial function: optimizing the complex world model shaped through active and perceptual inference during the waking hours by reducing its complexity [Hobson and Friston, 2012]. This speaks to the parsimony principle known as the Occam's razor: any explanation (or model) should be as simple as possible, while still providing sufficient accuracy. Indeed, an extremely accurate and complex model would both be computationally expensive and not generalize well to novel situations (a situation known as overfitting in statistics). One of the strengths of the active inference approach is that model accuracy *and* complexity are jointly optimized under the free-energy minimization scheme. In this specific case, the reduction of complexity would be carried out during REM sleep by pruning redundant synaptic connections, a proposal largely in line with the synaptic homeostasis hypothesis of sleep [Tononi and Cirelli, 2006, Tononi and Cirelli, 2014]. Hobson and Friston (2012) additionally proposed that dreams would be a manifestation of priors-driven inferential processes where top-down predictions continue to be emitted but sensory prediction errors and motor behavior (except for rapid-eyes movements or during sleep related disorders) are suppressed by a circadian shift from aminergic to cholinergic neuromodulation inhibiting sensory and motor neurons [Hobson and Friston, 2012]. In this modality, the brain would entertain and optimize virtual scenarios of its world based on the recombination of statistical regularities sampled during wake. As summarized by the authors: “sleep provides an opportunity to eliminate the complexity and redundancy accumulated by experience-dependent learning during the day” [Hobson and Friston, 2012].

Pagnoni (2017) first proposed an interesting parallel between REM sleep and meditation by noting that both states imply a reduction in the complexity of the sensorium (darkness for sleep, calm environment for meditation) and an inhibition of overt behavior (automatic in sleep, voluntary in meditation) [Pagnoni and Guareschi, 2017]. Similarly to REM sleep, meditation could also support a form of model optimization, a notion reminiscent of the emphasis on “unlearning” dear to contemplative teachings as exemplified by the famous quote of Shunryu Suzuki in his book *Zen Mind, Beginner's Mind*: “In the beginner's mind there are many possibilities, but in the expert's mind there are few.” [Suzuki, 2020]. More specifically, predictive processing accounts of meditation argue that this optimization of our internal model would target in particular the relationships between mental activity and its bodily correlates [Pagnoni, 2019, Lutz et al., 2019, Pagnoni and Guareschi, 2021]. By sitting still in a non-

judgmental attitude toward anything that may arise, meditation would allow the concrete re-enacting of stimulus-response behaviors occurring in daily life but in a non-reactive, non-confirming way. For instance, a saddening thought or memory comes to mind and is quickly accompanied by a host of bodily reactions: the posture bends forward, a churning sensation appears in the belly, the chest gets tighter, breathing becomes shallower, etc. It is possible that with sufficient meditative training one can observe these natural processes “at a distance”, without trying to stop them, neither judging nor reacting. Such “in-active” stance would steer the inferential process in the direction of pruning maladaptive but long-ingrained reactive priors or “patterns, to chasing after what we like (including ‘enlightenment’) and avoiding what we don’t like” [Pagnoni and Guareschi, 2021]. In this sense contemplative *in-action* can be thought of as a virtuously constrained “wholesome” model optimization toward a long-term reduction of suffering, which may mediate its positive effects on pathological conditions [Wielgosz et al., 2019].

Predictive model optimization during sleep was further extended to “day-dreaming” or mind-wandering by seeing these moments of abstract reflection secluded from the sensory world as “fact-free learning”, potentially leading to insights [Friston et al., 2017, Laukkonen and Slagter, 2021]. From this perspective, mind-wandering episodes naturally and recurrently occurring during meditation can be seen as opportunities, rather than nuisances, “for the generative model that we are to learn about itself” [Pagnoni and Guareschi, 2021]. Indeed, as noted by Pagnoni (2019), during REM sleep our mental content is fully transparent, we do not know that we are dreaming (except during lucid dreaming), and similarly — although to a lesser extent — during day-dreaming we are engrossed in our streams of thoughts, oblivious to the cognitive processes generating them. Meditation, on the other hand, may render mental processes more opaque [Pagnoni, 2019, Lutz et al., 2019], thereby facilitating the *bodily* experience of thoughts and emotions as constructed processes, amenable to regulation and perspective taking. By allowing the observation of the unfolding of mind-wandering, meditation could amplify its insight-generating nature or as cleverly expressed by Pagnoni and Guareschi (2021): “meditative in-action [can] be regarded as a skilled optimization of the ‘default’ auto-epistemic modality that characterizes spontaneous mindful rest” [Pagnoni and Guareschi, 2021]. Once again, the non-reactivity to spontaneous thoughts, instead of reinforcing them as in pathological cases of rumination (e.g., in depression), would reduce their salience. Furthermore, the immobility prescribed by the posture would compel the system (us) to find alternative ways of minimizing the prediction errors associated to unexpected mental states (e.g., anxiety) and their autonomic/bodily correlates (e.g. stress arousal), rather than resorting, as we often do, to performing distracting actions that only temporarily attenuate the consequences of the latent generating issue. During meditation one can clearly see how disturbing thoughts are impermanent and quickly disappear when not actively sustained, while their associated bodily reactions also vanish. Pagnoni and Guareschi (2021) went further by proposing that “meditation

is not a strategy to assuage anxiety by reducing uncertainty, but rather a way of constraining dynamics against habitual reactive patterns so that tolerance to uncertainty can be learned” [Pagnoni and Guareschi, 2021]. While such view may seem at odds with the predictive processing axiom of surprise minimization, the contradiction is only apparent: transient increases in surprise represents opportunities for the internal model to incorporate novel information, which may be necessary to minimize the average surprise in the long term.

In summary, we identified three related ways in which meditative practice may influence predictive processing. A first important step would be the precision up-weighting of low-level, temporally thin, sensory (interoceptive, proprioceptive and exteroceptive) information and the resulting relative precision down-weighting of high-level, temporally thick, abstract mental content, via a disciplined training of attention and postural immobility. The subsequent release of attentional focus in open-monitoring practices would correspond to a more equitable redistribution of precision among the possible contents of experience, a “flattening” of the distribution of prior beliefs over expected body-mind states (e.g., mental habits and their bodily correlates), promoting a non-judgmental attitude. Additionally, active monitoring of proprioceptive inputs and voluntary stillness could induce a shift of the inferential process from the active, confirmatory modality (changing the world to increase the likelihood of our desired states) to the perceptual modality (revising the model to make it more aligned to the current evidence). A parallel outcome, naturally stemming from the previous changes, would be an increased metacognitive access to the processes underlying any mental phenomena, be them percepts, emotions, thoughts, up to the felt sense of an independent and permanent self. In time, such heightened meta-awareness is deemed to foster the “dereification” of mental content, which then comes to be seen as mind-constructed entities, impermanent and dependent on various factors, abating the natural tendency to automatically react to them in maladaptive ways. Finally, contemplative *in-action* and the mindful regulation of day-dreaming episodes naturally appearing during its practice, could lead to the optimization of our embodied generative model, ultimately cultivating further equanimity and insights.

The theoretical account we just developed does not have the ambition to be fully exhaustive and explain every aspect of meditative experience and its effects. Rather, we tried to extract the core components on which most of the previous accounts agree and which could be reliably based on known — and in some cases, experimentally proven — aspects of predictive processing. The main and significant caveat of such proposal is the current lack of experimental evidence to support it. If previous results from the meditation literature were ingeniously reinterpreted with predictive processing lenses in a postdictive manner, experimental works genuinely testing the hypotheses developed above were almost absent at the onset of this PhD project (2018), except from an initial work from our team on a mismatch negativity EEG paradigm [Fucci et al., 2018], and a preliminary study on reward-based

prediction errors [Kirk and Montague, 2015]. Since then we have identified only four studies which tried to test the effect of meditation practice on cerebral and behavioral markers of predictive processing: a follow-up fMRI study on the effect of mindfulness practice on reward prediction errors [Kirk et al., 2019], two studies on pain conditioning [Taylor et al., 2018] and pain expectations [Vencatachellum et al., 2021], and an experimental work on multi-sensory integration [Guthrie et al., 2022]. In addition, a modeling work carried out in collaboration with our team simulated meta-awareness and attentional control processes within an active inference framework [Sandved-Smith et al., 2021]. While these pioneering works are certainly valuable, more experimental research is needed and the present PhD thesis work was aimed precisely at this.

We will now detail simple predictions that can be drawn from the predictive processing account of meditation we just reviewed, along with the experiments and study we designed to test them during this PhD.

3.2 From theoretical to active “scientific” inference: predicting and testing the effects of meditative practice on predictive processing

3.2.1 Heightened sensory precision and decreased prior formation? Insights from the study of auditory deviance detection under meditation

The auditory mismatch negativity (MMN) response, which we described earlier (section 2.2), is hypothesized to be a correlate of prediction errors within the cortical hierarchy and to be enhanced by attention through precision-weighting. According to the predictive processing account presented previously, meditation would be expected to influence the MMN, potentially in different ways depending on the specific meditative practice. FA meditation most likely fosters a global state of high-sensory precision as potential distractors (e.g. out of focus sensations, mental contents) are monitored and disengaged from. This enhanced attentional state, even if not usually directed to the auditory modality (focus on the breath or body sensations), would render unexpected stimuli (deviant tones) more surprising, leading to an increased MMN. A meditation focusing specifically on the auditory modality should further amplify this effect. During OM meditation, attention is more distributed than in FA but higher precision is still given to sensory streams compared to higher-order mental content. An effect specific to OM practice may concern the formation of prior expectations. Indeed the magnitude of the MMN is hypothesized to be dependent not only on the precision afforded to unexpected tones (deviants) but also on the strength (confidence) of the prediction about expected incoming tones (standards). This prior expectation is formed through the repetition of standard tones. Advanced OM meditation could potentially decrease this regularity extraction process and therefore lead to a reduced MMN, as the latter reflects the difference between “expected” and actual tone. In other words, deviant tones would be less surprising and so not elicit a strong prediction error (MNN). Interestingly, these predicted effects should also depend on the amount of practice, with experts practitioners showing stronger effects than novices for example. FA in experts could also gradually resemble aspects of OM. These predicted effects could also exhibit subtle differences following the meditative instructions. For instance, applying FA on the auditory modality versus the visual modality, or on focusing on the content of experience versus its monitoring aspect could impact sensory precision or prior formation. A schematic summary of these predictions about sensory precision and prior formation in general (not restricted to the MMN) is given in Figure 3.4. Notably, we do not expect a strong trait effect of meditative practice in this regard (i.e., when off the cushion) — for example, when comparing experts

and novices in an MMN paradigm during a rest condition — as permanently perturbing these low levels perceptual processes would impair normal functioning (e.g, in schizophrenia MMN is abnormally reduced [Damaso et al., 2015, Maekawa et al., 2012, Näätänen and Kähkönen, 2009]).

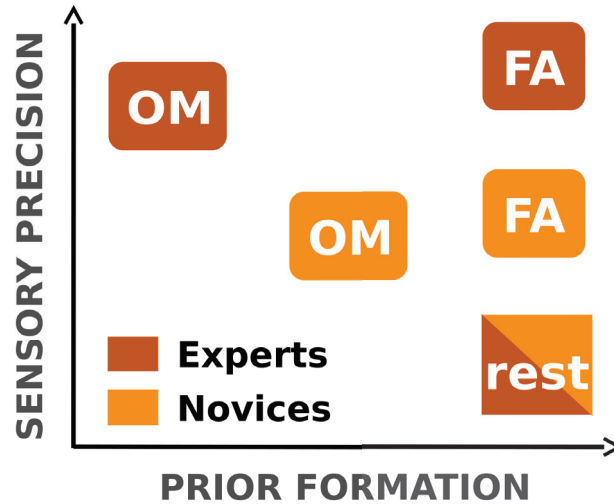


Figure 3.4: Expected effect of different meditative practices and degree of experience on sensory precision and prior formation processes. At rest, we would expect both novices and experts practitioners to show normal sensory precision and prior formation. During focused attention (FA), sensory precision would increase as an effect of attentional focus, which should be more stable in experts. During open monitoring (OM), sensory precision should still be higher than at rest and prior formation should decrease, with again stronger effects in experts than in novices.

To our knowledge, only three experimental works have examined so far the state (meditation vs. control) and trait (experts practitioners vs. novices or naive participants) effects of contemplative practices on MMN. The first research study on the topic compared Kriya Yoga practitioners to meditation and yoga naive participants [Srinivasan and Baijal, 2007]. The authors measured the MMN of yoga practitioners at rest and after three kinds of yogic practices culminating on a concentrative meditation on the breath (Sahaj Samadhi, FA style), and compared it to the MMN of control participants (naive to any yoga or meditation practice) recorded at rest and during three repetitions of a “relaxation” condition (reading a book). They found a main effect of group with the MMN of yoga practitioners being on average higher (more negative) than naive participants over all conditions (rest and meditation/reading). In yoga practitioners the concentrative condition further enhanced the MMN compared to rest and preliminary practices done before. These results are in agreement with our prediction of increased MMN during FA practice but the overall trait effect of meditation is quite surprising. However, as argued by Biedermann et al. (2016), this first study confounded different conditions between “experts” and novices [Biedermann et al., 2016], potentially leading to artificially enhanced overall MMN in experts. To remedy this potential bias, they compared experts practitioners in various contemplative traditions (Zen, Chan, Tibetan, all with more than 10 years of meditation practice) to an

age-matched control group of non-meditators, and importantly they recorded them during the same two conditions: an FA breath counting task, and a control task (imagining building a tree house). This study also revealed a general trait effect of meditation: an increased MMN in both meditation and non-meditation condition, but no state effect as the interaction between group and condition was not significant [Biedermann et al., 2016]. Of note, the trait effect was present at the Pz electrode but not at Fz (where MMN is usually stronger), and this study did not control for a potential condition-order bias, as the breath counting task always followed the control one. While this study seems to confirm a potential trait effect of meditation on the MMN, it did not replicate the state effect of FA practice found in the previous one, which was actually more expected.

Recent findings from our research group may be help to shed further light on these mixed results. Fucci et al. (2018) assessed the MMN during FA and OM meditations, as well as during a control reading task, in 16 highly experienced practitioners from the Tibetan tradition ($> 10,000$ hours of practice) and 15 age-matched meditation naive participants who received a brief meditation training and practiced for 7 days before the experiment [Fucci et al., 2018]. This design overcomes some limitations from the previous studies by comparing experts to control participants not totally naive to meditation, by using similar meditation and control conditions whose presentation order was randomized, and by analyzing the MMN averaged over an extended fronto-central region where it was shown to be maximal. Interestingly, Fucci et al. (2018) found a significant expertise by condition interaction, with post-hoc tests revealing trends for a higher MMN during FA but not OM (labelled OP in the article for open presence, an advanced form of open monitoring performed by experts) compared to the control condition in experts and, conversely, a higher MMN during OM but not FA, compared to the control condition, in novices. The MMN during FA and OM was not significantly different in neither groups of participants, and there was no group difference during OM and control conditions but experts presented a higher MMN than novices in FA meditation ([Fucci et al., 2018] and see Figure 3.5 for the main results).

Overall the results from these three preliminary studies on the effect of meditation on the MMN seem to support the hypothesis of an increased MMN during FA practice as an effect of enhanced sensory precision. However the only study investigating advanced OM practice (open presence) did not observe a decrease of the MMN compared to a control condition [Fucci et al., 2018], a state effect that we would expect if such practice effectively reduced prior formation (Figure 3.4). Besides there is no clear evidence for a genuine trait effect of meditation on the MMN as, contrary to previous studies, Fucci et al. (2018) found no differences between highly experienced practitioners and meditation novices during a control reading condition [Fucci et al., 2018]. The trait effect reported in previous studies could have artificially arisen from the use of different control conditions [Srinivasan and Bajjal, 2007], as well as comparing experienced practitioners to participants totally naive to meditation [Biedermann

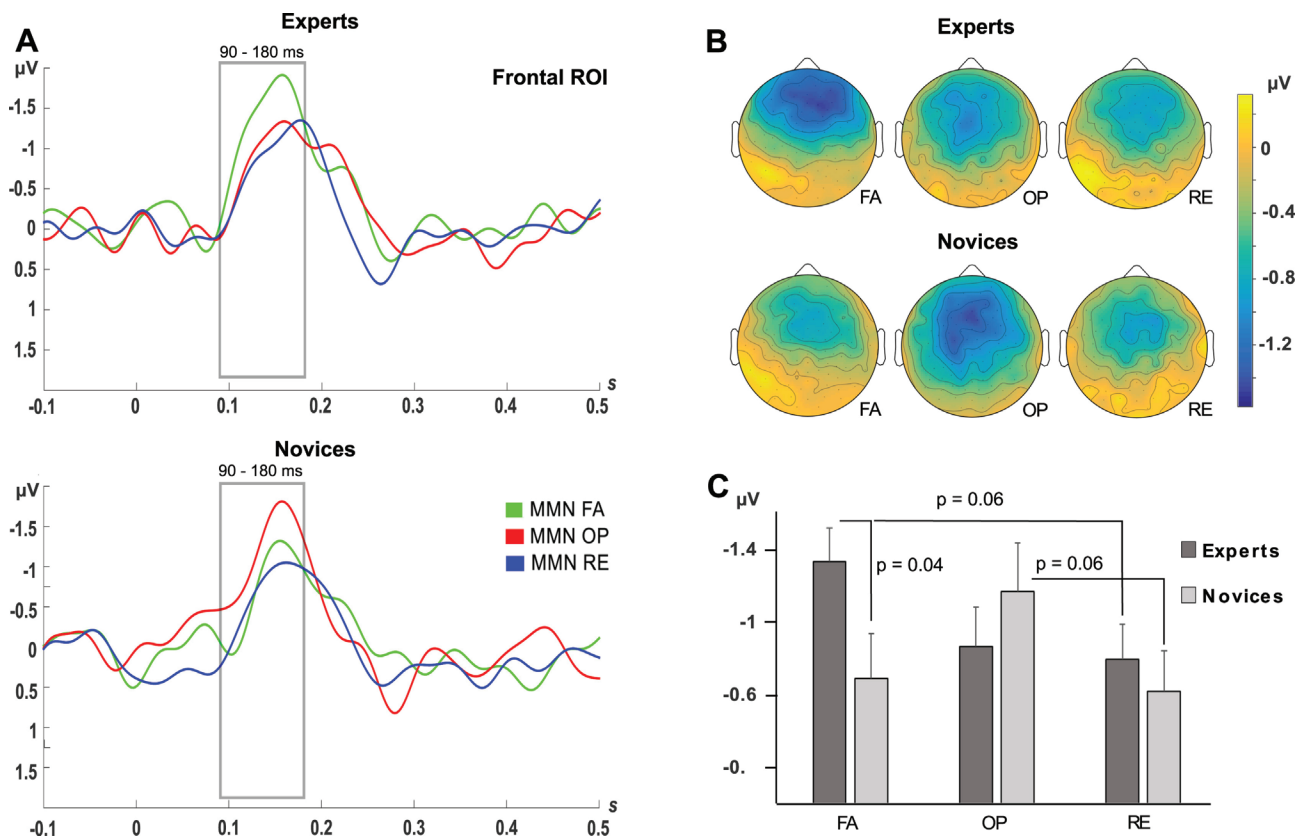


Figure 3.5: Main results from [Fucci et al., 2018]. The figure displays the MMN waveform at a frontal region of interest (ROI) for the different conditions and expertise groups (A), the related average scalp topography between 90 and 180ms (B), and the mean value of the MMN over the frontal ROI between 90 and 180ms +/- the standard error of the mean (C). *FA: Focused Attention, OP: Open Presence, RE: Reading.*

et al., 2016, Srinivasan and Baijal, 2007].

The replication article presented as Study 1 in this manuscript attempted to resolve these ambiguities by comparing again meditative states (FA, OM) and a control condition (video watching instead of reading), in highly experienced meditators and novice participants who followed a short meditation program based on similar teachings as the experts [Fucci et al., 2021]. Combining the results from both studies and using Bayesian statistics to assess the evidence for null effects, we found substantial evidence for an absence of differences in the MMN amplitude between meditative and control states as well as expertise group. The implication of these results on our hypotheses concerning the effect of meditation on predictive processing in the context of auditory deviance detection will be discussed in Chapter 6.

3.2.2 Seeing prior expectations as impermanent: pain processing under meditation

We saw earlier how meditative practice can modulate behavioral and neuronal correlates of painful experience, in particular its affective component (Section 1.2.4). From the perspective we have adopted, meditation effects on pain regulation could be mediated by the attentional shift toward exteroceptive and interoceptive sensations and away from (pre-)cognitive expectations — notably about incoming painful stimuli — a stance cultivated in meditation. Moreover, the increased opacity of mental processes promoted by contemplation, may allow the practitioner to become aware of the tendency to cognitively amplify physically aversive events. With practice, an increased awareness and the development of a non-judgmental, equanimous attitude towards anything that arises in experience may allow one to defuse habitual catastrophizing tendencies (“it is painful, it will never stop”, etc...), thus largely deflecting the “second arrow of pain” — i.e., the cognitive intensification of pain-related sensory data. This hypothesis was suggested by Porro and Pagnoni in 2014, who noted that brain imagery results on meditative analgesia indicated an increased activation in regions of the pain salience network (thalamus, anterior insula, MCC) but also a modulation of activity in areas implicated in affective/cognitive processes (ACC, OFC, dlPFC) [Pagnoni and Porro, 2014].

According to the Bayesian take on perception as inference, and our hypothesis on the mechanisms of contemplative practice, meditation should bias *posterior* pain experience toward the actual level of nociceptive input, decreasing the influence of short-term *prior* expectations about it. More specifically, we would expect meditation to increase the weight (precision) of the likelihood term in Bayes’ rule and reduce the weight of the prior. This does not necessarily imply that meditators would experience more intense pain and, in fact, most studies reported similar or reduced pain intensity ratings in meditators compared to controls (see Section 1.2.4). Also, a structural MRI study in long-term Zen practitioners revealed that decreased sensitivity to thermal pain (i.e., higher pain threshold) was associated with an increase of cortical thickness in the dorsal anterior cingulate cortex (aMCC) and secondary somatosensory cortex. Instead, we expect pain experience in meditators to be less vulnerable to intensification by affective and cognitive priors about pain, with a net effect of a decrease in pain unpleasantness (see Section 1.2.4 for some evidence of this). Another interesting, if somewhat counter-intuitive, corollary is that experienced meditators should be less susceptible to both nocebo and placebo effects [Pagnoni and Porro, 2014], a hypothesis that has begun to be put to experimental test (see below). Finally, if meditation does indeed help in reducing the habitual cognitive intensification of pain, it should also in turn build up a more general and long-lasting expectation of reduced suffering. Indeed, as discussed at length in the first chapter, this is the explicit aim of most contemplative practices, where suffering is taken in its broadest sense — i.e., concerning both oneself and other beings. Thus, the mounting confidence of such long-term belief of reduced suffering accrued through practice would conceptually

and experientially align with (i.e., provide evidence for) the soteriological high-level prior. In partial support of this notion, two studies from our research group showed that expert meditators reported less subjective pain catastrophism [Zorn et al., 2020] and employed different strategies to cope with pain compared to control participants [Poletti et al., 2021].

Despite an early scientific interest, to this day very few experimental work on meditation and pain expectations have been carried out. A first one from Taylor et al. (2018), replicated their previous fear-of-pain conditioning paradigm [Taylor et al., 2017] on a group of 11 experienced meditators (1050 to 9500 hours of practice) compared to a group of 51 meditation-naive participants (11 of which were selected to constitute an age-, gender- and education level matched control group, [Taylor et al., 2018]). While meditators significantly reported less intense pain to painful stimuli, no difference was found in the nociceptive flexion reflex (NRS), a measure of nocifensive spinal reflex, suggesting that pain-modulation by meditation may not be mediated by inhibitory descending control (suppression mechanism). Furthermore, no group difference in anticipatory skin conduction responses was detected, a results in apparent contradiction with previous findings of reduced brain activations (anterior insula, aMCC) in experts meditators during the anticipation of pain [Lutz et al., 2013, Brown and Jones, 2010]. However, these results can be potentially reconciliated if we assume that meditation may reduce the affective and cognitive anticipatory elaboration about incoming aversive events (as reflected in the anterior insular and cingulate cortice activity), while preserving survival-related physiological processes selected by evolution to prepare for a fight or flight response (as reflected by SCR). Finally, and of much interest for our current hypotheses, expert practitioners seemed to show a reduced influence of outcome expectations and their associated uncertainty on subsequent pain ratings [Taylor et al., 2018]. A more recent report suggested that exposing meditation-naive participants to a brief (10 min) mindfulness induction aiming at accepting incoming painful sensations reduced the effect of conditioned-hypoalgesia (placebo) compared to an experiential avoidance (suppression) induction ([Vencatachellum et al., 2021], Figure 3.6). Interestingly, this effect extended to both intensity and unpleasantness pain ratings but, contrary to the previous results from Taylor et al. (2018), was not present for conditioned hyperalgesia (nocebo). A possible explanation for the latter finding is that the modulation of increased pain expectations, compared to decreased pain expectations, may require a higher amount of meditative practice as they signal potentially harmful events.

While these preliminary results are interesting, and seem to point in the expected direction, they also have limitations, such as a low sample size of experienced practitioners [Taylor et al., 2018] or the use of a very brief mindfulness induction protocol [Vencatachellum et al., 2021], which is unlikely to reflect the refinement and modulatory power of a dedicated training over multiple years. Besides, they do not explicitly test the mechanistic predictions stated above concerning the potential effect of

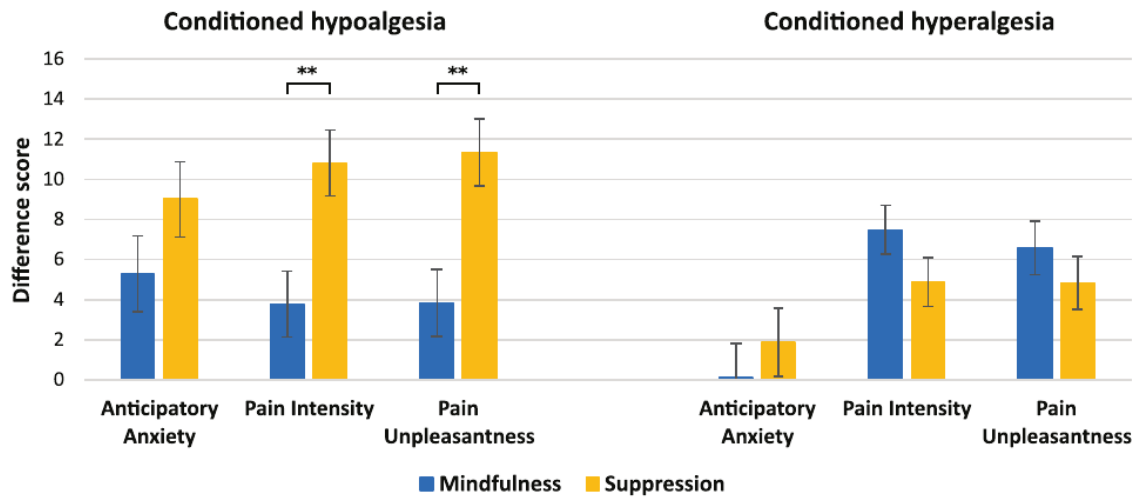


Figure 3.6: Main results from [Vencatachellum et al., 2021]. Hypoalgesia, measured as the difference in ratings between identical stimulation (medium level) following low-pain conditioned cues and novel cues (without conditioning), was present in the suppression (yellow) but not mindfulness (blue) group, for intensity and unpleasantness ratings. Both groups showed hyperalgesia, increased pain to high-pain conditioned cues compared to novel cues.

meditation on sensory and prior-expectations precisions. To remedy these shortcomings, we aimed to replicate an ingenious computational modeling work on a cue-based pain expectations paradigm [Hoskin et al., 2019], applying it in the context of meditation. Hoskin et al. (2019) presented their participants with two cues indicative of different pain intensity probabilities: for example, a 50% chance to receive either a level 4 or a level 7 stimulation, and 50% chance of receiving either a level 5 or a level 7 stimulation. Participants had to choose the option that best minimized the expected pain (e.g., in the case above, the option of a 50% chance to receive 4 or 7 should be picked). They then received a cutaneous electrical shock whose level depended on the probabilities corresponding to the chosen option, following which they rated their subjective pain experience in terms of felt intensity of the just received stimulation on a 0 to 100 visual analog scale.

The novelty in Hoskin et al. (2019) study is that the analysis of behavioral pain ratings and their dependency on expectations and uncertainty was conducted with computational models based on the predictive processing approach to pain perception. In a nutshell, and as depicted in Figure 3.7, Hoskin and colleagues modeled the *posterior* intensity ratings at each trial as the (Bayesian) integration of a two-component prior — a *cue-independent* (or trait-like) component about pain and a *cue-dependent* component reflecting the information provided by the cue about the stimulation level of the upcoming trial — and the actual stimulation level (i.e., the *likelihood*). Their models are twice Bayesian in the sense that priors and likelihood are combined by Bayes' rule and weighted by their respective precision (inverse variance); and that free parameters (encoding prior/likelihood mean or variances) are modeled as random variables with an associated probability distribution, which are therefore assigned a prior and whose posterior distribution can be obtained through Bayesian inference. The study found that a

“full” Bayesian model combining all three terms detailed above was better at explaining not only the data from their first experiment, but also those from a replication with a slightly different paradigm. More interestingly, a hierarchical version of the model with varying parameters per subject, proved even better at explaining both datasets. They notably found important differences between subjects in the cue-independent prior, which they related to a potential trait-like pessimism or optimism toward pain, at least in the context of their experiment.

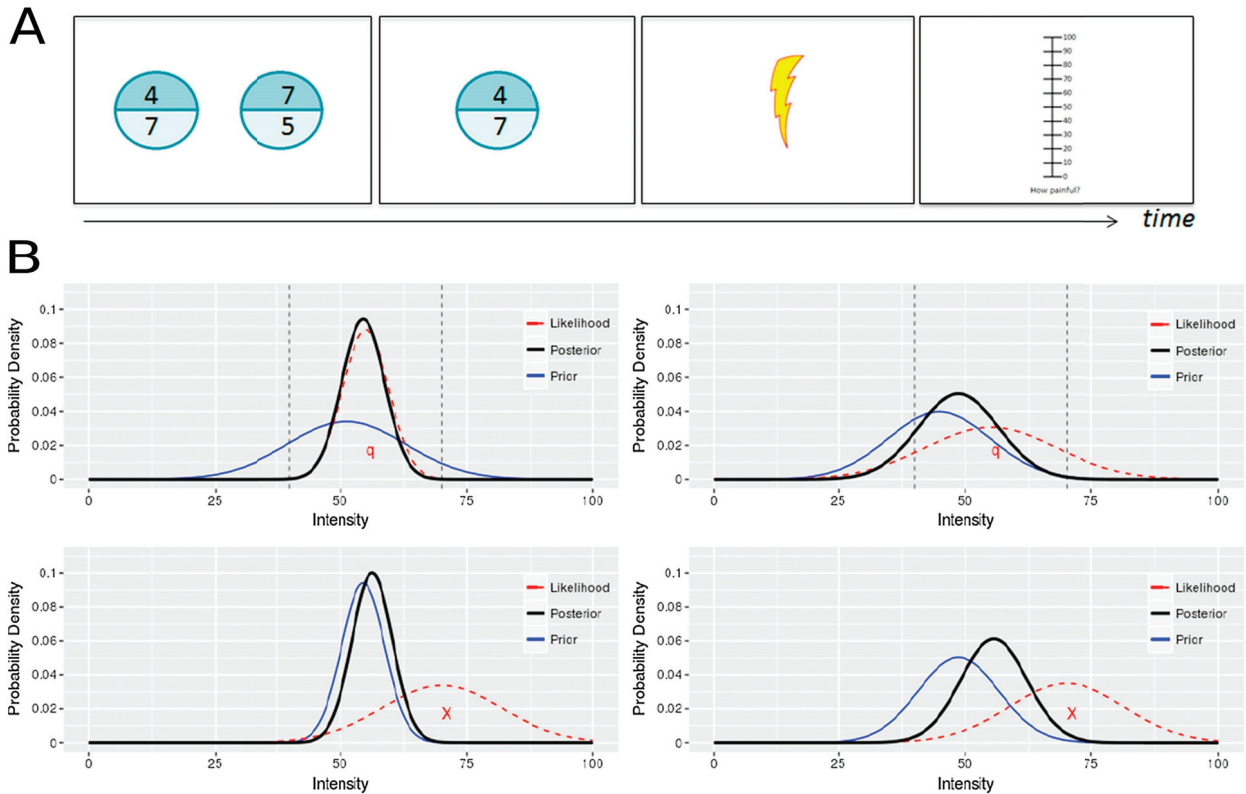


Figure 3.7: Experimental task and graphic representation of a Bayesian pain model from [Hoskin et al., 2019]. (A) The paradigm involved the choice by the participant of the “best” pain outcome based on two cues. In this example, one can either choose to receive a pain stimulation of level 4 or 7 with 50% chance each, or a stimulation of level 5 or 7 with 50% chance each. The participant should select the cue minimizing the expected pain, which is then displayed for 2s. The electrical stimulation is then delivered (in the illustrated case, either 4 or 7), and the participants rate the felt intensity on a 0 to 100 scale. (B) Graphical representations of Bayesian inference performed by the “full” hierarchical model at a given trial for two different subjects (left and right columns). In the upper panels, the probability distribution associated with the cue-independent prior (blue, different for the two participants) is first combined with the cue-dependent prior, which is treated as a likelihood for this computation (red; note that the mean expected stimulation given the cue, q , is identical for both subjects). The resulting posterior (black) forms the overall prior (lower panels, blue) to be combined with the delivered painful stimulation X (lower panels, red). The final posterior (lower panels, black) corresponds to the model-predicted value of the intensity ratings of the participants for this given trial. Since, in this example, the two participants had different (in mean and variance) cue-independent priors and different (in variance only) cue-dependent priors, they are characterized by different posterior ratings. *Adapted from [Hoskin et al., 2019].*

A main project of my PhD, the LONGIMED study, was to design an adapted version of Hoskin

et al. (2019) experimental task in the context of a longitudinal study of a 10-day intensive meditation retreat. The replication and extension of their modeling work to encompass also the affective dimension of pain (unpleasantness) is the topic of a first article on the baseline data of this study (Chapter 5, Study 3). Finally, we wanted to specifically test the effect of a 10-day meditation on predictive processing components of pain perception, as assessed by potential changes in Hoskin et al. (2019) Bayesian model's parameters encoding the stimulation level (likelihood) and cue-dependent prior variances (inverse precision) and the cue-independent prior mean and variance (Chapter 5, Study 4). The exact design of this randomized and controlled study which involved other experiments will be described in the next section. The analysis plan for the pain experiment was also preregistered online on the Open Science Framework (<https://osf.io/awtph>).

3.3 Studying the acute and lasting effects of intensive practice: insights from a 10-day mindfulness meditation retreat

3.3.1 Scientific studies of residential meditation retreats

Residential retreats are an important training component of many people’s ongoing meditation practice and, accordingly, they offer an ecologically valid research strategy for bridging empirical gaps between non-intensive interventions in novices and cross-sectional studies of expert practitioners.

[King et al., 2019]

As emphasized by King et al. (2019), who wrote the first theoretical piece on the topic not long ago, residential meditation retreats are not only a key component of most meditators’ practice, but also a very interesting and more ecological setting to scientifically study meditation in its more acute and traditional form [King et al., 2019]. It would be more accurate to talk about “forms” as there are many kinds of residential retreat, depending on the contemplative tradition, teaching, practice and culture. Yet, King et al. (2019) tried to extract the common ingredients to most forms of residential retreat [King et al., 2019]. First, a retreat imposes a (variably) rigorous schedule ranging from 6 to 11 hours of formal meditation practice, often completed by informal practices. Formal practice is performed during a dedicated period at a time (around 30 minutes to one hour, followed by a short break) and is usually done sitting in meditation posture with specificities depending on traditions and teachings (note for instance that some retreats allow practitioners to sit on a chair or lie down for formal practice). Informal practices are also often encouraged and involve the extension of the mental set brought to bear during formal practice (e.g., paying attention to internal or external objects, being aware, a non-judgmental, non-reactive attitude, etc.) to any other habitual activity (walking, eating, dish-washing, etc)⁷. To facilitate the adherence to such a demanding schedule and the extension of the meditative attitude outside of formal sitting, residential retreat are typically located in secluded places offering a serene and calm environment, where the practitioners’ basic needs are tended for⁸.

⁷Walking and other movement based activities, such as yoga or qi-gong, are sometimes taught as formal practices in their own right.

⁸Note that, depending on the type of retreat, essential tasks such as cooking or cleaning can be either (a) included in the costs and taken care of by the retreat centers’ employees, (b) performed in turn by the participants to the retreat (this is best exemplified by Zen Buddhism retreats where work is an essential part of the practice), or (c) performed by other meditators acting as volunteers, as is usually done in Vipassana retreats (with a similar emphasis as Zen on the integration of practice in daily activities and on the service rendered to

The creation of a setting favorable to the calm concentration required for practice is also the result of a specific attitude brought forth by the participants. “Noble” silence is most often the rule (people refrain from speaking except in case of urgent concerns, but do not communicate in any other ways or initiate eye contacts), and the use of electronic devices or other medias, intoxicants such as alcohol, smoking or drugs is proscribed, as is sexual intercourse. These restrictions are put in place not to enforce a punitive or authoritarian regime, but to minimize distractions that could feed the “monkey mind” and perturb meditative practice, while also helping to break the loop of ingrained habits. As beautifully put by King et al. (2019): “retreats are designed to facilitate uninterrupted attention and reflective awareness toward ongoing mental and sensory phenomena”, and “this training structure expands the definition of meditation — often conceptualized as a delimited, formal activity — into an embodied and intentional stance toward experience that informs one’s values, ethics, and attitudes in all areas of life.” According to these authors, and the teachings of most contemplative traditions, residential retreats establish the ideal setting for developing meditation as a “way of life” and help foster a deeper impregnation of its multiple benefits to perceptual, affective, cognitive and social domains of human life — against all odds ⁹.

As reported by King et al. (2019) in their review article (Figure 3.8), the effect of meditation interventions most likely depends on their intensity (hours per day), their duration (number of days), and the initial expertise of the participants (totally naive or already acquainted to meditation). Residential meditation retreats, no matter their length, are inevitably more intense and recruit more experienced practitioners than experimental induction treatments and standardized mindfulness-based interventions (MBIs, app-based or 8-week live or online courses). Concerning duration, the picture is more complex as MBIs usually last longer (typically 8 weeks) than most intensive retreats (7 days to 1 month), yet their daily “dose” of meditation is much lower. Interestingly, King et al. (2019) also reported the proportional number of longitudinal scientific studies that explored the effects of each of this different intervention formats (blue circles in Figure 3.8): induction and 8-week courses are by far the more studied cases. Indeed, scientifically studying meditation through a direct short induction in the lab or through standardized interventions in a therapeutical setting is much easier than during an intense residential retreat held in a remote place and in total silence. Yet, as noted earlier, residential retreats are a really important ingredient in most practitioners meditative development and their study may inform us on the “dosage” requirements for some effects to occur, the importance of context (setting, teaching, silence, etc.), the mechanisms specific to intensive practice and their individual variations, which may be difficult to reproduce in a non-ecological laboratory setting [King

other as compassion in action).

⁹Indeed and similarly to the pragmatic value of contemplative “in-action” discussed previously (section 3.1.2), it is at first counter-intuitive to see how a calm and silent environment may help to deal with “real life” stressful and often violent problems.

et al., 2019]. Last but not least, a retreat setting offers an almost perfectly controlled environment: all participants have the same schedule and diet, are exposed to the same environmental disturbances (which are rare, due to the seclusion of retreat centers), and share the same meditation schedule with the same instructions and teacher.

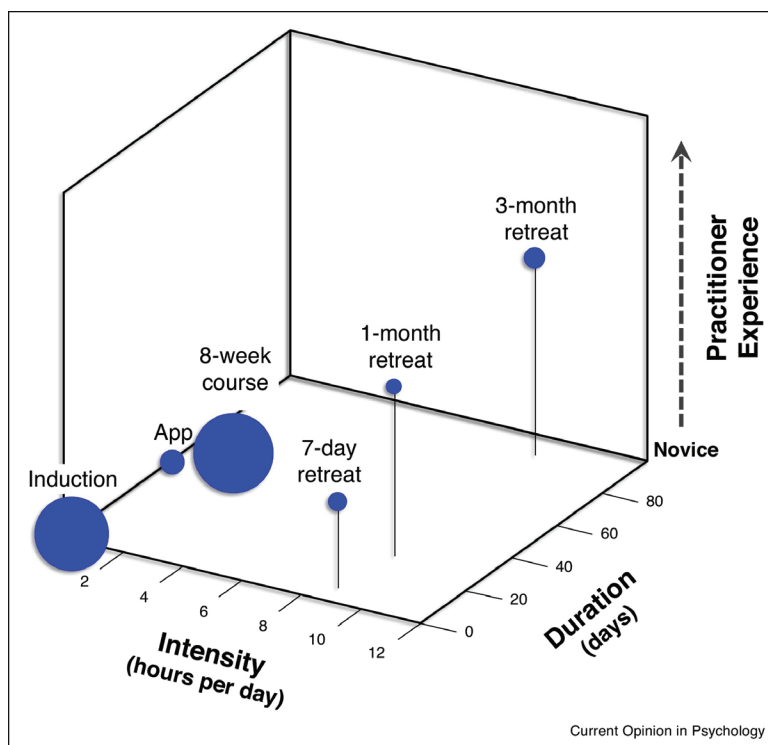


Figure 3.8: Intensity, duration and expertise attributes of meditation interventions in the literature. Typical meditation interventions used in scientific studies are mapped onto a matrix with three dimensions: intensity in hours per day, duration in number of days, and initial expertise of the participants (composite score depending on cumulative lifetime hours, years of practice and number and length of attended residential retreats). The size of blue circles represent the proportional number of published studies for each kind of intervention. “Induction” are single-session meditation protocols carried out in the laboratory, “App” includes mobile-based interventions and “8-week” refers to non-intensive standardized mindfulness-based interventions (MBSR, MBCT and derivatives). *Reproduced from [King et al., 2019].*

One of the most impressive study in the field to date is the Shamatha Project [Saron, 2013] which aimed to investigate the effects of traditional *shamatha* or “calm abiding” Tibetan Buddhist meditation (FA style), as taught by Alan Wallace [Wallace, 1999] during two 3-month residential retreats. This controlled and randomized longitudinal study measured 30 participants with ongoing meditation experience before, during (beginning, middle and end) and after (at 6-months, 18-months, and up to 7-years later) the first residential retreat which included about 6 hours of individual practice a day as well as meditation instructions, teacher talks and discussions twice a day. A second sample of 30 participants matched in age, gender and meditation experience was recruited into a waiting-list control group (group assignment was randomized) and also assessed longitudinally before participating to a second 3-month retreat held after the first one. A large number of experimental measures were

collected in this study, including changes in attention, perception, emotional regulation and their brain correlates (EEG), psychological scores (questionnaires and interviews), physiological variables (heart rate, respiration, skin conductance), and biological markers of stress, inflammation and aging (blood and saliva samples). Taking place in spring and fall of 2007, the Shamatha study led to an impressive number of publications, with the first one in 2010 [MacLean et al., 2010] and results still being published today [Shields et al., 2020, Zanesco et al., 2021]. Summarizing all their results is out of the scope of this manuscript ¹⁰, but King et al. notably clarified how some of them (as well as results from other intensive residential retreats) were specific to the intensity of practice brought forth in these “high dose” interventions. Of particular interest to us in the current work are findings of increased sustained and selective attention [MacLean et al., 2010, Slagter et al., 2007, Lutz et al., 2009, Zanesco et al., 2019], response inhibition [Sahdra et al., 2011], executive control [Zanesco et al., 2013, Shields et al., 2020] and emotional regulation [Rosenberg et al., 2015], but decreased mind-wandering [Zanesco et al., 2016] during and following intensive retreats. Importantly, authors from the Shamatha project found that some of their initial effects (sustained attention and response inhibition) were partially maintained seven years later, and that this maintenance was modulated by the continuity of meditation practice after the retreats [Zanesco et al., 2018].

A more general review of all studies on residential meditation retreat up to 2017 found large effects on psychological measures of anxiety, depression and stress, and moderate effects on measures of emotional regulation and quality of life [Khoury et al., 2017], with the increase in mindfulness strongly mediating these effects. However this result is based on just 21 studies testing meditation effects during or after a retreat and therefore should be evaluated *vis-à-vis* the much larger amount of evidence on mindfulness-based interventions (see Figure 3.8). Furthermore, studies able to implement on-site measures *during* residential retreats as the Shamatha project are even scarcer. A primary objective of my PhD was to partly fill this gap by testing the effects of a 10-day intensive mindfulness meditation retreat on various perceptual and cognitive processes. We will now describe this controlled and randomized longitudinal study titled “LONGIMED”.

¹⁰But see <https://saronlab.ucdavis.edu/shamatha-project.html> for an overview of the study, their main findings and a list of all publications.

3.3.2 The LONGIMED study

Scientific aspects: design, experiments and questionnaires

In order to investigate the longitudinal effect of intensive meditation practice on predictive processing mechanism we recruited 54 participants with variable meditation experience. Inclusion and exclusion criteria are detailed at length in the clinical trial record of the study (<https://clinicaltrials.gov/ct2/show/NCT04449913>), which received approval from an independent ethics committee on human research (CPP Est IV 2020-A00669-30). Important inclusion criteria were: age between 18 and 67, speaking french fluently, having been trained in a formal meditation practice (mindfulness, MBSR, Buddhist traditions), having a regular practice of at least 20 minutes per day three times a week for at least one year, and having already participated to a silent meditation retreat of at least two days with a minimum of 6 hours of meditation per day. Stringent exclusion criteria were: neurological or psychiatric antecedents, chronic pain or any other medical condition causing acute pain or increased sensitivity to pain, motor or sensory deficit in the hands, severe auditory loss, the regular use of opioids or antidepressant drugs, or more generally any drug or prescription acting on the central nervous system. Participants were randomly assigned to an “active” or to a “control” group in equal proportion (27 each, age- and gender-matched), whose differences will be explained below. Upon their first visit to the laboratory, they received a detailed explanation about the study and the different measures they will undergo before signing an informed consent and passing an interview with the author of this manuscript to assess their meditation experience. More specifically, the following information was collected: (1) the first type of meditation they practiced, (2) the one they were currently practicing, (3) an estimation of the time (in hours) of practice every year since they started meditating (lifetime practice), (4) the duration of their current daily practice (in minutes), (5) an estimation of the overall time (in hours) spent in meditation retreats up to the time of entering the study, and (6) what was the meditation type they have practiced the most during retreats.

All participants took part in three behavioral and EEG experiments with different timelines depending on group assignment. We will briefly present below the experimental paradigms and hypotheses for two of those experiments (on auditory and tactile perception) whose data are still under analysis and thus will not be presented in this thesis. The first one is an auditory experiment adapted from [Demarchi et al., 2019] that presented participants with two tone sequences of different predictability while recording their EEG. Specifically, sequences differed in their presentation of four pure tones (200 Hz, 431 Hz, 928 Hz, 2000 Hz; 100ms duration, 3 Hz rate) following each other with 75% probability in the “ordered” sequence or being presented in a completely stochastic fashion (equal 25% probability for all tones) in the “random” sequence. Participants received no explicit information about the sequences

and were instructed to either perform a focused attention (FA) meditation on the tones, an open monitoring (OM) meditation without particular focus on the auditory modality, or a control condition of watching a calm marine documentary. They were delivered four tone sequences of four minutes each (two ordered and two random) per condition and were asked to fill phenomenological scales about their experience after each sequence (tone clarity, sequence rhythmicity, attentional focus, meditation stability, meta-awareness, and equanimity to the sounds). The presentation order of sequences and meditation conditions was randomized within (longitudinal measures) and between participants. In their initial article on MEG data, [Demarchi et al., 2019] observed differences between predictability conditions (four conditions, with graded predictability from ordered to random) in brain activity patterns preceding a sound (or an omission), which were interpreted as markers of auditory predictions. They further used this protocol to assess how patients with tinnitus differed from healthy subjects and found stronger anticipatory predictions in this pathology [Partyka et al., 2019]. According to the hypotheses proposed in this thesis, we would expect our FA meditation condition to also increase the magnitude of such markers of auditory predictions. Indeed the attentional focus set on the auditory stream should amplify auditory prediction errors, thereby improving the precision of subsequent predictions and cause stronger pre-stimulus signals. On the other hand, we expect OM meditation to decrease the formation of anticipatory predictions, leading to lower pre-stimulus signals.

The second experiment is an adaptation of the force-matching task [Shergill, 2003] taking into account recent developments [Bays et al., 2005, Walsh et al., 2011, Kilteni and Ehrsson, 2017]. Participants were presented with pressure forces of varying intensity (1 to 3.5 Newtons, fixed 3s duration) on the fingertip of their left index finger via a motorized lever, and were then asked to reproduce the felt pressure after each trial in three difference conditions. First they had to press their right finger 20 cm apart from the left one to control (through a synchronized sensor/motor coupling) the lever and replicate the previously experienced pressure. In a second condition they had again to replicate the experienced force, but this time by pressing directly on their left index finger with their right one. Finally, the third condition asked participants to reproduce once more the force, but this time through the lever which remained on top of their left index. Two devices measured the forces applied by the volunteers: a ring sensor newly developed by the Tribology and Systems Dynamics Laboratory (LTDS UMR 5513, Ecole Centrale Lyon, “Touchy Finger”), which directly assessed the deformation of the right index finger pad (force application), and a more commonly used pressure sensor placed under the left index finger (force reception). Importantly, participants sat in a meditative posture (on their knee) to perform the task and were prompted to meditate using their finger sensations as the attentional target for one minute before the start of each condition (30 trials). They were instructed to replicate the delivered force intensity as best as they could, and maintain it for about 5s to obtain a “clean” measure. Conditions order was randomized within and between participants.

The two first experimental manipulation described above exhibited consistent differences, with the

replication of the force with a direct contact between fingers leading to a stronger force applied by the participants [Bays et al., 2005, Walsh et al., 2011, Kilteni and Ehrsson, 2017]. This effect, known in the literature as a process of sensory attenuation, is thought to be a crucial aspect of active inference [Brown et al., 2013], namely the precision down-weighting of sensory prediction errors to allow the fulfillment of motor predictions by an overt motor act (see section 2.1.3 for a more detailed explanation). We adopted this experimental paradigm to test specific hypotheses about active inference regulation through meditation. In a nutshell, we posited that the “in-active” meditative stance would affect such sensory attenuation mechanisms. The higher-order policy of immobility set at the beginning of a meditation practice would lead to a down-weighting of movement-related predictions, while the heightened attention (either sharply focused in FA or more distributed in OM) to bodily sensory signals would increase their precision. This synergy of amplified sensory prediction errors and decreased strength (precision) of motor predictions, would facilitate in turn meditative stillness. We expect this effect to be enhanced during an intensive 10-day meditation retreat and measurable through the force-matching paradigm as a decrease of sensory attenuation.

Our third and last behavioral experiment is a modified version of Hoskin et al. (2019) cued pain task [Hoskin et al., 2019]), which we already described in Section 3.2.2), and whose results will be presented in detail in Study 3 (Chapter 5).

The main difference between our two groups of participants is the time at which they underwent the three experiments. “Active” subjects were first tested “on-site” upon their arrival at the retreat center, one to two days before the beginning of the retreat (baseline visit/measure, V1 in Figure 3.9), then 7 days later at days 6 or 7 of the retreat (V2, retreat measure) depending on the date of their first measure, and a last time at a minimum of three weeks after their second measure (V3, follow-up measure, between 21 and 30 days, except for one participant who could only come back 40 days after the end of the retreat because of medical issues). This last measure was not collected as the two previous ones “on-site” at the retreat center, but at the laboratory. To control for a potential habituation to our experimental tasks that would explain a difference in results at V2, we also tested “control” subjects three times with the same delay between measures as the “active” group (7 days between V1 and V2 and between 21 and 30 days between V2 and V3) but in the laboratory and before their participation to the retreat (Figure 3.9).

Importantly, “control” participants were recruited with the same criteria as “active” ones and the partaking of both groups to the retreat ensured a similar motivation to participate to the study and perform the experimental tasks — group attribution was simply indicated as group 1 or group 2 to the participants. Some additional measures were also collected independently of group attribution and analyzed for the 54 subjects sample as a whole (“Both” line in Figure 3.9), further deconstructing the potential devaluation that could have come from a “control” vs. “active” dichotomy. This included

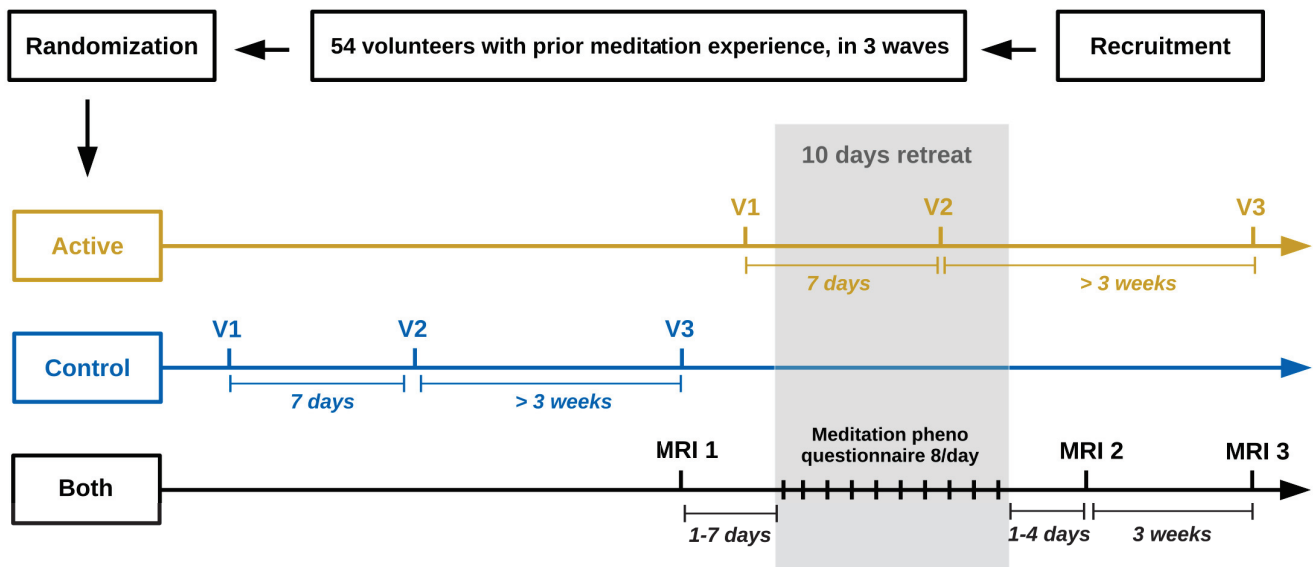


Figure 3.9: Design of the LONGIMED study. 54 volunteers with ongoing meditation experience were recruited and randomly attributed to an “active” or “control” group (27 each). “Active” participants underwent the behavioral and EEG experiments just before (V1, -2 to -1 days), during (V2 at day 6 or 7 of the retreat) and at least three weeks after a 10-day intensive mindfulness retreat (V3, +21 to +30 days). “Control” participants took part in the experiments with the same timeline but before the retreat. Finally, we collected structural and functional MRI scans for both groups before (MRI 1, -7 to -1 days), just after (MRI 2, +1 to +3 days) and minimum three weeks after the retreat (MRI 3, +21 to +31 days). All participants to the retreat also completed a battery of psychometric questionnaires during their first visit in the laboratory (V1 for “control” and MRI 1 for “active”), as well as a phenomenological questionnaire after every meditation sessions of the retreat (8 sessions per day, for 10 days).

one structural MRI and three fMRI scans (resting state, FA and OM meditation conditions) which took place before (MRI 1, -7 to -1 days), just after (MRI 2, +1 to +3 days), and at least three weeks after the retreat (MRI 3, +21 to +30 days). Participants also completed a battery of psychometric questionnaires (French versions from validated or internal translations) during their first visit at the laboratory (V1 for “control” subjects and MRI 1 for “active” subjects) which included (in alphabetic order):

- BFI: Big Five Inventory (short version) [Plaisant et al., 2010]
- BIS–BAS: Behavioral Inhibition/Activation Scale [Caci et al., 2007, Carver and White, 1994, Demianczyk et al., 2014, Morean et al., 2014, Studer et al., 2016]
- DDS: Drexel Defusion Scale [Forman et al., 2012]
- FFMQ: Five Facets Mindfulness Questionnaire [Baer et al., 2006, Heeren et al., 2011]
- FLANDERS: Flinder Handedness Survey [Nicholls et al., 2013]
- MAIA: Multidimensional Assessment of Interoceptive Awareness [Mehling et al., 2012, Edwige, 2014]

- PHQ: Patient Health Questionnaire [Kroenke et al., 2010, Norton et al., 2007, Spitzer et al., 1999]
- PCS: Pain Catastrophizing Scale [Sullivan et al., 1995, French et al., 2005]
- PTQ: Perseverative Thinking Questionnaire [Ehring et al., 2011, Devynck et al., 2017]
- RRS: Ruminative Response Scales [Treynor et al., 2003]
- SWBQ: Spiritual Well-Being Questionnaire [Gomez and Fisher, 2003]

Participants additionally completed an adapted version of the MEDEQ (Meditation Depth Questionnaire, [Piron, 2001]) at the end of the retreat, assessing their global experience of the retreat, the Life Change Inventory Revisited (LCI-R, [Greyson and Ring, 2004]) during the follow-up visit (MRI 3), as well as some of the questionnaires already completed at baseline to evaluate potential changes occurred since the retreat (BIS-BAS, DDS, FFMQ, MAIA, PCS, PTQ, RRS, SWBQ).

During the retreat, no distinction was made between participants (except when “active” subjects had to take part in the V2 measure at day 6 or 7) and they all completed a 5-minute phenomenological questionnaire on tablets after every meditation session of the 10-day retreat (8 times/day). Since, similarly to behavioral experiments 1 (auditory perception) and 2 (tactile perception), the data from this questionnaire are still under evaluation, they will not be included in this thesis but we will nonetheless briefly describe their content and goal.

The questionnaire was based on previous phenomenological inquiries of meditation [Kok and Singer, 2017, Abdoun et al., 2019, Petitmengin et al., 2017], published phenomenological models [Lutz et al., 2015], as well as extensive experience of members from our team (both in day-to-day and retreat settings) in various contemplative practices (Zen, Vipassana, Tibetan, mindfulness). It notably received valuable feedback from the meditation instructor of our retreats who spent himself 15 years in Tibetan Buddhist residential retreats and had over 80,000 hours of meditation experience. A preliminary version of the questionnaire was first tested by a member of our team during a 10-day Vipassana retreat and during a mindfulness meditation retreat on experienced practitioners, and was revised according to acquired feedback.

The final version was composed of seven domains. The first one evaluated some general information about the meditation session which had just been completed: (*a*) the overall instructions given to participants, (*b*) an estimation of the percentage of time they were actively trying to meditate, (*c*) as opposed to sleeping or dozing, (*d*) their degree of fatigue, and (*e*) how much doubt or confidence they felt in the practice “efficacy” and their ability to meditate. Then, other five main domains were inquired : affect, physical stability, mental stability, attention/awareness, and thoughts. Each domain contained between 4 and 10 questions exploring various dimensions proper to this domain (e.g. for affect: main

emotions felt, pain intensity and unpleasantness, equanimity). Finally, the last domain assessed more occasional experiences which may not happen at every session. First, we asked if participants had any “insights” (defined as “the impression of gaining some new understanding” about their meditative practice, the teaching, their mind, their life, or anything else) and, if so, how intense was it. Second, potential adverse events that could occur during an intensive retreat, as well as unusual positive experiences, were also investigated based on previous work [Lindahl et al., 2017]. If participants reported such experiences during the last-performed session, they also had to indicate its intensity, emotional valence (negative or positive), and how unusual it was for them, and were encouraged to talk with the meditation instructor about it if they felt the need to do so.

This study had a very low dropout rate, with only the third measure not collected on two participants. One control participant withdrew from the study after the second experimental measure (V2) for personal reasons and one active participant had to quit the study after completion of the retreat due to an adverse event unrelated to meditation.

We just reviewed most scientific aspects of the LONGIMED study also to underscore its complexity, which required a large amount of dedicated time (more than one year) and problem solving. We will now describe the meditation retreat which also represented an important organizational and logistic challenge, especially given its coincidence with the COVID-19 pandemic.

Contemplative aspects: residential retreat logistics and teachings

Due to multiple logistic constraints — in particular, the need to record a large number of participants in a short time period during and after the retreat, and the ongoing Covid-19 epidemic at that time — we organised three separate 10-day mindfulness meditation retreats between fall 2020 and spring 2021 in the IZARIAT retreat center (<https://izariat.com/>) at Le Poizat-Lalleyriat (Ain, France). All retreats followed the same structure but, because of initial difficulties in recruitment and Covid-19’s sanitary restrictions, they varied in the number of participants: 7 “active” (6 females, 1 males) and 4 “control” participants (2 females, 2 males) took part in the first one, 11 “active” (7 F, 4 M) and 12 “control” (8 F, 4 M) in the second, and 9 “active” (5 F, 4 M) and 10 “control” (5 F, 5 M) in the third and last retreat (one control participant enrolled in the third retreat withdrew from the study before participating to it). The first retreat had to stop one day earlier, due to the Covid-19 pandemic peak at the end of October 2020 and the announcement of a second national lockdown by the French government. We were able to still perform most of our scientific measures, but participants to this first retreat (who were only 11) only had 9 days of intensive practice instead of 10.

The three retreats were supervised by Stephane Offort, a qualified meditation instructor trained

in the Karma-Kagyü tradition of Tibetan Buddhism under the guidance of Guendune Rinpoché. He notably took part in a traditional 3-year 3-month 3-day Tibetan Buddhist retreat and further developed his meditative practice during twelve years of residential retreat. He has since been teaching mindfulness meditation in both its secular and traditional forms to a large European audience for the past 10 years, being also qualified as an MBSR instructor. For these particular “scientific” mindfulness retreats he gave a fully secular teaching, with instructions mainly following the typical program of mindfulness-based interventions. The retreats took place in silence, communication was restricted to the bare minimum (except during discussion with the meditation teacher and necessary interactions with the research team), and participants were encouraged not to use electronic devices, especially smartphones. The daily schedule included approximately 7 hours of formal group meditation split in eight 45 to 60 minutes sessions. Participants had the choice to meditate on a cushion or a chair and a special emphasis was given on finding a comfortable yet active posture. Most sessions were continuously guided by the meditation instructor except for a third (beginning of the retreat) to a half (end of retreat) of them, during which practitioners were given instructions beforehand and left to meditate in group without the instructor. A one-hour group session of questions and answers with the instructor was provided at the end of each day.

Briefly, the first three days of retreat consisted in developing attentional focus while calming the mind by directing one’s attention to breathing and body sensations (FA family of practices). Other sensory modalities such as hearing and vision were then progressively inserted from day 4. Starting at day 5, the instructions encouraged participants to widen and loosen the attentional focus to encompass any experience that may arise in the field of awareness, a practice described as “open consciousness or presence” (OM practice). Participants were further instructed in the next days to allow thoughts and emotions to come and go in this open consciousness. Specific instructions on how to let go of difficult and particularly “sticking” thoughts and emotions were given, for example by pointing to their impermanent and elusive nature. Participants were also encouraged to adopt an equanimous stance toward mental and emotional experiences, without judging or reacting to them or, if they did so, by remaining aware of them and of their mental and physical correlates. During the last days (around days 7 to 10), the instructor would let more sessions go on unguided or with little guidance to give some independence to the participants, with the general instruction to remain in an open consciousness or, alternatively, to focus on a given sensory target if the mind was too distracted. Finally, brief instructions of compassion toward oneself and other beings were sparsely provided throughout the retreat, with added emphasis during the last days. One or two sessions of meditation in motion, inspired by yoga and qi gong, were also given by the instructor every day of the retreat, with again an emphasis on being aware of movements and the sensations they engendered.

The general instructions, duration, and guidance (guided or not) of every meditation session was recorded by the instructor.

“Active” participants who underwent the three behavioral and EEG experiments during the retreat had to miss around 3 to 4 hours of meditation at day 6 or 7, depending on when they underwent their first measure. We arranged to have the experiments distributed throughout the day so that participants could maintain a continuity of practice. The choice of recording the “retreat” measure at day 6 or 7 and not the last day of the retreat was based on personal experience from the investigators of the study (PhDs Antoine Lutz, Giuseppe Pagnoni, Oussama Abdoun and myself) who all participated to 10-day and longer intensive retreats, and completed by the teacher experience of the meditation instructor Stephane Offort. We posited that during a 10-day meditation retreat the “peak” effect of the practice is attained around these days, after a usual halfway “low point” (day 5), and not at the end of the retreat, as could be expected from an higher cumulative “dose” of meditation. From our students’ and teacher’s experience, in the last days participants to a retreat already project themselves past the end of the retreat, potentially diminishing the effects of intensive practice.

The scientific and contemplative elements of the LONGIMED study should be kept in mind when we get to the description of the experimental methods, results and interpretation of the pain task in Chapter 5 (Study 3 and 4).

Part II

Experimental work

Chapter 4

Predictive coding and meditation in auditory perception

Following the previous introductory chapters, we now have all the theoretical elements and hypotheses needed to fully apprehend our experimental work and their associated results. In this fourth chapter, we will first present a manuscript, published in a peer-reviewed journal, where we investigated the effects of two meditation practices and of meditation expertise on auditory deviance detection, as measured by an auditory oddball EEG data [Fucci et al., 2021]. This dataset was collected by another PhD student in our research group during the cross-sectional study composing the first part of the ERC grant BrainandMindfulness (for details of the protocol see [Abdoun et al., 2019]). We could not replicate some of the findings reported in a similar dataset [Fucci et al., 2018]. Instead, we reported a null state effect of FA and OM meditation practices on the mismatch negativity (MMN), as well as a null trait effect on the MMN when comparing novices and experienced practitioners (Study 1).

To better understand this null result, we used a computational modeling approach of the MMN to refine our understanding of these processes. We first modeled the control state of the EEG datasets from Fucci et al. (2018) and Fucci et al. (2022), which was not a meditative condition. We compared two main hypotheses about MMN generation present in the literature, namely adaptation and predictive coding accounts. By confronting the predictions from computational models simulating these hypotheses with observed EEG data, we were able to show that both adaptation and predictive coding-like mechanisms are needed to explain the MMN. This work is the object of a submitted publication at *PLOS Computational Biology* (Study 2). The modeling of meditation states will be continued after this PhD.

4.1 Study 1: No effect of focused attention and open monitoring meditation on EEG auditory mismatch negativity in expert and novice practitioners



Contents lists available at ScienceDirect

International Journal of Psychophysiology

journal homepage: www.elsevier.com/locate/ijpsycho

No effect of focused attention and open monitoring meditation on EEG auditory mismatch negativity in expert and novice practitioners

Enrico Fucci^{a,*}, Arnaud Pouban-Couzardot^{b,1}, Oussama Abdoun^{b,1}, Antoine Lutz^{b,*}

^a Institute for Globally Distributed Open Research and Education (IGDORE)

^b Lyon Neuroscience Research Center, Lyon, France

ARTICLE INFO

Keywords:
Mindfulness
EEG
MMN
Replication
Alpha power
Meditation

ABSTRACT

The auditory mismatch negativity (MMN) is a well characterized event-related potential component which has gained recent attention in theoretical models describing the impact of various styles of mindfulness meditation on attentional processes and perceptual inference. Previous findings highlighted a distinct modulation of the MMN amplitude by different meditation practices and degrees of expertise. In the present study, we attempted to replicate results from the recent literature with a data sample that allowed for increased statistical power compared to previous experiments. Relying on traditional frequentist analysis, we found no effect of focused attention and open monitoring meditation on the auditory MMN amplitude compared to a control condition (silent movie) in expert or novice practitioners (all $p > 0.17$), providing a non-replication of our previous work (Fucci et al. 2018). Using a Bayesian approach, we found strong evidence against an interaction effect on the MMN amplitude between expertise groups and meditation practices ($BF_{01} = 11.0$), strong evidence against effects of either meditation practices compared to the control condition (BF_{01} between 11.9 and 16.1) and moderate evidence against an effect of expertise during meditation (BF_{01} between 5.3 and 7.9). On the other hand, we replicated previous evidence of increased alpha oscillatory power during meditation practices compared to a control state ($p < 0.001$). We discuss our null findings in relation to factors that could undermine the replicability of previous research on this subject, namely low statistical power, use of flexible analysis methods and a possible publication bias leading to a misrepresentation of the available evidence.

1. Introduction

The understanding of the neurocognitive processes involved in mindfulness meditation has increased considerably over the last few years. Neuroimaging studies have identified neuroanatomical and neurophysiological markers of attention and emotion regulation as correlates of state and trait effects in different meditation styles and groups of practitioners (see Deolindo et al., 2020; Fox et al., 2016; Lutz et al., 2008; Tang et al., 2015 for extensive and critical reviews). The investigation of biological markers of mindfulness meditation plays a major role in apprehending the mechanisms underlying mindfulness-based interventions in clinical and other applied settings. Neurophysiological data is also essential to the operationalization of meditation practices into psycho-cognitive models and to support a dialogue between contemplative and scientific inquiry on fundamental research questions (e.g. Dahl et al., 2015; Lutz et al., 2015). Based on these

premises, assessing the validity and replicability of the evidence for the modulation of relevant neural markers by meditation states and expertise is a necessary and critical task. In this article, we report the outcome of a replication study assessing the reliability of the auditory mismatch negativity (MMN) as an electrophysiological marker of attentive and predictive processes influenced by different styles and degrees of experience in mindfulness meditation.

The MMN is a stimulus-evoked neurophysiological response elicited by a sudden change in a stream of repeating sensory stimuli. It can be recorded using electroencephalography (EEG) through experimental paradigms, named “oddball task”, consisting of sequences of repeating stimuli (standard) perturbed by infrequent stimuli (deviant) that can differ in terms of frequency, intensity, duration and other features (Näätänen et al., 2007). In general, the MMN is obtained by subtracting the average response to the last standard stimulus before a deviant from the response to the deviant stimulus. In its auditory form, this event-

* Corresponding authors.

E-mail addresses: enrico.fucci@igdore.org (E. Fucci), antoine.lutz@inserm.fr (A. Lutz).

¹ Arnaud Pouban-Couzardot and Oussama Abdoun have contributed equally to the present work.

<https://doi.org/10.1016/j.ijpsycho.2022.03.010>

Received 16 September 2021; Received in revised form 11 March 2022; Accepted 23 March 2022

Available online 28 March 2022

0167-8760/© 2022 Published by Elsevier B.V.

related potential (ERP) component peaks between 100 and 200 ms after stimulus onset (Sams et al., 1985) and is distributed over frontocentral scalp areas, prominently in frontal regions. The auditory MMN has been widely studied in the field of neurophysiology for more than three decades, also due to the ease in recording it using non-invasive techniques in different populations and developmental stages. As a neural marker of sensory and attentive processing, it has been informing neuroscientific models of cognition (Sussman and Shafer, 2014; Näätänen et al., 2007) and has been used as an index of cognitive dysfunction in psychiatric and neurological patients (Näätänen et al., 2014). More recently, its underlying mechanisms and its functional significance have been reframed in the context of the overarching psychobiological theory of predictive coding (Friston and Kiebel, 2009). Within the predictive coding framework, the MMN is described as a marker of precision-weighted prediction error signals involved in implicit perceptual learning processes (Garrido et al., 2009). This account of the MMN has been supported by neuroanatomical and neurocomputational models describing bottom-up and top-down neural pathways conveying predictions and prediction-error signals within a hierarchical structure that encompasses the auditory cortex as well as prefrontal brain regions (Garrido et al., 2008; Wacongne et al., 2012). This has also led to a reevaluation of the mechanisms underlying MMN modulation in psychiatric conditions such as schizophrenia (Rentzsch et al., 2015).

Up to date, few studies reported an effect of meditation on the amplitude of the auditory MMN. Increased MMN was observed during forms of focused attention (FA; Lutz et al., 2007) meditation as a state effect in both expert (Srinivasan and Bajjal, 2007) and novice practitioners (Braboszcz and Delorme, 2011) and as a trait effect in expert practitioners compared to novices (Srinivasan and Bajjal, 2007; Biedermann et al., 2016). In a recent study, Fucci et al. (2018) reported an interaction between meditation experience and styles of practice, showing a larger MMN amplitude in expert practitioners during FA compared to a control condition and novice practitioners. However, no effect on the MMN amplitude was observed in the same expert practitioners during a form of advanced, objectless practice considered to entail a state of nondual awareness (open presence, OP; Lutz et al., 2007).

The results from these studies have been quoted as supporting evidence of increased attentional and perceptual performance by meditation practice (e.g. Josefsson and Broberg, 2011; Ramsburg and Youmans, 2014; Vago and Silbersweig, 2012) and in the context of reviewing how the MMN can be influenced by specific states of consciousness (Näätänen et al., 2012). More recently, the results from Fucci et al. (2018) have been quoted in theoretical works proposing a predictive coding account of meditation practices (Laukkonen and Slagter, 2021; Lutz et al., 2019; Pagnoni and Guareschi, 2017). Within this framework, larger MMN amplitude during FA would reflect the monitoring of task-unrelated events enhancing the brain responses to deviations in the auditory environment. The state of OP, in turn, would entail a reduction in the formation of perceptual predictions and, as a result, a decreased MMN amplitude even when attention is not diverted from the auditory stream.

Previous studies, however, were conducted on small samples of participants, an issue that has a detrimental effect on reproducibility (Button et al., 2013) and that has been argued to represent a recurrent limitation in meditation research (Davidson and Kaszniak, 2015). Moreover, as it is the case for a large part of the literature in neuroscience and psychology, experiments implementing similar paradigms may have not analyzed the MMN amplitude or may have not reported null findings (Ferguson and Brannick, 2012). Finally, previous evidence was obtained using traditional frequentist approaches to statistical analysis, which may cause an overestimation of effect sizes or even errors in effect direction when applied to underpowered studies (Gelman and Hill, 2006).

In the present work, we sought to address these limitations and replicate previous findings showing a differential modulation of the

MMN amplitude by FA and OP meditations in experts compared to a control state and novice practitioners. We did so in the context of a passive auditory oddball task similar to the one implemented in Fucci et al. (2018) and studying an adequately powered sample size of expert and novice meditation practitioners during FA and OP styles of meditation. As in the previous experiment, we expected to observe an interaction between groups and meditation states. Specifically, experts would show a larger MMN amplitude during FA compared to a control state and OP, as well as a larger MMN during FA compared to novices. We tested this hypothesis using both frequentist and Bayesian statistical approaches.

As part of our replication effort, we analyzed EEG oscillatory activity in the alpha (8–12 Hz) frequency band. We expected to observe, in line with Fucci et al. (2018), increased power over this frequency band during meditative states compared to a control condition in both expert and novice practitioners. Increased alpha power during meditation practice has been frequently reported (Ahani et al., 2014; Lagopoulos et al., 2009; Takahashi et al., 2005; Yu et al., 2011), but null or opposite results were also found (Amihai and Kozhevnikov, 2014; Braboszcz et al., 2017; Cahn et al., 2010; Kubota et al., 2001; see Cahn and Polich, 2006; Lee et al., 2018 and Lomas et al., 2015 for exhaustive reviews). In the same way, the presence of a trait effect of meditation experience on alpha power is reported inconsistently depending on meditation traditions and degrees of expertise (Cahn and Polich, 2006). For example, Fucci et al. (2018) did not observe any significant differences between long term meditators and novices. In the current report we explored the effect of meditation expertise on alpha oscillations in a larger sample of participants.

2. Methods and materials

2.1. Subjects

Thirty expert meditation practitioners (52 ± 7.8 years old, 13 females, right-handed, 3 ± 2.62 years of higher education, all white Caucasians) and thirty-six control healthy participants (52 ± 7.6 years old, 17 females, right-handed, 3 ± 2.33 years of higher education, all white Caucasians) participated in the study. Control participants (hereafter referred to as “novices”) were recruited locally through flyers and posters in public places, mailing lists, social media and notifications to research participant databases. Novices underwent a weekend-long introduction to meditation and practiced at home before participating in the experiment (see Section 2.2 for details). Expert practitioners ($>10,000$ h of practice and at least one 3-years intensive retreat) were recruited from multiple meditation centers, predominantly in France but also internationally, with the help of a long-term practitioner with extensive contacts within the Nyingma and Kagyü communities of Tibetan Buddhism. The present study was part of a larger research project that included several experimental sessions. A detailed description of the general inclusion and exclusion criteria for expert and novice practitioners is provided in a separate document (Abdoun et al., 2018; accessible online at osf.io/dbwch). For this specific study, relevant exclusion criteria were: use of psychoactive medications, history of neurological or psychiatric conditions, history of chronic pain or other conditions involving sensitization to pain, personal or family history of epilepsy, severe hearing loss. All participants provided proof of affiliation to social security, signed an informed consent before the beginning of the experiment, and were paid for their participation. Ethical approval was obtained from the appropriate regional ethics committee on Human Research (CPP Sud-Est IV, 2015-A01472-47).

2.2. Meditation practices and training

The distinction between focused attention (FA) and open monitoring (OM) styles of meditation has been extensively discussed in previous theoretical work (Lutz et al., 2008). In the present study, during the FA

practice, participants were asked to focus on a fixation cross as the object of meditation, in line with our previous research (Fucci et al., 2018). The choice of this specific object was done to allow for an unbiased comparison with the control state, when subjects watched a silent documentary on the same screen. During the OM practice, a fixation cross was present on the screen and subjects were asked to keep their eyes open. Practicing meditation with eyes open is a common feature of the Nyingma and Kagyü schools, which all expert practitioners who participated in this research belonged to. It is also suggested in the training program undergone by control participants, which derives from the same traditions of Tibetan Buddhism. In the present study, keeping the eyes open helped avoiding the contamination of the EEG signal by alpha oscillatory activity.

It is important to notice that novice and expert practitioners differ in their style of practicing OM meditation. Specifically, the present group of expert meditators was intensively trained in practices aiming at recognizing and sustaining the “nature of the mind” (Tib. Wylie, *sems nyid* or *sems kyī rang bzhin*) or one’s “fundamental awareness” (Tib. Wylie, *rig pa*). In their case, practicing meditation without object is related to maintaining a state previously referred to as open presence (OP) (e.g. Lutz et al., 2007). Within this state, an awareness of whatever emerges in experience continues without the effortful vigilance that characterizes OM styles of practice (see the notion of “reflexive awareness” in previous works, e.g., Coseru, 2015; Dunne, 2011). However, as we mentioned in previous research (Fucci et al., 2018), the state of OP is characterized by a high degree of variability between subjects and within a specific session. In the present study, experts were recruited based on their capacity to recognize and stabilize the state of OP (following teachers’ recommendations and peer validation facilitated by direct contact with the communities of practice) and were instructed to engage with this specific practice. Novice practitioners were introduced to the practices of FA and OM meditations during a weekend-long training program and were asked to practice at home during the period preceding the experiment (see Abdoun et al., 2019 for a detailed account of subjects’ degree of compliance to the home practice).

2.3. Task design and stimuli

The experiment consisted in a passive auditory oddball task (e.g. Näätänen et al., 2004). In this task, variable sequences of standard tones (880 Hz; 80 ms duration; 10 ms rise and fall) were followed by the presentation of a frequency deviant (988 Hz; 20% of all auditory stimuli). The stimuli were presented binaurally (fixed inter-stimulus interval = 500 ms). To increase the number of observations per subject while keeping recording time reasonably short, the paradigm comprised two sessions. Each session included three experimental conditions (named “states” hereafter): focused attention meditation (FA), open presence (OP) meditation (or open monitoring [OM] meditation for novices) and a control state during which subjects were asked to watch a silent movie (extracts from the documentary “Baraka”, 1992; <https://www.barakasmara.com>) and ignore the auditory stimuli (VD). We informed the participants that this condition was a control state where the mind lets itself be absorbed by the content of the silent movie and is thought not to exhibit spontaneous meditative qualities. We used an active control state which entailed attending a visual object to minimize confounds when comparing it with the two meditation conditions (both involving a fixation cross). This was also in line with conditions commonly used in passive oddball paradigms where participants are instructed not to pay attention to the auditory stimuli. State order was randomized within each session. Each state-block (~6 min long) started with a 2 min period with no auditory stimuli. During this time, subjects were asked to meditate or watch the silent movie, depending on the state instructions. Subsequently, oddball sequences were embedded in alternating 30s periods (8 periods per block) in which participants were informed of the possibility of receiving an electric shock (threat periods, $n = 4$) or that no shocks would be delivered (safe periods, $n = 4$) [as in Cornwell et al.,

2007]. Safe and threat periods consisted of the same number of stimuli ($n = 56$ deviants and $n = 224$ standards when combining the four periods of each state). For the aim of the present research, where we sought to replicate previous findings on the effect of meditation states and expertise on event-related potentials and brain oscillatory activity in an emotionally neutral condition, we conducted statistical analysis excluding threat periods. Results from the analysis of the effect of threat on event-related potentials in novice practitioners are reported in Fucci et al. (2019). At the end of each block, participants answered (on a 1 to 7 Likert-item) questions about their subjective experience during the task. The results of the analysis of self-reports will be described elsewhere.

2.4. EEG recordings

EEG was recorded with the ActiveTwo system (BioSemi, Amsterdam, Netherlands) and sampled at 512 Hz. The recording system consisted of 64 active electrodes positioned on a cap following the standard 10/20 coordinates. The horizontal and vertical EOG was measured by placing flat electrodes on the outer canthi and above and below the subject’s left eye. Electrodes were kept within an offset of 50 mV (+/–25 mV) using the Biosemi ActiView data acquisition system for measuring signal quality. Recordings were performed in a highly shielded Faraday chamber.

2.5. EEG data preprocessing

EEGlab (Delorme and Makeig, 2004) and in-house Matlab scripts (version R2015a) were used for data pre-processing. Raw recordings were downsampled to 250 Hz and re-referenced offline based on the electrodes placed on the mastoids (average activity of the two channels). Bad channels (affected by muscle-related activity over long periods of time) were identified by means of visual inspection and subsequently marked for being interpolated. We applied Independent Component Analysis (ICA) to the continuous data using the Runica algorithm (Makeig et al., 2002), distinctly for the two recording sessions. Data that underwent ICA were manually cleared of large movement-related artifacts following examination of the raw EEG (performed by the first author of the manuscript, who had previous experience in EEG data processing), did not include channels previously marked for interpolation and were filtered between 1 and 20 Hz (using *eegfiltnew* function, which applies a Hamming windowed sinc FIR filter. High-pass and low-pass filters were applied separately. High-pass 1 Hz order = 1650; low-pass 20 Hz order = 166). ICA was applied only on recordings from the control video state (VD). This choice was made to be able to clearly identify saccades and blinks across all experimental states while allowing for faster calculations. Resulting ICA matrices were then applied to the original raw data of each experimental state within a single session. Subsequently, ICA components reflecting the activity caused by blinks and saccades were rejected. Despite the 1 Hz initial filter, slow frequency artifacts (“drifts”), presumably due to increased sweating caused by the threat-induction procedure, were affecting the final ERP waveforms. As an additional step, we applied a high-pass 2 Hz filter (2 Hz, filter order = 414) on the ICA corrected data, successfully removing those artifacts. Bad channels were interpolated and the *Cleanline* algorithm was applied to remove 50 Hz noise. Finally, epochs were generated for standard and deviant stimuli (–200 to 500 ms after stimulus onset) and baseline-corrected (–100 ms baseline). Epochs that included artifacts were removed by means of visual inspection. Visual inspection was complemented by an automatic rejection of those epochs that included data points exceeding a $-/+70$ μ V amplitude threshold. Epochs that included artifacts were removed if the number of deviants in the safe or threat condition was lower than thirty-five after rejection of bad epochs. Based on this criterion, all data from one expert practitioner and data from specific experimental conditions (FA, OP, VD) and blocks were excluded from further analysis (see Supplementary Information 1.1 for a summary table and Supplementary Information 1.2 for a summary of the mean

number of epochs in each session separately for group, state and stimulus type. Available online at osf.io/3f9e7). Outer ring channels were rejected for all participants due to occasional high-frequency noise caused by muscle-related artifacts identified in some subjects, (see Fig. S1 for a visual layout of the remaining 41 channels. Available online at osf.io/3f9e7). Finally, a low-pass filter of 20 Hz was applied to the epoched data for the analysis of the evoked responses.

2.5.1. Mismatch negativity

To identify the MMN, we computed the difference waveforms from the grand average between all subjects, separately for each group, state (FA, OP, VD) and session. We conducted the analysis using an a priori region of interest (ROI) which included the electrode Fz and four surrounding electrodes (Fig. S1). This ROI was in line with our previous research (Fucci et al., 2018) and with the literature on auditory MMN (e.g., Duncan et al., 2009). The MMN amplitude was computed based on a 20 ms time-window centered around the most negative peak, between 90 and 200 ms after stimulus onset, of the difference waveform for each session, state and group (in line with previous studies, e.g., Näätänen et al., 2004).

2.5.2. Alpha spectral power

For spectral analysis, we limited the analysis to two spatial regions of interest (ROI), frontal and occipital, defined similarly to Fucci et al. (2018) (see Fig. S1 for a description of channels included in the ROIs, available online at osf.io/3f9e7). Cleaned epoched data were imported into MNE-Python (Gramfort et al., 2013) and power spectral density (PSD) was computed for each state (FA, OP and VD) and epoch with the multitapers method (Percival et al., 1993), using default parameters of the `psd_multitaper` function from MNE-Python (version 0.22.1). PSDs were then averaged across all channels of a ROI and alpha absolute power was computed as the integral approximation of the spectrum between 8 and 12 Hz using Simpson's rule (`simps` function from `scipy` 1.4.1, Python 3.8). Finally, alpha power was averaged for every subject across all epochs (deviants and standards), states and sessions, and log-transformed before statistical analyses.

2.6. Statistical analysis

R Studio (version 3.4.2 [R core team, 2017]) was used for all statistical analyses. R code used in the analysis can be found at osf.io/3f9e7.

2.6.1. Statistical power

The present study was part of the ERC-funded, Brain and Mindfulness project (Abdoun et al., 2018). In the context of this larger multimodal, neurocognitive project, multiple factors determined the final sample sizes, independently of the goal of the current research (i.e., the estimation of meditation effects on the MMN). Nonetheless, our sample size is larger than any other used in previous, similar studies, including Fucci et al. (2018). We evaluated the statistical power of our current sample through simulations performed using the R library `simr` (v1.0.5; Green and MacLeod, 2016). Simulations were based on all model estimates (fixed and random effects) from Fucci et al. (2018) study. The variance of residuals was scaled up to take into account the fact that less trials were available to calculate the average ERP (112 deviant stimuli per state in the current study vs. 150 in Fucci et al., 2018). Two different values were explored for the effect size of interest, namely the group by state interaction: the original effect size from the 2018 study, as well as half of it. Using 10,000 simulations for each scenario, we found that our replication analysis had 89% power to detect the original effect size, but only 33% to detect half of it.

2.6.2. Linear mixed models

We used linear mixed-effects models (R package `lme4`, Bates et al., 2014) to test the effect of group and state on the MMN amplitude and on

the alpha log power separately (for alpha log power the analysis was performed independently for parietal and occipital regions of interest). For every model, group, state and their interaction were included as fixed effects and subject was included as random intercept. All models included unaveraged data from the two experimental sessions. The main effects and interactions of the model were estimated using a type III F-test with Satterthwaite's method for approximation of degrees of freedom. Paired *t*-tests, corrected for multiple comparisons using Tukey honestly significant difference test (HSD), were used as post-hoc tests comparing estimated marginal means using the R package `emmeans` (Lenth et al., 2020). In the results section, we report estimates of pseudo- R^2 as proposed by Nakagawa and Schielzeth (2013) for the full models and unstandardized effects sizes (pairwise differences in estimated marginal means and their associated confidence intervals) for pairwise comparisons. Notice that the computation of effect sizes in linear mixed models is not unambiguous and should be handled with consideration.

2.6.3. Bayesian statistics

2.6.3.1. Model structure. The MMN amplitude was reanalyzed in the Bayesian framework using the R package `rstanarm` (v2.21.1; Goodrich et al., 2020). Data from the original experiment (Fucci et al., 2018) and the present study were modeled together within a single statistical model, the same used for frequentist analysis (see section above), with the minor addition of a fixed main effect of Experiment to account for differences in recorded MMN amplitude due to differences in stimuli properties and recording devices. Interestingly, the Bayes factors obtained from a model combining two experiments is equal to the product of the Bayes factor obtained from the first experiment, by the Bayes factor obtained from the second experiment where the posterior distributions of the first experiment are used as priors of the replication (Ly et al., 2019):

$$BF(exp_1, exp_2) = BF(exp_1) \times BF(exp_2|exp_1)$$

2.6.3.2. Prior distributions. We followed the principled Bayesian workflow proposed by Schad et al. (2021) to define priors that are compatible with domain expertise (see Supplementary Information 2 at osf.io/3f9e7). Prior predictive checks revealed that the default, weakly informative priors of `rstanarm` were unrealistic as they were compatible with highly implausible datasets. Therefore, we used an alternative set of narrower priors that remain broad enough to reflect very general and uncontroversial elements of domain expertise regarding the MMN component, i.e., knowledge about the likely characteristics of the MMN component obtained from previous research such as plausible ranges of amplitude and effect sizes. Notably, all priors for the regression coefficients were normally distributed, centered on 0 (except the prior for the intercept, which was centered on the mean of the data). Standard deviations of normally distributed priors (regression coefficients) and rates of exponentially distributed priors (random intercept and residual variance) were conveniently set to 2 and 1 μ V, respectively.

2.6.3.3. Statistical inference. 10,000 samples were drawn from every posterior distribution. Contrast samples were extracted using the R package `emmeans` (Lenth et al., 2020). Bayes factors and support intervals were calculated using the R package `bayestestR` (v0.11.0 Makowski et al., 2019).

3. Results

3.1. MMN amplitude

3.1.1. Frequentist analysis

Here we sought to replicate previous findings on the different modulation of the MMN amplitude by meditation states and expertise. Fig. 1. A shows the time-course of the mean difference waveforms, across

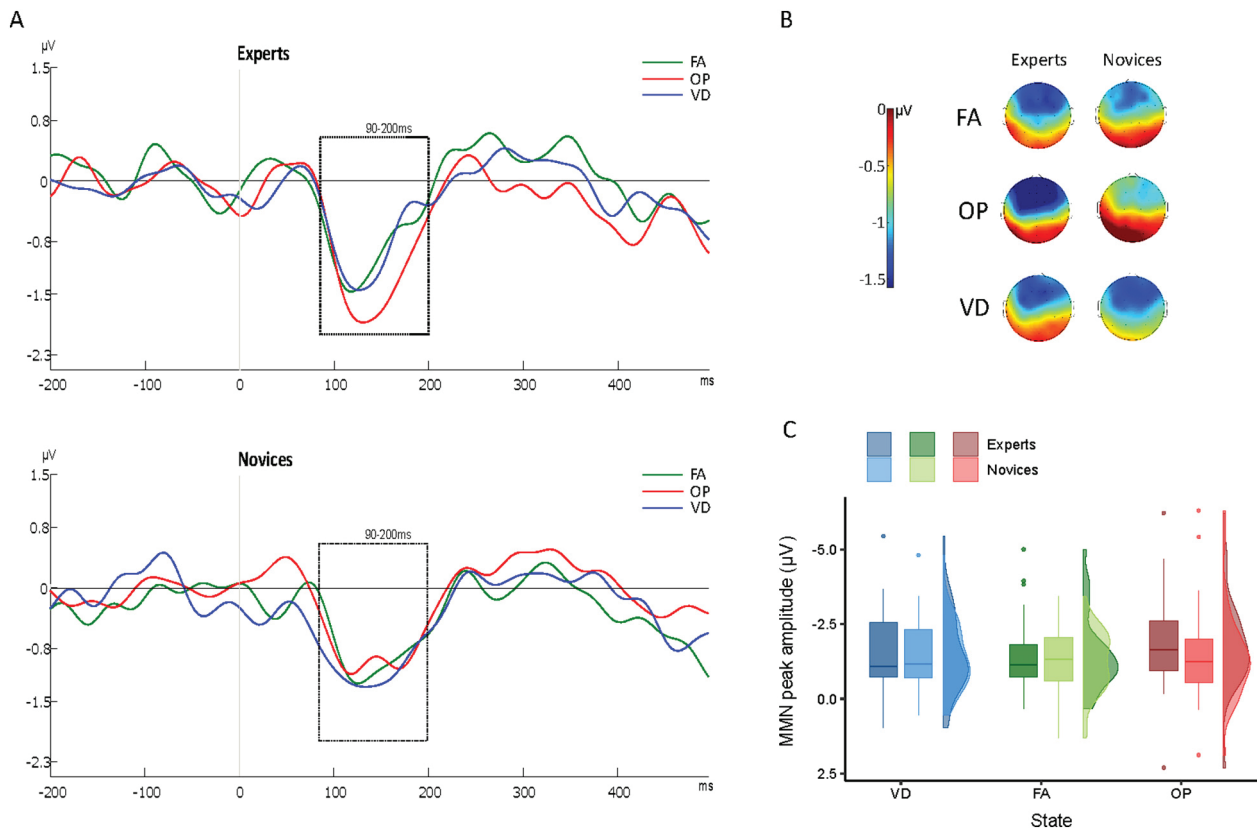


Fig. 1. Meditation states and degrees of experience do not modulate the amplitude of auditory mismatch negativity (MMN). (A) Subtraction (deviant minus standard, i.e., MMN) waveforms at frontal ROI (see Fig. S1) for FA, OP and VD states (FA: focused attention meditation, OP: open presence meditation, VD: watching a silent movie) in expert (top) and novice (bottom) practitioners. (B) Average voltage scalp maps of MMN in a 20 ms window centered around the peak amplitude [between 90 and 200 ms after stimulus onset] for experts (top) and novices (bottom) during FA, OP and VD. (C) Box plots and half-violin plots displaying median values and distribution of the MMN peak amplitude averaged over sessions across subjects at frontal ROI.

participants and sessions during FA, OP and VD, separately for expert and novice participants. Fig. 1.B shows the respective scalp topographies of the mean MMN peak amplitude for the above-mentioned states and groups. We did not find a significant interaction between group and state on the MMN amplitude ($F(2) = 0.51$; $p = 0.6$) [Fig. 3.C; Pseudo- R^2 (full model) = 0.08, Pseudo- R^2 (fixed effects only) = 0.01]. A main effect of state or group was not present either ($F(2) = 0.57$; $p = 0.57$ and $F(1) = 1.95$; $p = 0.17$ for state and group, respectively). Additionally, we tested an alternative model at the frontal ROI which included MMN amplitude measures derived from the mean value in a fixed time-window of 90–180 ms after stimulus onset (in line with Fucci et al., 2018). Both models did not highlight any significant interaction or main effect of state and level of expertise.

When adding session information as a factor, the main model yielded a statistically significant main effect of session ($F(2) = 4.68$; $p = 0.03$; higher MMN amplitude during the second session). However, adding this factor did not influence the other results.

3.1.2. Bayesian analysis

Traditional frequentist statistical tests yield limited information and are inconclusive by nature when they fail to reject the null hypothesis. In contrast, Bayesian factors provide a measure of relative likelihood of two competing hypotheses and may be used to assess the evidence in favor or against the null or a theory's predictions (Dienes, 2014). Moreover, calculating Bayesian factors on aggregated data naturally integrates the evidence provided by an experiment and its replication, where the inference of the replication builds on the conclusions of the original study (Ly et al., 2019). We have used this property to assess the accumulated evidence of our original study (Fucci et al., 2018) and the

current replication, integrating the two studies in a single statistical model fitted and evaluated in the Bayesian framework. We focused on five effects of particular interest: i) the difference in MMN amplitude between states in experts compared to novices, which was the main finding of the original study; ii) the increase of MMN during FA as an effect of expertise, which is the most commonly reported result in the literature on meditation and MMN; the somewhat less studied iii) state effect of FA meditation (compared to a control condition) on MMN, here evaluated across all participants. Additionally, and in a more exploratory fashion, we tested iv) the effect of expertise on the MMN during the practice of OM/OP, as well as v) the main effect of OM/OP meditation across all participants. The posterior distributions of the five tested contrasts are represented in Fig. 2 (top plots) along with the mean, 50% and 95% confidence intervals. For each contrast, hypothesis testing was carried out using Bayesian factors in two different ways. First, and following a common practice, we evaluated the evidence *in favor* (or *against*) the null hypothesis BF_{01} , compared to the alternative hypothesis that there is an effect of unspecified sign. This measure is particularly useful to reject the null hypothesis when it is true, but it is poorly informative when the evidence in favor of the null is not overwhelming (or even against it) as it says nothing about the direction and magnitude of plausible effect sizes. Therefore, we also report the range of effect sizes that received strong ($BF > 10$) evidence from the data, alias the *support interval* (Wagenmakers et al., 2020). Fig. 2 summarizes the results, with the plots on the bottom row displaying the BF for the null hypothesis (i.e., BF_{01}) along with the $BF > 10$ and $BF > 3$ support intervals.

For the interaction effect between levels of expertise and meditation states, we found strong evidence in favor of the null ($BF_{01} = 11.0$)

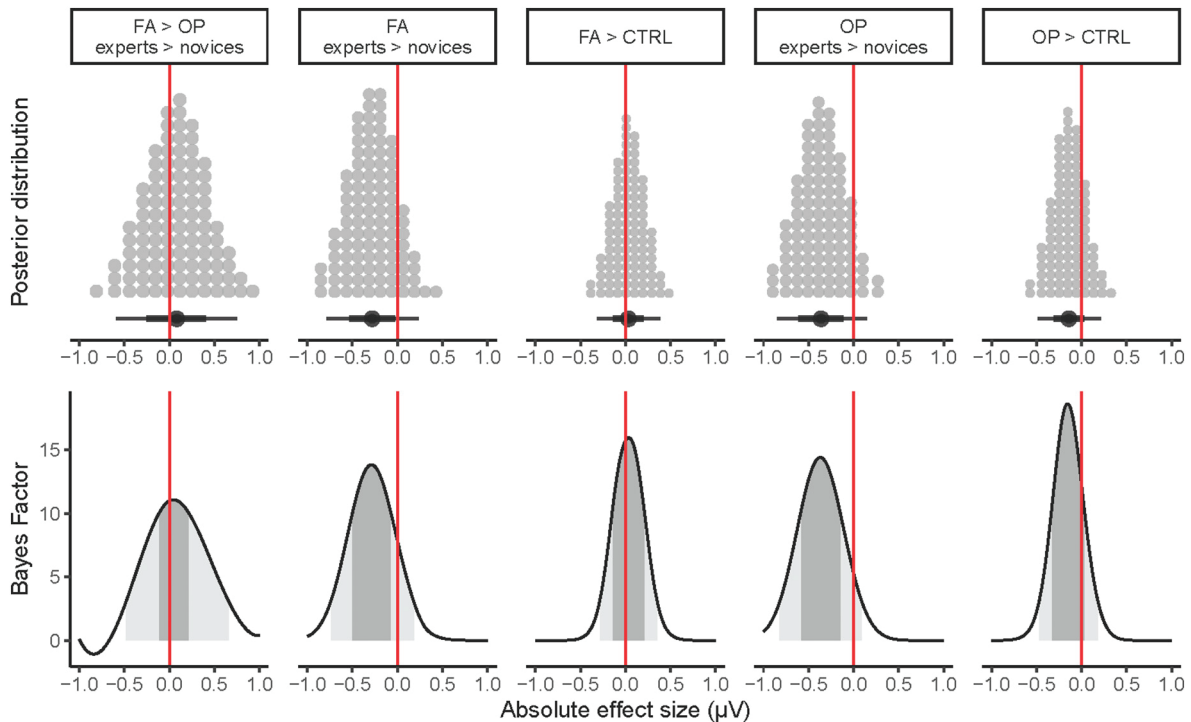


Fig. 2. Bayesian analysis of aggregated data provides evidence for null effects of meditation states and expertise on the MMN. Data from Fucci et al., 2018 and the current study were fitted in a single model to assess the cumulative evidence obtained from the original study and its replication. Labels at the top indicate the contrasts that have been tested, with “A > B” meaning “A minus B”. (Top) Posterior distributions of contrasts of interest. Thick and thin lines represent 50% and 95% highest density intervals respectively. (Bottom) Bayes factors of point hypotheses. For each value δ , BF_{δ} is the Bayes factor of the hypothesis that the absolute effect size of the contrast (i.e., its amplitude in μV) is equal to δ vs. the alternative, complementary hypothesis that it is different from δ . Note that BF_{01} , the Bayesian factor of the null hypothesis, is the special case of $\delta = 0$. Light and dark shaded areas highlight the support interval for $BF > 3$ and $BF > 10$, respectively, i.e., values of the contrast estimate that receive substantial and strong evidence, respectively, from the data when pitted against the alternative, complementary hypothesis. The null hypothesis falls within the substantial support interval for expertise effects during FA and OM (second and fourth plots, respectively) and within the strong support interval for all other tested contrasts.

compared to the alternative hypothesis of a differential effect of meditation states on MMN between novice and long-term practitioners. Accordingly, the range of absolute effect sizes supported by strong evidence ($BF > 10$) extended in both directions around 0, from -0.05 to $0.25 \mu\text{V}$.

For the expertise effect in FA, we found moderate evidence in favor of the null ($BF_{01} = 7.9$) compared to the alternative hypothesis of a difference in MMN between novice and long-term practitioners. Absolute effect sizes supported by strong evidence ($BF > 10$) ranged from -0.48 to $-0.08 \mu\text{V}$, corresponding to Cohen's d from 0.06 to 0.33 for a larger MMN amplitude in experts than novices.

For the state effect of FA compared to the control condition across all participants, the combined data provides strong evidence in favor of the null ($BF_{01} = 16.1$). Absolute effect sizes supported by strong evidence ($BF > 10$) ranged from -0.15 to $0.21 \mu\text{V}$, an interval that is roughly centered on 0.

For exploratory purposes, we extended main effects analyses to the practice of OM/OP. We found moderate evidence in favor of a null effect of expertise ($BF_{01} = 5.3$; support interval = $[-0.55 \mu\text{V}, -0.16 \mu\text{V}]$) corresponding to Cohen's d from 0.09 to 0.31 for a larger MMN amplitude in experts than novices), and stronger evidence in favor of a null effect of state ($BF_{01} = 11.9$; support interval = $[-0.31 \mu\text{V}, 0.02 \mu\text{V}]$).

Taken altogether, the results from the Bayesian reanalysis of 96 participants suggest a) that meditation expertise does not interact with meditation states with respect to the MMN amplitude; b) that extended lifetime meditation experience is unlikely to be associated to enhanced MMN component during meditation states and, if it does, it is to a small extent at best; and c) that state effects of meditation on the magnitude of the MMN, if they exist, are even smaller than expertise effects.

3.2. Alpha spectral power

Investigation of spectral power differences in the alpha frequency band (8–12 Hz) was carried out across expertise groups and meditative states, for two spatial ROIs: frontal and occipital (Fig. S1). Results were similar in both ROIs (Fig. 3): there was no interaction between group and state [frontal ROI: $F(2) = 2.6$, $p = 0.076$, pseudo- R^2 (full model) = 0.84 (fixed effects only = 0.05); occipital ROI: $F(2) = 2.68$, $p = 0.07$, pseudo- R^2 (full model) = 0.86 (fixed effects only = 0.04)], and no main effect of group [frontal ROI: $F(1) = 1.1$, $p = 0.30$; occipital ROI: $F(1) = 0.63$, $p = 0.43$] but a main effect of state [frontal ROI: $F(2) = 35.9$, $p < 0.001$; occipital ROI: $F(2) = 44.4$, $p < 0.001$]. Post-hoc tests revealed higher alpha power during focused attention (FA) and open presence (OP) meditations compared to the video control state (VD) (frontal ROI: VD-FA t -ratio(306) = -5.8 , $p < 0.001$, estimate = -0.21 [-0.29 , -0.12], VD-OP t -ratio(307) = -8.2 , $p < 0.001$, estimate = -0.3 [-0.38 , -0.21]; occipital ROI: VD-FA t -ratio(306) = -6.7 , $p < 0.001$, estimate = -0.29 [-0.39 , -0.19], VD-OP t -ratio(306) = -9.1 , $p < 0.001$, estimate = -0.4 [-0.5 , -0.29]) and also during OP compared to FA meditation (FA-OP, frontal ROI: t -ratio(306) = -2.47 , $p = 0.037$, estimate = -0.089 [-0.17 , 0]; occipital ROI: t -ratio(306) = -2.40 , $p = 0.040$, estimate = -0.1 [-0.21 , 0]). Notice that the higher alpha power observed during meditation states compared to the control condition could be due to differences in visual stimulation (i.e., a video stimulus versus a fixation cross). This issue is discussed in Section 4.4.

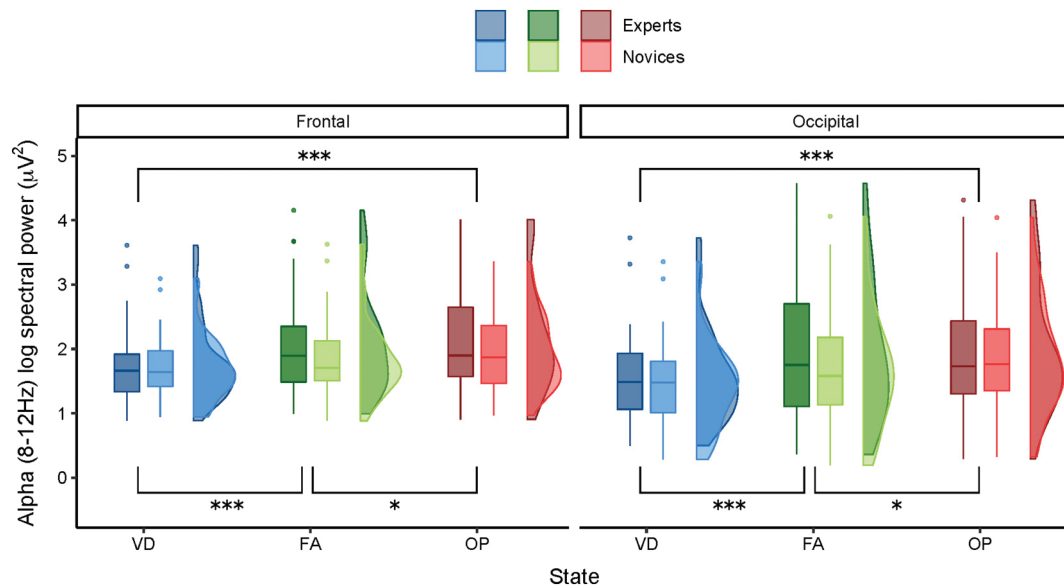


Fig. 3. Increased alpha power in two different meditation practices compared to a control state. Box plots and half-violin plots displaying median values and distributions of alpha (8–12 Hz) log spectral power averaged over sessions at frontal (left) and occipital (right) ROIs for FA, OP and VD states in expert (dark shade) and novice (light shade) practitioners. Statistical analyses revealed significant main effects of state in both ROIs regardless of expertise, with FA and OP showing higher alpha than VD (*: $p < 0.05$, ***: $p < 0.001$, p-values from post-hoc tests of linear mixed models).

4. Discussion

4.1. Frequentist analysis: non replication of meditation states and expertise effects on MMN amplitude

In the present study, we sought to replicate previous findings on the differential modulation of the auditory MMN amplitude by meditation states in novice and expert practitioners, as observed in Fucci et al. (2018). The replication attempt was conducted in the context of a similar, although not identical, experimental paradigm, using the same frequentist models for statistical analysis and with a high enough statistical power to detect the effects reported in Fucci et al. (2018). As a result, we did not find any statistically significant interaction effect between meditation states and groups, nor differences between groups or states taken independently (Fig. 1). This result was unexpected because previous research has shown an impact of meditation on the auditory MMN amplitude. Different studies reported increased MMN amplitude during FA meditation as a state effect in experts or novice practitioners (Braboszcz and Delorme, 2011; Srinivasan and Bajjal, 2007) or as a trait effect in experts compared to novices (Biedermann et al., 2016; Srinivasan and Bajjal, 2007), while one study reported increase MMN amplitude during FA in experts, compared to novices, but no difference between a control condition and a state of nondual, open presence (OP) meditation (Fucci et al., 2018). The non-replication of previously reported results could be related to factors undermining research replicability in the field of contemplative science and in the broader fields of cognitive neuroscience and experimental psychology. First, previous studies were conducted on sample sizes that are likely inadequate to detect reasonable effect sizes (Srinivasan and Bajjal, 2007 $n = 10$ [experts] and $n = 10$ [novices]; Braboszcz and Delorme, 2011 $n = 12$; Biedermann et al., 2016 $n = 12$ [experts] and $n = 14$ [novices]; Fucci et al., 2018 $n = 16$ [experts] and $n = 15$ [novices]). Low statistical power is a well-known issue affecting neuroscientific studies (Button et al., 2013). It can not only reduce the chance of detecting moderate and low effects, but also increase the likelihood that statistically significant differences overestimate (Gelman and Carlin, 2014) or do not represent true effects (in this case affecting the positive predictive value [PPV], which is the probability that a “positive” finding represents a true effect; Ioannidis, 2005). These issues are reinforced by the selective

reporting of positive findings (so-called “publication bias”). Secondly, some of the previous studies implemented analysis approaches that deviates from standard practices without a strong methodological justification. For instance, Biedermann et al. (2016) reported increased MMN amplitude in experts, compared to novice practitioners at Pz electrode. This is an unconventional scalp area for the analysis of the auditory MMN amplitude, which normally peaks at frontocentral scalp regions (Näätänen et al., 2012). In Srinivasan and Bajjal (2007), increased MMN amplitude during a form of FA meditation in expert practitioners was reported following a post-hoc analysis carried out specifically on this group. However, since no interaction between groups and meditation states was present in the main model tested, the validity of the post-hoc analysis is questionable. More generally, the high degree of flexibility characterizing analysis in neuroscientific studies is prone to increase the chance of false positives (Gelman and Loken, 2013; Pol-drack et al., 2017).

Together with small sample sizes, these issues could have increased the prevalence of Type I errors in the research on MMN modulation by meditation states and traits and could represent an explanation for the discrepancy between the null findings reported in the present study and the previous literature. The participants sample studied here was at least double the size of those used in previous research and allowed for the detection of an effect as large as the one reported in Fucci et al. (2018) with a 89% power. Overall, the present study represents, in our opinion, the most reliable analysis of the modulation of auditory MMN by meditation states and traits to date.

This is the first report of null findings on this subject. However, it could be the case that other studies did not report null results on the influence of meditation on the MMN amplitude or did not analyze the MMN component in auditory oddball paradigms. While trying to make sense of the differences between our null findings and the existing literature, we identified at least seven studies that implemented similar experimental designs (i.e. passive or active auditory oddball tasks with tones presented binaurally), groups of participants and meditation practices (Atchley et al., 2016; Barnes et al., 2018; Cahn and Polich, 2009; Delgado-Pastor et al., 2013; Isbel et al., 2019; Klee et al., 2020; Yoshida et al., 2020). They reported results of analyses on auditory event-related EEG components, but not on the MMN amplitude even though this was feasible. We contacted the first and corresponding

authors of these studies to ask whether they had the opportunity to analyze the MMN component. Up to date, only the authors of [Cahn and Polich \(2009\)](#) replied, saying that no difference in MMN amplitude was found between FA meditation and a control state in expert practitioners in their study. This adds one element to the list of null findings on the subject of MMN modulation by meditation states and traits. Formally estimating whether the literature on this topic is significantly affected by selective reporting was beyond the scope of the present study, but there are reasons to believe that it could be the case. Publication bias is highly prevalent in psychology and social sciences ([Fanelli, 2010](#)) and it has been recently identified in the context of mindfulness-based interventions research ([Coronado-Montoya et al., 2016](#)).

4.2. Bayesian analysis of accumulated evidence: assessing the likelihood of meditation state and trait effects on the auditory MMN

Frequentist analyses have failed to reject any null hypothesis pertaining to the effect of meditation states or expertise on the amplitude of the MMN, prompting the use of Bayesian measures of evidence to quantify the support for such null hypotheses. We focused in particular on the interaction between meditation expertise and meditation states, as well as on simple effects of expertise during meditation states, and main effects of meditation states across all participants.

First, we found that the effect of expertise during FA meditation, which has been reported multiple times in the literature ([Srinivasan and Bajjal, 2007](#); [Biedermann et al., 2016](#); [Fucci et al., 2018](#)), was in the expected direction - i.e., larger MMN amplitude for expert meditators - but compatible with null effects and moderate at best. A similar conclusion could be drawn for OM/OP meditation. Second, the aggregated data provided strong evidence in favor of an absence of meditation (both FA and OM/OP) state effects across participants and levels of expertise, as well as an absence of interaction between meditation expertise and meditation states with respect to the MMN. The same conclusions were reached using Bayes Factors and hypothesis testing based on posterior distributions. Overall, these findings disqualify the auditory MMN as a marker dissociating meditation states from the FA and OM families, or even meditation states from baseline conditions. Our data remain compatible with the theory that meditation practices impact perceptual learning, especially in experienced practitioners, but they suggest that the size of effect is too modest for the MMN amplitude to be a practically useful index of meditation state and/or expertise. Future research on this topic may benefit from the development of more sensitive biomarkers, potentially driven by theoretical considerations on the phenomenology of meditation states.

4.3. Modulation of alpha spectral power by meditative states and expertise

Consistently with previous literature ([Cahn and Polich, 2006](#); [Lomas et al., 2015](#); [Lee et al., 2018](#)), we observed an increase in the power of alpha (8–12 Hz) EEG oscillatory activity during focused attention (FA) and open presence (OP) meditations, compared to a control state, in both frontal and occipital scalp regions and for both expert and novice practitioners. This partially replicates findings on oscillatory power from [Fucci et al. \(2018\)](#), with the addition of a significant difference between FA and OP in both frontal and occipital ROIs that was only present as a trend in the previous study. This last result should be taken cautiously due to its relatively small effect size compared to the larger difference between meditation and control conditions.

Even when increasing the sample size compared to our previous study ([Fucci et al., 2018](#)), we did not find differences in alpha power between groups (experts vs novices). This null finding is in contrast to previous research where trait effects were observed (see [Cahn and Polich, 2006](#) for a detailed review). As pointed out in a recent opinion paper ([Schoenberg et al., 2018](#)), one must be careful when interpreting results from a single measure of frequency band power to avoid

oversimplifying the multi-dimensionality of meditative states that are known to be phenomenologically very complex ([Lutz et al., 2015](#)). Increased alpha power has notably been linked to breath-control ([Zaccaro et al., 2018](#)) and sustained and selective attention processes in the visual ([Clayton et al., 2015](#); [Foxe and Snyder, 2011](#); [Klimesch, 2012](#)) but also auditory modality ([Strauß et al., 2014](#); [Weisz and Obleser, 2014](#)). While meditation practices are likely to affect these processes ([Gerritsen and Band, 2018](#); [Kerr et al., 2013](#)), most paradigms and analyses from the literature, including ours, do not allow to clearly disentangle them. Overall, due to the lack of specificity and unclear functional explanations concerning the engagement of alpha power in meditation, we would suggest not to rely on this marker alone to characterize meditative states and traits; a warning already raised by Cahn and colleagues ([Cahn et al., 2010](#)). Thus, we would like to emphasize that more advanced analyses involving compositional markers are needed to fully depict the diversity of attentional and perceptual processes at stake during contemplative practices and how they are shaped by long-term training. Examples of novel analyses of brain oscillatory activity during meditation can be found in the recent literature ([Rodriguez-Larios et al., 2020](#); [Schoenberg et al., 2018](#); [van Lutterveld et al., 2017](#)).

4.4. Limitations

The present study is affected by some limitations that might deserve further investigation. First, the non-replication of our previous findings might have been influenced by differences in the experimental design between the present and our previous study: [Fucci et al. \(2018\)](#) used a random inter-stimulus interval (I.S.I.) normally distributed between 800 and 1200 ms, while we used a shorter and fixed one of 500 ms (consistently with previous literature in the field of auditory MMN, e.g. [Näätänen et al., 2004](#)). Additionally, in our previous study novices received only a meditation instruction before the task while in the present study they underwent an introductory practice weekend and practiced daily before the experiment (see [Abdoun et al., 2018](#)). Enhancing the meditative skill of novices could have decreased our ability to detect a group by state interaction but would have increased the study's potential for detecting a state effect. Moreover, participants underwent short oddball sequences embedded in safe or threat periods. Previous studies on MMN modulation by meditation states and traits did not explore this marker in the context of mood induction. Here, we excluded data acquired during periods of threat, which were separated from neutral periods by a pause of several seconds. In a separate report ([Fucci et al., 2019](#)), we describe subjective and electrophysiological responses to the different periods in novice practitioners during a control state, showing how they can be independently characterized and how threat periods do not seem to influence the MMN amplitude. However, it is possible that emotional contexts masked effects that would otherwise be present in a neutral situation through a lingering effect.

Secondly, our studies made use of the classical, simple oddball task and our conclusions are restricted to that specific class of experimental designs. In particular, our data and analyses do not exclude the possibility that meditation alters perceptual or statistical learning in more elaborate settings.

Thirdly, the critical literature review included in the discussion section of the present study cannot be considered as reliable as a meta-analysis. No formal approaches have been implemented to determine the prevalence of a publication bias in the literature. As such, the explanations for the non-replication proposed in the discussion section should be considered as opinions informed by an empirical approach rather than statements derived from formal analyses.

Concerning the results on alpha oscillations, as suggested by one reviewer, we cannot entirely exclude that the modulation of alpha power between experimental conditions was partly due to differences in visual stimuli (silent movie for VD and fixation cross for FA/OP). Nonetheless, the results of the present study are supported by the

difference found between meditation styles involving the same visual stimuli. We also acknowledge that the current design and analysis, replicating Fucci et al. (2018), do not allow us to disentangle the effects of meditation from the effects of the oddball task and their interaction on alpha power. Finally, a reviewer pointed out that differences in alpha power, interpreted as change in the periodic component of oscillatory activity, could also be due to a confounding difference in aperiodic components (1/f like, for example offset and slope). We visually inspected the spectral power density distributions to make sure that the differences in alpha power were not conflated by such aperiodic activity (see Supplementary Fig. S3). However, more rigorous methods exist to differentiate periodic and aperiodic components (Donoghue et al., 2020) and they will be worth using in future analyses.

4.5. Future directions

In conclusion, while the MMN is a robust ERP component and reliable neural correlate of attentional processes and perceptual inference in the sensory domain, we warn about its use to back up models of meditation influence on attention and perceptual inference until there is more clarity on the effects of expertise and practices on this marker. Future studies on this subject should design adequately powered experiments and implement rigorous and standardized methods of analysis (or justify alternative methods with valid arguments). To support this, we encourage pre-registrations of both new experiments and future replication attempts (Nosek et al., 2019). Replications of existing studies and reporting of null findings should be a priority for neuroimaging investigations of contemplative practices and recent publications suggest that this field might not be unaffected by the general replicability crisis in neuroscience and psychology research (Barnes et al., 2018; Payne et al., 2019). Additionally, future studies could implement recent analysis approaches in the Bayesian framework to determine the validity of initial hypotheses and produce estimates of plausible effect sizes for a given effect of interest. Future exploratory research could also benefit from electrode-wise cluster-based methods of analysis to identify possible effects in unconventional scalp regions and time windows, as well as from recent developments in EEG data processing that address limitations in the use of baseline correction and can determine how much this step contributes to observed differences in post-stimulus ERPs (Alday, 2019).

Finally, refining the phenomenological granularity of meditation instructions, for instance comparing FA on the auditory sounds to FA on a visual distractor (or to the widely practiced FA on bodily sensations) during auditory oddball paradigms could also deepen our understanding of these processes.

Data availability statement

Single-subject preprocessed EEG data (MMN amplitude and alpha power) used for statistical analysis are available, along with analysis code, at a repository of the Open Science Framework (osf.io/3f9e7). The repository also contains single-subject MMN amplitude data from Fucci et al. (2018), which was used when conducting analysis in the Bayesian framework. The manuscript contains detailed information that should allow a replication of the findings. If you would like to obtain raw EEG data, or need further information, please contact the corresponding author. The experiment was not preregistered.

CRediT authorship contribution statement

EF and AL designed the experiment. EF collected data. EF, OA and APC analyzed data. EF, OA, APC and AL wrote the manuscript.

Acknowledgments

This study was supported by a European Research Council grant

ERC-Consolidator 617739-BRAINandMINDFULNESS to Antoine Lutz. The authors have no conflict of interests. We thank the numerous students who have been involved in the data acquisition and data analysis processes during their internships. In particular Kelly Assouly, Alix Arnas, Arthur Sevennec, Maximilien Metaireau and Gansheng Tan. We thank an anonymous reviewer for numerous detailed and constructive feedback.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ijpsycho.2022.03.010>.

References

- Abdoun, O., Zorn, J., Fucci, E., Perraud, E., Aarts, K., Lutz, A., 2018. Brain & mindfulness project manual. OSF. <https://osf.io/dbwch/>.
- Abdoun, O., Zorn, J., Poletti, S., Fucci, E., Lutz, A., 2019. Training novice practitioners to reliably report their meditation experience using shared phenomenological dimensions. *Conscious. Cogn.* 68, 57–72. <https://doi.org/10.1016/j.concog.2019.01.004>.
- Ahani, A., Wahbeh, H., Nezamfar, H., Miller, M., Erdogmus, D., Oken, B., 2014. Quantitative change of EEG and respiration signals during mindfulness meditation. *J. NeuroEng. Rehabil.* 11 (1), 87. <https://doi.org/10.1186/1743-0003-11-87>.
- Alday, P.M., 2019. How much baseline correction do we need in ERP research? Extended GLM model can replace baseline correction while lifting its limits. *Psychophysiology* 56 (12), e13451. <https://doi.org/10.1111/psyp.13451>.
- Amihai, I., Kozhevnikov, M., 2014. Arousal vs. Relaxation: a comparison of the neurophysiological and cognitive correlates of vajrayana and theravada meditative practices. *PLOS ONE* 9 (7), e102990. <https://doi.org/10.1371/journal.pone.0102990>.
- Atchley, R., Klee, D., Memmott, T., Goodrich, E., Wahbeh, H., Oken, B., 2016. Event-related potential correlates of mindfulness meditation competence. *Neuroscience* 320, 83–92. <https://doi.org/10.1016/j.neuroscience.2016.01.051>.
- Barnes, L.J., McArthur, G.M., Biedermann, B.A., de Lissa, P., Polito, V., Badcock, N.A., 2018. No meditation-related changes in the auditory N1 during first-time meditation. *Int. J. Psychophysiol.* 127, 26–37. <https://doi.org/10.1016/j.ijpsycho.2018.03.003>.
- Bates, D., Mächler, M., Bolker, B., Walker, S., 2014. Fitting Linear Mixed-effects Models Using lme4. *ArXiv Preprint ArXiv:1406.5823*.
- Biedermann, B., de Lissa, P., Mahajan, Y., Polito, V., Badcock, N., Connors, M.H., Quinto, L., Larsen, L., McArthur, G., 2016. Meditation and auditory attention: an ERP study of meditators and non-meditators. *Int. J. Psychophysiol.* 109, 63–70. <https://doi.org/10.1016/j.ijpsycho.2016.09.016>.
- Braboszcz, C., Delorme, A., 2011. Lost in thoughts: neural markers of low alertness during mind wandering. *NeuroImage* 54 (4), 3040–3047. <https://doi.org/10.1016/j.neuroimage.2010.10.008>.
- Braboszcz, C., Cahn, B.R., Levy, J., Fernandez, M., Delorme, A., 2017. Increased gamma brainwave amplitude compared to control in three different meditation traditions. *PLOS ONE* 12 (1), e0170647. <https://doi.org/10.1371/journal.pone.0170647>.
- Button, K.S., Ioannidis, J.P.A., Mokrysz, C., Nosek, B.A., Flint, J., Robinson, E.S.J., Munafò, M.R., 2013. Power failure: why small sample size undermines the reliability of neuroscience. *Nat. Rev. Neurosci.* 14 (5), 365–376. <https://doi.org/10.1038/nrn3475>.
- Cahn, B.R., Polich, J., 2006. Meditation states and traits: EEG, ERP, and neuroimaging studies. *Psychol. Bull.* 132 (2), 180–211. <https://doi.org/10.1037/0033-2909.132.2.180>.
- Cahn, B.R., Polich, J., 2009. Meditation (Vipassana) and the P3a event-related brain potential. *Int. J. Psychophysiol.* 72 (1), 51–60. <https://doi.org/10.1016/j.ijpsycho.2008.03.013>.
- Cahn, B.R., Delorme, A., Polich, J., 2010. Occipital gamma activation during vipassana meditation. *Cogn. Process.* 11 (1), 39–56. <https://doi.org/10.1007/s10339-009-0352-1>.
- Clayton, M.S., Yeung, N., Cohen Kadosh, R., 2015. The roles of cortical oscillations in sustained attention. *Trends Cogn. Sci.* 19 (4), 188–195. <https://doi.org/10.1016/j.tics.2015.02.004>.
- Cornwell, B.R., Baas, J.M.P., Johnson, L., Holroyd, T., Carver, F.W., Lissek, S., Grillon, C., 2007. Neural responses to auditory stimulus deviance under threat of electric shock revealed by spatially-filtered magnetoencephalography. *NeuroImage* 37 (1), 282–289. <https://doi.org/10.1016/j.neuroimage.2007.04.055>.
- Coronado-Montoya, S., Levis, A.W., Kwakkenbos, L., Steele, R.J., Turner, E.H., Thombs, B.D., 2016. Reporting of positive results in randomized controlled trials of mindfulness-based mental health interventions. *PLOS ONE* 11 (4), e0153220. <https://doi.org/10.1371/journal.pone.0153220>.
- Coseru, C., 2015. *Perceiving Reality: Consciousness, Intentionality, and Cognition in Buddhist Philosophy*. Oxford University Press.
- Dahl, C.J., Lutz, A., Davidson, R.J., 2015. Reconstructing and deconstructing the self: cognitive mechanisms in meditation practice. *Trends Cogn. Sci.* <https://doi.org/10.1016/j.tics.2015.07.001>.
- Davidson, R.J., Kaszniak, A.W., 2015. Conceptual and methodological issues in research on mindfulness and meditation. *Am. Psychol.* 70 (7), 581–592. <https://doi.org/10.1037/a0039512>.

- Delgado-Pastor, L.C., Perakakis, P., Subramanya, P., Telles, S., Vila, J., 2013. Mindfulness (Vipassana) meditation: effects on P3b event-related potential and heart rate variability. *Int. J. Psychophysiol.* 90 (2), 207–214. <https://doi.org/10.1016/j.ijpsycho.2013.07.006>.
- Delorme, A., Makeig, S., 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134 (1), 9–21. <https://doi.org/10.1016/j.jneumeth.2003.10.009>.
- Deolindo, C.S., Ribeiro, M.W., Aratanha, M.A., Afonso, R.F., Irrmischer, M., Kozasa, E.H., 2020. A critical analysis on characterizing the meditation experience through the electroencephalogram. *Front. Syst. Neurosci.* 14, 53. <https://doi.org/10.3389/fnsys.2020.00053>.
- Dienes, Z., 2014. Using bayes to get the most out of non-significant results. *Front. Psychol.* 5 <https://doi.org/10.3389/fpsyg.2014.00781>.
- Donoghue, T., Haller, M., Peterson, E.J., Varma, P., Sebastian, P., Gao, R., Noto, T., Lara, A.H., Wallis, J.D., Knight, R.T., Sheshyuk, A., Voytek, B., 2020. Parameterizing neural power spectra into periodic and aperiodic components. *Nat. Neurosci.* 23 (12), 1655–1665. <https://doi.org/10.1038/s41593-020-00744-x>.
- Duncan, C.C., Barry, R.J., Connolly, J.F., Fischer, C., Michie, P.T., Näätänen, R., Polich, J., Reinvang, I., Van Petten, C., 2009. Event-related potentials in clinical research: guidelines for eliciting, recording, and quantifying mismatch negativity, P300, and N400. *Clin. Neurophysiol.* 120 (11), 1883–1908. <https://doi.org/10.1016/j.clinph.2009.07.045>.
- Dunne, J., 2011. Toward an understanding of non-dual mindfulness. *Contemp. Buddhism* 12 (1), 71–88. <https://doi.org/10.1080/14639947.2011.564820>.
- Fanelli, D., 2010. “Positive” results increase down the hierarchy of the sciences. *PLOS ONE* 5 (4), e10068. <https://doi.org/10.1371/journal.pone.0010068>.
- Ferguson, C.J., Brannick, M.T., 2012. Publication bias in psychological science: prevalence, methods for identifying and controlling, and implications for the use of meta-analyses. *Psychol. Methods* 17 (1), 120–128. <https://doi.org/10.1037/a0024445>.
- Fox, K.C.R., Dixon, M.L., Nijeboer, S., Girn, M., Floman, J.L., Lifshitz, M., Ellamil, M., Sedlmeier, P., Christoff, K., 2016. Functional neuroanatomy of meditation: a review and meta-analysis of 78 functional neuroimaging investigations. *Neurosci. Biobehav. Rev.* 65, 208–228. <https://doi.org/10.1016/j.neubiorev.2016.03.021>.
- Foxe, J., Snyder, A., 2011. The role of alpha-band brain oscillations as a sensory suppression mechanism during selective attention. *Front. Psychol.* 2, 154. <https://doi.org/10.3389/fpsyg.2011.00154>.
- Friston, K., Kiebel, S., 2009. Predictive coding under the free-energy principle. *Philos. Trans. R. Soc.* B 364 (1521), 1211–1221. <https://doi.org/10.1098/rstb.2008.0300>.
- Fucci, E., Abdoun, O., Caclin, A., Francis, A., Dunne, J.D., Ricard, M., Davidson, R.J., Lutz, A., 2018. Differential effects of non-dual and focused attention meditations on the formation of automatic perceptual habits in expert practitioners. *Neuropsychologia*. <https://doi.org/10.1016/j.neuropsychologia.2018.07.025>.
- Fucci, E., Abdoun, O., Lutz, A., 2019. Auditory perceptual learning is not affected by anticipatory anxiety in the healthy population except for highly anxious individuals: EEG evidence. *Clin. Neurophysiol.* 130 (7), 1135–1143. <https://doi.org/10.1016/j.clinph.2019.04.010>.
- Garrido, M.I., Friston, K.J., Kiebel, S.J., Stephan, K.E., Baldeweg, T., Kilner, J.M., 2008. The functional anatomy of the MMN: a DCM study of the roving paradigm. *NeuroImage* 42 (2), 936–944. <https://doi.org/10.1016/j.neuroimage.2008.05.018>.
- Garrido, M.I., Kilner, J.M., Stephan, K.E., Friston, K.J., 2009. The mismatch negativity: a review of underlying mechanisms. *Clin. Neurophysiol.* 120 (3), 453–463. <https://doi.org/10.1016/j.clinph.2008.11.029>.
- Gelman, A., Carlin, J., 2014. Beyond power calculations: assessing type S (Sign) and type M (Magnitude) errors. *Perspect. Psychol. Sci.* 9 (6), 641–651. <https://doi.org/10.1177/1745691614551642>.
- Gelman, A., Hill, J., 2006. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.
- Gelman, A., Loken, E., 2013. In: *The Garden of Forking Paths: Why Multiple Comparisons can be a Problem, Even When There is no “Fishing Expedition” or “p-hacking” and the Research Hypothesis was Posited Ahead of Time*. Department of Statistics, Columbia University, p. 348.
- Gerritsen, R.J.S., Band, G.P.H., 2018. Breath of life: the respiratory vagal stimulation model of contemplative activity. *Front. Hum. Neurosci.* 12, 397. <https://doi.org/10.3389/fnhum.2018.00397>.
- Goodrich, B., Gabry, J., Ali, I., Brilleman, S., 2020. rstanarm: Bayesian applied regression modeling via Stan. <https://mc-stan.org/rstanarm>.
- Gramfort, A., Luessi, M., Larson, E., Engemann, D., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L., Hämäläinen, M., 2013. MEG and EEG data analysis with MNE-python. *Front. Neurosci.* 7, 267. <https://doi.org/10.3389/fnins.2013.00267>.
- Green, P., MacLeod, C.J., 2016. SIMR: an R package for power analysis of generalized linear mixed models by simulation. *Methods Ecol. Evol.* 7 (4), 493–498. <https://doi.org/10.1111/2041-210X.12504>.
- Ioannidis, J.P.A., 2005. Why Most published research findings are false. *PLoS Med.* 2 (8), e124. <https://doi.org/10.1371/journal.pmed.0020124>.
- Isbel, B.D., Lagopoulos, J., Hermens, D.F., Summers, M.J., 2019. Mental training affects electrophysiological markers of attention resource allocation in healthy older adults. *Neurosci. Lett.* 698, 186–191. <https://doi.org/10.1016/j.neulet.2019.01.029>.
- Josefsson, T., Broberg, A., 2011. Meditators and non-meditators on sustained and executive attentional performance. *Ment. Health Relig. Cult.* 14 (3), 291–309. <https://doi.org/10.1080/13674670903578621>.
- Kerr, C.E., Sacchet, M.D., Lazar, S.W., Moore, C.I., Jones, S.R., 2013. Mindfulness starts with the body: somatosensory attention and top-down modulation of cortical alpha rhythms in mindfulness meditation. *Front. Hum. Neurosci.* 7 <https://doi.org/10.3389/fnhum.2013.00012>.
- Klee, D., Colgan, D.D., Hanes, D., Oken, B., 2020. The effects of an internet-based mindfulness meditation intervention on electrophysiological markers of attention. *Int. J. Psychophysiol.* 158, 103–113. <https://doi.org/10.1016/j.ijpsycho.2020.10.002>.
- Klimesch, W., 2012. Alpha-band oscillations, attention, and controlled access to stored information. *Trends Cogn. Sci.* 16 (12), 606–617. <https://doi.org/10.1016/j.tics.2012.10.007>.
- Kubota, Y., Sato, W., Toichi, M., Murai, T., Okada, T., Hayashi, A., Sengoku, A., 2001. Frontal midline theta rhythm is correlated with cardiac autonomic activities during the performance of an attention demanding meditation procedure. *Cogn. Brain Res.* 11 (2), 281–287. [https://doi.org/10.1016/S0926-6410\(00\)00086-0](https://doi.org/10.1016/S0926-6410(00)00086-0).
- Lagopoulos, J., Xu, J., Rasmussen, I., Vik, A., Malhi, G.S., Eliassen, C.F., Arnsten, I.E., Sæther, J.G., Hollup, S., Hølen, A., Davanger, S., Ellingsen, Ø., 2009. Increased theta and alpha EEG activity during nondirective meditation. *J. Altern. Complement. Med.* 15 (11), 1187–1192. <https://doi.org/10.1089/acm.2009.0113>.
- Laukkonen, R.E., Slagter, H.A., 2021. From many to (n)one: meditation and the plasticity of the predictive mind. *Neurosci. Biobehav. Rev.* 128, 199–217. <https://doi.org/10.1016/j.neubiorev.2021.06.021>.
- Lee, D.J., Kulubya, E., Goldin, P., Goodarzi, A., Girgis, F., 2018. Review of the neural oscillations underlying meditation. *Front. Neurosci.* 12, 178. <https://doi.org/10.3389/fnins.2018.00178>.
- Lenth, R., Singmann, H., Love, J., Buerkner, P., Herve, M., 2020. emmeans: estimated marginal means, aka least-squares means (1.4.5) [computer software]. <https://CRAN.R-project.org/package=emmeans>.
- Lomas, T., Ivtzan, I., Fu, C.H., 2015. A systematic review of the neurophysiology of mindfulness on EEG oscillations. *Neurosci. Biobehav. Rev.* 57, 401–410.
- Lutz, A., Dunne, J.D., Davidson, R.J., 2007. Meditation and the neuroscience of consciousness. In: *Cambridge Handbook of Consciousness*, pp. 499–555.
- Lutz, A., Slagter, H.A., Dunne, J.D., Davidson, R.J., 2008. Attention regulation and monitoring in meditation. *Trends Cogn. Sci.* 12 (4), 163–169. <https://doi.org/10.1016/j.tics.2008.01.005>.
- Lutz, A., Jha, A.P., Dunne, J.D., Saron, C.D., 2015. Investigating the phenomenological matrix of mindfulness-related practices from a neurocognitive perspective. *Am. Psychol.* 70 (7), 632–658. <https://doi.org/10.1037/a0039585>.
- Lutz, A., Mattout, J., Pagnoni, G., 2019. The epistemic and pragmatic value of non-action: a predictive coding perspective on meditation. *Curr. Opin. Psychol.* 28, 166–171. <https://doi.org/10.1016/j.copsyc.2018.12.019>.
- Ly, A., Etz, A., Marsman, M., Wagenmakers, E.-J., 2019. Replication bayes factors from evidence updating. *Behav. Res. Methods* 51 (6), 2498–2508. <https://doi.org/10.3758/s13428-018-1092-x>.
- Makeig, S., Westerfield, M., Jung, T.-P., Enghoff, S., Townsend, J., Courchesne, E., Sejnowski, T.J., 2002. Dynamic brain sources of visual evoked responses. *Science* 295 (5555), 690–694. <https://doi.org/10.1126/science.1066168>.
- Makowski, D., Ben-Shachar, M.S., Lüdtke, D., 2019. BayesestR: describing effects and their uncertainty, existence and significance within the Bayesian framework. *J. Open Source Softw.* 4 (40), 1541.
- Näätänen, R., Pakarinen, S., Rinne, T., Takegata, R., 2004. The mismatch negativity (MMN): towards the optimal paradigm. *Clin. Neurophysiol.* 115 (1), 140–144. <https://doi.org/10.1016/j.clinph.2003.04.001>.
- Näätänen, R., Paavilainen, P., Rinne, T., Alho, K., 2007. The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clin. Neurophysiol.* 118 (12), 2544–2590. <https://doi.org/10.1016/j.clinph.2007.04.026>.
- Näätänen, R., Kujala, T., Escera, C., Baldeweg, T., Kreegipuu, K., Carlson, S., Ponton, C., 2012. The mismatch negativity (MMN) – a unique window to disturbed central auditory processing in ageing and different clinical conditions. *Clin. Neurophysiol.* 123 (3), 424–458. <https://doi.org/10.1016/j.clinph.2011.09.020>.
- Näätänen, R., Sussman, E.S., Salisbury, D., Shafer, V.L., 2014. Mismatch negativity (MMN) as an index of cognitive dysfunction. *Brain Topogr.* 27 (4), 451–466. <https://doi.org/10.1007/s10548-014-0374-6>.
- Nakagawa, S., Schielzeth, H., 2013. A general and simple method for obtaining R² from generalized linear mixed-effects models. *Methods Ecol. Evol.* 4 (2), 133–142. <https://doi.org/10.1111/j.2041-210x.2012.00261.x>.
- Nosek, B.A., Beck, E.D., Campbell, L., Flake, J.K., Hardwicke, T.E., Mellor, D.T., van ’t Veer, A.E., Vazire, S., 2019. Preregistration is hard, and worthwhile. *Trends Cogn. Sci.* 23 (10), 815–818. <https://doi.org/10.1016/j.tics.2019.07.009>.
- Pagnoni, G., Guareschi, F.T., 2017. Remembrance of things to come: a conversation between zen and neuroscience on the predictive nature of the mind. *Mindfulness* 8 (1), 27–37. <https://doi.org/10.1007/s12671-015-0438-z>.
- Payne, J.R., Baell, O., Geddes, H., Fitzgibbon, B., Emmonson, M., Hill, A.T., Van Dam, N.T., Humble, G., Fitzgerald, P.B., Bailey, N.W., 2019. Experienced meditators exhibit no differences to demographically matched controls in theta phase synchronization, P200, or P300 during an auditory oddball task. *Mindfulness*. <https://doi.org/10.1007/s12671-019-01287-4>.
- Percival, D.B., Walden, A.T., B, P.D., T, W.A., 1993. *Spectral Analysis for Physical Applications*. Cambridge University Press.
- Poldrack, R.A., Baker, C.I., Durnez, J., Gorgolewski, K.J., Matthews, P.M., Munafò, M.R., Nichols, T.E., Poline, J.-B., Vul, E., Yarkoni, T., 2017. Scanning the horizon: towards transparent and reproducible neuroimaging research. *Nat. Rev. Neurosci.* 18 (2), 115–126. <https://doi.org/10.1038/nrn.2016.167>.
- Ramsburg, J.T., Youmans, R.J., 2014. Meditation in the higher-education classroom: meditation training improves student knowledge retention during lectures. *Mindfulness* 5 (4), 431–441. <https://doi.org/10.1007/s12671-013-0199-5>.
- Rentsch, J., Shen, C., Jockers-Scherübel, M.C., Gallinat, J., Neuhaus, A.H., 2015. Auditory mismatch negativity and repetition suppression deficits in schizophrenia explained by irregular computation of prediction error. *PLoS ONE* 10 (5). <https://doi.org/10.1371/journal.pone.0126775>.

- Rodriguez-Larios, J., Faber, P., Achermann, P., Tei, S., Alaerts, K., 2020. From thoughtless awareness to effortful cognition: alpha - theta cross-frequency dynamics in experienced meditators during meditation, rest and arithmetic. *Sci. Rep.* 10 (1), 5419. <https://doi.org/10.1038/s41598-020-62392-2>.
- Sams, M., Paavilainen, P., Alho, K., Näätänen, R., 1985. Auditory frequency discrimination and event-related potentials. *Electroencephalogr. Clin. Neurophysiol.* 62 (6), 437–448. [https://doi.org/10.1016/0168-5597\(85\)90054-1](https://doi.org/10.1016/0168-5597(85)90054-1).
- Schad, D.J., Betancourt, M., Vasishth, S., 2021. Toward a principled Bayesian workflow in cognitive science. *Psychol. Methods* 26 (1), 103.
- Schoenberg, P.L.A., Ruf, A., Churchill, J., Brown, D.P., Brewer, J.A., 2018. Mapping complex mind states: EEG neural substrates of meditative unified compassionate awareness. *Conscious. Cogn.* 57, 41–53. <https://doi.org/10.1016/j.concog.2017.11.003>.
- Srinivasan, N., Bajjal, S., 2007. Concentrative meditation enhances preattentive processing: a mismatch negativity study. *Neuroreport* 18 (16), 1709–1712. <https://doi.org/10.1097/WNR.0b013e3282f0d2d8>.
- Strauß, A., Wöstmann, M., Obleser, J., 2014. Cortical alpha oscillations as a tool for auditory selective inhibition. *Front. Hum. Neurosci.* 8, 350. <https://doi.org/10.3389/fnhum.2014.00350>.
- Sussman, E.S., Shafer, V.L., 2014. New perspectives on the mismatch negativity (MMN) component: an evolving tool in cognitive neuroscience. *Brain Topogr.* 27 (4), 425–427. <https://doi.org/10.1007/s10548-014-0381-7>.
- Takahashi, T., Murata, T., Hamada, T., Omori, M., Kosaka, H., Kikuchi, M., Yoshida, H., Wada, Y., 2005. Changes in EEG and autonomic nervous activity during meditation and their association with personality traits. *Int. J. Psychophysiol.* 55 (2), 199–207. <https://doi.org/10.1016/j.ijpsycho.2004.07.004>.
- Tang, Y.-Y., Hölzel, B.K., Posner, M.I., 2015. The neuroscience of mindfulness meditation. *Nat. Rev. Neurosci.* 16 (4), 213–225. <https://doi.org/10.1038/nrn3916>.
- Vago, D.R., Silbersweig, D.A., 2012. Self-awareness, self-regulation, and self-transcendence (S-ART): a framework for understanding the neurobiological mechanisms of mindfulness. *Front. Hum. Neurosci.* 6 <https://doi.org/10.3389/fnhum.2012.00296>.
- van Lutterveld, R., van Dellen, E., Pal, P., Yang, H., Stam, C.J., Brewer, J., 2017. Meditation is associated with increased brain network integration. *NeuroImage* 158, 18–25. <https://doi.org/10.1016/j.neuroimage.2017.06.071>.
- Wacongne, C., Changeux, J.-P., Dehaene, S., 2012. A neuronal model of predictive coding accounting for the mismatch negativity. *J. Neurosci.* 32 (11), 3665–3678. <https://doi.org/10.1523/JNEUROSCI.5003-11.2012>.
- Wagenmakers, E.-J., Gronau, Q.F., Dablander, F., Etz, A., 2020. The support interval. *Erkenntnis*. <https://doi.org/10.1007/s10670-019-00209-z>.
- Weisz, N., Obleser, J., 2014. Synchronisation signatures in the listening brain: a perspective from non-invasive neuroelectrophysiology. *Hear. Res.* 307, 16–28. <https://doi.org/10.1016/j.heares.2013.07.009>.
- Yoshida, K., Takeda, K., Kasai, T., Makinae, S., Murakami, Y., Hasegawa, A., Sakai, S., 2020. Focused attention meditation training modifies neural activity and attention: longitudinal EEG data in non-meditators. *Soc. Cogn. Affect. Neurosci.* 15 (2), 215–224. <https://doi.org/10.1093/scan/nsaa020>.
- Yu, X., Fumoto, M., Nakatani, Y., Sekiyama, T., Kikuchi, H., Seki, Y., Sato-Suzuki, I., Arita, H., 2011. Activation of the anterior prefrontal cortex and serotonergic system is associated with improvements in mood and EEG changes induced by zen meditation practice in novices. *Int. J. Psychophysiol.* 80 (2), 103–111. <https://doi.org/10.1016/j.ijpsycho.2011.02.004>.
- Zaccaro, A., Piarulli, A., Laurino, M., Garbella, E., Menicucci, D., Neri, B., Gemignani, A., 2018. How breath-control can change your life: a systematic review on psychophysiological correlates of slow breathing. *Front. Hum. Neurosci.* 12 <https://doi.org/10.3389/fnhum.2018.00353>.

4.2 Study 2: Time-resolved dynamic computational modeling of human EEG recordings reveals gradients of generative mechanisms for the MMN response

Time-resolved dynamic computational modeling of human EEG recordings reveals gradients of generative mechanisms for the MMN response

Arnaud Poublan-Couzardot¹, Françoise Lecaigard¹, Enrico Fucci², Richard Davidson³⁻⁶, Jérémie Mattout¹, Antoine Lutz^{1‡}, Oussama Abdoun^{1‡*}

1 Lyon Neuroscience Research Centre (INSERM U1028, CNRS UMR5292, Lyon 1 University), Lyon, France

2 Institute for Globally Distributed Open Research and Education (IGDORE), Fuerteventura, Spain

3 Center for Healthy Minds, University of Wisconsin – Madison

4 Department of Psychology, University of Wisconsin – Madison

5 Waisman Laboratory for Brain Imaging and Behavior, University of Wisconsin – Madison

6 Department of Psychiatry, University of Wisconsin – Madison

‡ AL and OA are joint Senior Authors.

* oussama.abdoun@protonmail.com

Abstract

Over ten years after the first review trying to unify the different theoretical accounts of the mismatch negativity (MMN) under the predictive coding framework, there is still an ongoing debate on the neurophysiological mechanisms responsible for the MMN. Low-level models of neuronal adaptation to recurrent stimuli are able to explain most of the MMN features. On the other hand, several modeling studies reported evidence for Bayesian learning models to explain human event-related potentials. However, adaptation models were not systematically included in model comparison or could not be rejected decisively. Based on reports indicating spatial and temporal dissociation of physiological mechanisms within the timecourse of mismatch responses in animals, we hypothesized that different computational models could underlie different temporal phases of the human MMN. Using electroencephalography data from two independent studies of a simple auditory oddball task (82 subjects in total after pre-processing rejections), we compared adaptation and Bayesian learning models' ability to explain the trial-to-trial dynamics of auditory deviance detection in a time-resolved fashion. We first ran simulations to evaluate the capacity of our design to dissociate the tested models and found that they were indistinguishable below a certain level of signal-to-noise ratio (SNR). In subjects with a sufficient SNR, our time-resolved approach revealed a temporal dissociation between the two model families, with high evidence for adaptation during the early MMN window (from 90 to 150-190 ms post-stimulus depending on the dataset) and for Bayesian learning later in time (170-180 ms or 200-220ms). Comparison of parameterized models within the adaptation family revealed a gradient of adaptation rates. This finding evokes an analogous anatomical gradient found in animal studies within the auditory cortical hierarchy.

Author summary

The ability to detect and adapt to changes in the environment is an essential feature for survival of living beings. Two main theories have been proposed to explain how our brain performs such an automatic task in the auditory domain. The first one, adaptation, gives prominence to the ability of auditory cortical and sub-cortical neurons to attenuate their response to repeated stimuli, which renders the brain more sensitive to deviations from expected sensory inputs. The second one, Bayesian learning, further involves higher-level cortical regions, which would update their predictions about incoming tones, depending on their performance at predicting previous ones. These two views may not be mutually exclusive, but few experimental work tried to directly compare them. We used computational models inspired from both accounts to assess which view may provide a better fit of two independent electrophysiological datasets from similar auditory experiments.

Evidence from 82 subjects provided a complex picture, with adaptation processes seemingly dominating the early phase of auditory brain response, and Bayesian learning processes appearing later on. Our results converge with other recent works and emphasize the necessary reconciliation of those two theories for a better understanding of human auditory perception.

Introduction 1

The MMN, a widely used tool in neuroscientific research 2

The auditory mismatch negativity (MMN) is a well-investigated electrophysiological marker of automatic neural detection of changes occurring in the auditory environment [1]. The MMN is typically induced using an oddball paradigm where sequences of frequent standard tones are presented interspersed with rare deviant tones. It is most often observed in humans through electroencephalography (EEG) or magnetoencephalography (MEG) and is defined as the difference between the deviant and standard event-related potentials (ERPs) [2]. There are multifarious deviance features including pitch, duration, intensity, spatial location, etc. The MMN is even observed in response to violations of abstract regularities that bear no relationship with the physical features of the stimuli [3]. 10

The MMN is a negative wave generally peaking over fronto-central scalp electrodes between 100 and 250 milliseconds, with variations depending on the experimental design including the dimension of the deviating feature [1]. Discovered more than 40 years ago [4], it was investigated in research areas as broad as neurolinguistic [5], attention [6], altered states of consciousness [7,8], aging [9], psychiatry [10,11], emotions [12], psychedelics [13] or meditation [14]. However despite its use in clinical research or as an experimental measure for many cognitive functions [15], its generative mechanisms are still widely debated. 16

An old debate around the generative mechanisms of the MMN 17

Recent reviews described in details the different mechanistic hypotheses trying to explain mismatch responses [16,17] and we will only briefly described them here. The MMN was originally viewed as a marker of sensory memory by its discoverers [18]. Yes this simplistic *memory-based* account could not explain the MMN response to abstract deviations [3] and was refined under the *model-adjustment hypothesis* [19,20]. In this view the main mechanism behind the MMN is not only the detection of changes in some physical features of a stimulus, but more generally the detection, representation and updating of regularities in the auditory stream. Being more flexible, the model adjustment account can explain most of the MMN properties, from 24

omission responses to violations of abstract rules. 25

Another prevailing explanation for the MMN is the *adaptation* hypothesis which rests upon electrophysiological observations of stimulus-specific adaptation (SSA) in auditory brain regions at the origin of deviance detection [21–23]. Indeed midbrain [24], thalamic [25] and cortical [26–29] auditory neurons tuned to frequently repeated tones adapt faster than neurons tuned to rare deviant tones. This differential adaptation process observed at the neuronal level is supposed to account for the scalp N1 ERP to frequent tones (standards) to be attenuated compared to the N1 response to novel fresh stimuli (deviants), and thus giving rise to the negative deflection called the MMN [30–32]. If critics were raised concerning the explanatory power of the adaptation hypothesis and its relation to the MMN [33–35], most of them were fairly answered with simulations works [30,31,36,37] and detailed reviews [22,38]. As May et al. [22] highlighted, memory-based and adaptation accounts fall in two different categories: psychological and physiological, respectively, which are not mutually exclusive explanations of the MMN but justify further efforts in the development of a more encompassing theory of deviance detection. 37

Bayesian inference, and its implementation in the brain as *predictive coding*, has been used extensively in the last two decades as a general model of perception [39–43], with particular emphasis on audition [44–46]. In predictive coding, the MMN is seen as an electrophysiological marker of the discrepancy between top-down predictions based on repeated stimuli (i.e. standards) and bottom-up unexpected sensory inputs (i.e. deviants): a prediction error [40]. Propagated through forward connections, prediction errors would trigger the updating and optimization of an internal model of sensory causes at higher cortical areas, in agreement with the model-adjustment hypothesis. Adjusted predictions would then be sent backward to explain away or suppress prediction errors at lower cortical areas. In this framework, suppression of the MMN when deviant tones are repeated and thus become standards, as in the roving standards paradigm [47], is interpreted as a minimization of prediction errors, the internal model having learnt a new statistical regularity. Additionally, hierarchical minimization of prediction errors would also be weighted by the relative precision of the predictions and sensory inputs. Physiologically, changes in local, intrinsic connectivity of prediction error units would mediate this weighting process in auditory cortices via modifications of synaptic gains, reminiscent of the adaptation account. Using Dynamic Causal Modelling (DCM) on human EEG data, Garrido and colleagues showed that model-adjustment or adaptation hypotheses alone are not sufficient to explain changes in extrinsic and intrinsic connectivity specific to the MMN in both classical [48–51] and roving [52] auditory oddball paradigms. Converging with previous attempts to unify [53], they proposed that both hypotheses are necessary to fully account for N1 and MMN brain responses, and are reconcilable under predictive coding [54]. Empirical support for this integrated account has come from the previously mentioned 56

DCM studies on EEG [48–52] and MEG data [55,56] as well as EEG-MEG fusion [57] and electrocortigraphy (ECoG) recordings [56]; but also from more classical ERP and sources reconstruction analyses using refined paradigms with EEG and MEG [58] or ECoG [59], to reveal the hierarchical organization of deviance detection processes in auditory, associative and frontal cortices.

In parallel to above MMN studies in humans, an extensive field of research has developed on deviance detection in the animal brain. The possibility to use more invasive methods on cats and rodents allows to gain additional insights on the neural mechanisms of deviance processing.

Insights from animal studies

Animal studies of deviance detection have focused on the neuronal dynamics and location of such effects, identifying SSA as the core neuronal mechanism at the microscopic level [26,27], and defining the MMN as the macroscopic correlate of SSA summed over multiple neuronal populations [21]. However theoretical and methodological issues were encountered while trying to integrate SSA and MMN under the same deviance detection mechanism [21,60]. More recently, following human research, results in animal studies have been reinterpreted under the predictive coding framework with convincing explanatory power [16,61]. Building on new paradigms allowing to disentangle mere adaptation (repetition suppression mechanism) from prediction error [16] and single-neuron recordings, researchers were able to demonstrate empirically the hypothesized hierarchical organization of predictive coding in the auditory system at the neuronal level [62]. If neuronal mismatch activity was predominantly explained by repetition suppression, a significant part could also be accounted for by “genuine” deviance detection (prediction error). Notably such prediction error–like activity was found as soon as the inferior colliculus in the midbrain, and gradually increased when ascending to thalamic and cortical areas [62], a result challenging the classical view of predictive coding as a cortex restricted mechanism. In addition, the neurons in the non-lemniscal divisions of the recorded brain areas systematically showed higher prediction error–like activity than their lemniscal counterparts at each level of the auditory hierarchy. This empirical finding is consistent with the known anatomical and functional connectivity of the non-lemniscal pathway. Indeed, Carabajal and Malmierca (2018) have argued that the fact that broadly tuned non-lemniscal neurons receive the bulk of descending cortical signals, as well as inputs from the neighbouring, sharply tuned and tonotopically organized lemniscal neurons, make it a good candidate for hierarchical predictive coding [16]. Interestingly, in a subsequent study, the same authors tested again this dichotomy at frontal areas, confirming the existence of genuine mismatch responses in the prefrontal cortex (PFC) of anesthetized rats [63]. Those responses differ in multiple aspects from auditory cortex responses, adding

further evidence for the hypothesis of divergent but complementary roles of frontal and sensory areas in predictive coding. Indeed responses at prefrontal areas were mostly explained by a prediction error index, and occurred later than auditory cortex responses dominated by repetition suppression mechanisms.

Insights from computational modeling

Computational modeling has been of great help to try to better understand MMN generation mechanisms under adaptation and predictive coding accounts. Independent modeling efforts have designed models of adaptation at the neuronal [30, 64, 65] and cortical level [37, 66], while a predictive coding model was implemented with a neuronal architecture simulating a thalamo-cortical network [67]. Additionally, more phenomenological models linked auditory perception as hierarchical Bayesian inference to electrophysiological responses such as the MMN [68, 69]. However, if these independent models successfully accounted for critical SSA or MMN features, few studies directly compared models derived from divergent MMN theories. A first attempt found a “Bayesian surprise” model, consistent with predictive coding accounts, to outperform classical “stimulus change” (or change detection) models at predicting somatosensory mismatch EEG responses at the source level [70]. This work was extended to the auditory MMN in fused EEG-MEG [71] and electrocortigraphy (ECoG) [72] modalities, confirming the previous results and designing new Bayesian learning models of auditory deviance.

Only one study directly compared all major MMN theories: change detection, adaptation, and three flavours of Bayesian inference derived from the Free Energy Principle [42]—predictive coding, model adjustment and novelty detection [73]. In their 2013 modelling study of auditory MMN responses elicited by a roving oddball paradigm [73], Lieder and colleagues found substantially higher evidence for a set of Bayesian models compared to a set of alternative models (adaptation and change detection models). However, this result hinged on the model comparison being carried out at the level of families aggregating multiple models, and no conclusion could be drawn at the level of individual models, possibly due to an insufficient sample size (8 participants). Critically, the analysis was performed on a time-averaged signal, rendering it blind to any temporal differences in MMN generation processes.

The current study

In the current study, we used two independent EEG datasets of 28 and 54 subjects from similar auditory oddball paradigms to compare five model families, each inspired by prevailing MMN theories (change detection, adaptation, predictive coding, model adjustment and novelty detection). Building on previous

modeling studies [70,71,73], we defined an extensive set of models and employed a time-resolved dynamic modeling approach to test a potential temporal dissociation in the computational mechanisms explaining the MMN, as suggested by recent electrophysiological findings [63]. As a first validity check and methodological improvement from former studies, we ran confusion analyses on simulated data with varying signal-to-noise ratio (SNR). Under our design and models set, model recovery was highly dependent on SNR and allowed us to test the veracity of later model comparison results. Indeed Bayesian model fitting and comparison revealed a clear distinction in models' families dominance between early and late windows of the MMN, that could not be explain by confusion between the models. We could also identify a temporal gradient within adaptation models time constant, reminiscent of recent electrophysiological results in the animal literature [29,63].

Materials and methods

For the present trial-by-trial modeling study, we used datasets from two published EEG experiments investigating the impact of meditation states and meditation expertise on the amplitude of the MMN evoked potential [14], [74]. The original experiments have been extensively described in the corresponding articles; below we report only the materials and methods relevant to the current study. In particular, only the data from the control conditions (defined below) of the two original experiments have been used for modeling. To ensure the quality of the modeled data and thus of the results derived from it, participants with insufficient data after preprocessing or lacking a detectable MMN based on objective criteria were rejected. The last part of the methods section describes in details our modeling approach. We introduce the specific framework used and detail the computational implementations of the prevailing accounts offered as generative models of the MMN response. The parameters of the model fitting procedure are provided for replication purposes. The approach we implemented for inference based on Bayesian model comparison as well as model recovery analysis are described with their underlying rationale. A last section concerns the measure used to evaluate the absolute goodness-of-fit of the selected models.

Study 1

Participants

Data from 31 healthy participants (25 males, mean age of 42.9 ± 10.4 years old) were collected. The research protocol was approved by the UW-Madison Health Sciences Internal Review Board and participants provided written informed consent for all study procedures.

Experimental design

144

Subjects underwent a passive auditory oddball task [75] consisting of the variable repetition of a standard tone (1000 Hz; 60 ms duration; 10 ms rise and fall; 80 dB SPL) followed by the presentation of a frequency deviant tone (1200 Hz; 60 ms duration; 10 ms rise and fall; 80 dB SPL). Each block of the task contained 80% standard tones ($n = 200$) and 20% deviant tones ($n = 50$) with a variable inter-stimulus interval (ISI) of 800–1200 ms. Each subject underwent three blocks of three conditions each: two meditative practices and one control condition. During the control condition, subjects were instructed to read a newspaper and ignore the auditory stimulation. We refer the interested reader to the mentioned article [14] for details about the meditation instructions. The order of conditions within a block was randomized but all subjects had the same order, with the control condition being the first condition in the first block and then the third condition in the second and third block.

145
146
147
148
149
150
151
152
153
154

EEG data was collected with a 128-channel Geodesic Sensor Net (Electrical Geodesics, Eugene, OR), sampled at 500 Hz, and referenced to the vertex (Cz).

155
156

EEG data preprocessing

157

Data were pre-processed using the EEGLAB software [76] on Matlab (The Mathworks Inc.). Notch filtering was applied at 60 Hz to remove line noise, followed by the application of a band-pass filter between 0.5 and 100 Hz. Raw data were manually cleared of large movement-related artefacts and bad channels were identified and interpolated. Independent Component Analysis (ICA) was applied to the raw data of each participant (only on the non-interpolated channels) using runica (Infomax) algorithm [77] to identify and remove artefacts caused by eye blinks and saccades, as well as cardiac and muscular activity. After ICA correction, data were re-referenced offline to the average of both mastoids, a non-causal band-pass 1–60 Hz digital filter was applied, and two-second epochs centred on stimulus onset (-1 to 1 s) were extracted. Epochs were visually inspected and the ones still presenting muscular or eye movement artefacts were marked for rejection. Finally, the epoched data were baseline-corrected by subtracting the mean value of the signal during the 100 ms window preceding stimulus presentation.

158
159
160
161
162
163
164
165
166
167
168

We focused the present analysis on the same region of interest (ROI) as Fucci and colleagues [14], a frontal area comprising 12 electrodes selected based on the previous literature (including Fz). This ROI can be visualized in Fig S1A. The data fed to the models was always the average over these ROI electrodes.

169
170
171

Study 2

172

Participants

173

The second dataset used for modeling comes from a subsequent, replication study by the same authors [74]. 66 participants were included in this study (30 females, mean age of 52 ± 7.7 years old). All participants signed an informed consent before the beginning of the experiment, and were paid for their participation. Ethical approval was obtained from the appropriate regional ethics committee on Human Research (CPP Sud-Est IV, 2015-A01472-47).

174

175

176

177

178

Experimental design

179

The experiment consisted in a passive auditory oddball task with sequences of standard tones of variable length (880Hz; 80ms duration; 10ms rise and fall) followed by the presentation of a frequency deviant (988Hz; 20% of all auditory stimuli). The stimuli were presented binaurally with a fixed ISI of 500ms. A threat induction procedure [78] was added to the experiment. Participants were exposed to two experimental conditions while passively listening to the tones : “threat” periods during which they were informed they could receive an electrical shock on the wrist (whose amplitude was calibrated beforehand, based on the participant pain threshold), and “safe” periods where they knew no shocks would be delivered. Safe and threat periods were alternating randomly every 70 stimuli (35 seconds) and participants were cued to the change. This threat induction oddball paradigm was split into three blocks where subjects were either asked to: watch a silent movie (control condition), practice a focused attention meditation, or practice an open presence meditation. The order of the blocks were randomized between participants. 224 standards and 56 deviants were presented per block and per condition (4 safe and 4 threat periods per block). The full paradigm was repeated in a second session two hours later. For the purpose of the current article we will focus on EEG data from the control condition (watching a silent movie) and safe periods (no threat of electrical shock) only. EEG data was collected in a shielded Faraday chamber with a 64 electrodes Biosemi ActiveTwo system at a 512 Hz sampling frequency, offset was maintained within 50mV (± 25) as recommended in the Biosemi guidelines.

180

181

182

183

184

185

186

187

188

189

190

191

192

193

194

195

196

EEG data preprocessing

197

Pre-processing was done using EEGLAB [76] and in-house Matlab scripts (Matlab version R2017a, The Mathworks Inc). EEG data was downsampled to 250 Hz, re-referenced to the average of the mastoids and visually inspected, noisy channels (deficient electrodes, long-lasting muscular artefacts, etc) were marked

198

199

200

and removed temporarily. For every subject, and separately for the two sessions, Independent Component Analysis (ICA, with runica algorithm) was applied on 1–20 Hz filtered data (EEGLAB eegfiltnew function, Hamming windowed sinc FIR filter of order 826) to improve ICA decomposition. ICA weights hence obtained were applied to the unfiltered data and ICA components reflecting blinks and eye movements were manually selected and rejected. Subsequently a 2 Hz high-pass filter (Hamming windowed sinc FIR filter of order 414) was applied to remove low-frequency drifts most likely caused by stress-related perspiration induced by the threat of electrical shocks. Previously marked bad channels were interpolated using EEGLAB spherical splines interpolation and a line noise removal algorithm (Cleanline, linefreqs=50, bandwidth=2, tau=100, winsize=4, winstep=1) was applied.

Cleaned data were epoched between -200ms and 500ms around auditory tones. The baseline was chosen as the 100 ms period before stimulus onset, as recommended in most of the MMN literature, and removed for subsequent analyses. Finally epoched data were low-pass filtered at 60 Hz (order 56), a ± 75 mV rejection threshold was applied and epochs marked for rejection visually inspected. For this article where only data from the control state (watching a video) were analysed, the recording for one subject was excluded due to an excessive amount of high-frequency noise and two others because they ended up with less than 35 remaining good epochs per condition to compute the MMN (deviant and standard before deviant). This led to 63 participants remaining for Study 2 after EEG preprocessing.

For subsequent analyses we selected the same frontal ROI as [74] (Fig S1B), comprising only 5 electrodes but similarly located as the ones in Study 1 when taking into account the difference in electrodes layout (EEG system with 64 electrodes for Dataset 2 instead of 128 for Dataset 1). We always modeled the average over these 5 electrodes ROI.

Validation of individual MMN

The MMN is a rather robust component observable in virtually all human participants, but in some occasions it might fail to be detected even in healthy participants with intact sound discrimination [79,80]. We used a simple automatic procedure to identify those participants that did not show a detectable MMN. For each participant, a Gaussian function was fitted on the standards minus deviants ERPs difference, averaged over all selected ROI electrodes, by optimizing two free parameters—the offset level and the standard deviation—as well as two constrained parameters: the temporal location of the peak forced between 100 and 200ms and the amplitude of the peak, restricted to negative values. If the optimizer converged on a zero amplitude solution, it was interpreted as failing to detect a negative deflection in the signal and the participant was excluded.

The results of the Gaussian fit on the ERP procedure are represented in Fig S2. As a result of the procedure, 3 participants out of the 31 in study 1 were excluded, as well as 9 out of 63 in study 2.

Modeling

Modeling framework

The goal of this study was to compare alternative accounts of MMN generation, namely adaptation and Bayesian learning. Importantly, those accounts are based on the temporal dependency to previously heard tones, a feature that is difficult to explore with traditional average-based ERP analyses. Thus we used dynamic modeling to explore trial-by-trial fluctuations of deviance detection processes. Static models were also designed to serve as controls. To model MMN responses in a dynamic way we used variational Bayes inference as implemented in the VBA toolbox [81]. This framework uses non-linear state-space models to predict a *response* y (or measurement, e.g. EEG scalp activity) to experimental *inputs* u (e.g. auditory tones), given internal (hidden) states \mathbf{x} that are not directly observable. Such models are composed of two functions working together. First an **evolution function** f describes how internal *hidden states* \mathbf{x} evolve over time (here represented by trial index i):

$$\mathbf{x}_{i+1} = f(\mathbf{x}_i, u_i, \boldsymbol{\theta}) \quad (1)$$

where $\boldsymbol{\theta}$ is a vector of *evolution parameters* and u_i codes the identity of trial i (standard tones: $u = 0$, deviant tones: $u = 1$).

Second we need to define an **observation function** g that computes the *predicted response* generated by hidden states \mathbf{x} . The predicted response is connected to neural activity through a simple linear function:

$$y_i = \phi_0 + \phi_1 g(\mathbf{x}_i, u_i, \boldsymbol{\lambda}) + \epsilon_i \quad (2)$$

where ϕ_0 and ϕ_1 are (fitted) *observation parameters* and ϵ the Gaussian measurement noise which is always modeled but will be omitted in later description of the models for the sake of simplicity. $\boldsymbol{\lambda} = (\lambda_0, \lambda_1)$ is a model-dependent pair of offset and scale factor that normalize the observation function to a similar range across models. Unlike ϕ , $\boldsymbol{\lambda}$ is not fitted but estimated for each model using synthetic data, as such they are not model parameters *per se*. These normalization constants allow the parameters ϕ to share the same interpretation across models. In practice, we set $\boldsymbol{\lambda}$ so that ϕ_0 can be interpreted as the mean response to standard tones ($\bar{y}_{u=0}$), and ϕ_1 as the mean deviant minus standard difference potential ($\bar{y}_{u=1} - \bar{y}_{u=0}$).

Model inversion over all EEG trials was performed independently for every model, subject and 10ms averaged time window ranging from 45 to 255ms post-stimulus. It issued inversion diagnostics, model's posterior parameters estimation and goodness-of-fit measures which will be described later. We will now present in more detail the generative models used to formalize the MMN hypotheses at stake.

Generative models of brain responses

We will first describe the two static models used as controls, namely the *null model* and the *deviant detection*. By definition, static models have no evolution function.

The **Null model** predicts a constant response to any auditory stimulus.

$$g(u_i) = 0 \tag{3}$$

Deviant detection (DD) model predicts the MMN to be a response to deviant stimuli only:

$$g(u_i) = u_i \tag{4}$$

Adaptation models are based on the neuronal adaptation hypothesis [22] which sees the auditory MMN as resulting from a difference in adaptation of frequency-specific cortical neurons, a process often called stimulus specific adaptation (SSA). Neurons tuned to the frequency of standard tones adapt quickly with recurrent repetitions thus decreasing their responsiveness, while neurons responding preferentially to the frequency of deviant tones presented rarely will adapt less and therefore produce a stronger response than their standard counterparts. SSA has been observed with electrophysiology throughout the entire auditory hierarchy, particularly in cortical areas [23]. It is commonly and adequately modeled computationally as a process of exponential decay [72, 82, 83]. For the sake of simplicity we will abbreviate adaptation models to SSA, from the supposed underlying electrophysiological mechanism, yet one has to keep in mind that we do not model a biologically plausible implementation of SSA (see [30, 37, 64, 66] for such works) but only its phenomenological approximation as exponential decay.

In the implementation of adaptation dynamics two hidden states (x_0, x_1) represent neuronal responsiveness to standards and deviants, respectively. Adaptation and recovery are each modeled with exponential functions. Neural activity in response to the presented tones is then supposed to be proportional to the responsiveness of tone-specific neuronal populations:

$$f(x_i, u_i) : \begin{cases} \text{if } u_i = 0 \text{ (standard)} \\ \text{if } u_i = 1 \text{ (deviant)} \end{cases} \begin{cases} x_{0,i+1} = x_{0,i}K_a \\ x_{1,i+1} = 1 - (1 - x_{1,i})K_r \\ x_{0,i+1} = 1 - (1 - x_{0,i})K_r \\ x_{1,i+1} = x_{1,i}K_a \end{cases} \quad (5)$$

with $K_a = e^{-1/\tau_a}$ and $K_r = e^{-1/\tau_r}$

$$g(x_i, u_i, \boldsymbol{\lambda}) = \lambda_0 + x_{u_i,i} \lambda_1$$

Single-cell recordings in the cat auditory cortex (A1) have revealed the existence of multiple timescales of adaptation ranging from less than a second to several tens of seconds with recovery rates about twice as high [27]. Based on this literature we implemented six SSA models with different adaptation parameters $\tau_a \in [3, 10, 20, 30, 50, 100, 200]$. The recovery parameter τ_r was always set to be twice τ_a , as informed by electrophysiological data on auditory neurons of anesthetized cats [27].

Bayesian Learning (BL) models are based on the predictive coding view of the MMN, they assume the brain learns the probability to have a deviant. We inspired from computational models initially designed by Ostwald et al. [70], which treat the MMN as a measure of Bayesian surprise : the difference between prior and posterior distributions of beliefs on states of the world (here states being the category of tone stimulus : deviant or standard). Such models also implement an approximate learning of the statistical regularities of the environment by setting a “forgetting” constant. In their study they showed that these models predict better the somatosensory mismatch response than control models that do not learn. This finding was replicated in the auditory modality combining EEG and MEG measurements [71].

A common evolution function models auditory tones as the outcome of a *Bernoulli process* of parameter $\mu \in [0, 1]$, the probability to have a deviant. The belief over μ follows a *Beta distribution*: $\mu \sim \text{Beta}(\alpha, \beta)$ with parameters α and β , tracking the number of deviants and standards that have been heard so far, respectively. The particularity of BL models is to compare prior and posterior beliefs over μ . We thus assigned hidden states to α and β parameters of the Beta distribution over μ . $x_{1,i}$ and $x_{2,i}$ therefore define the **prior belief** over deviant’s probability before having heard the current tone u_i , while $x_{1,i+1}$ and $x_{2,i+1}$ define the **posterior belief**. We can see that the main characteristic of these hidden states is to implement a memory of past inputs. However to be biologically plausible, and take into account recent inputs history more representative

of ongoing dynamics, this memory should not be “infinite”. Following previous implementations [70, 71], we weighted the stimulus counts by a “forgetting” constant depending on a *temporal integration parameter* τ_t : $K_t = e^{-1/\tau_t}$, which is hypothesized to account for different temporal integration windows. Hence with varying forgetting constants, this model can be seen as having more or less “memory” or, in a predictive coding scheme, as having more precise prior beliefs and associated predictions to make its perceptual decision. Thus it can “learn” if τ_t is sufficiently high ($K_t \rightarrow 1$, slow forgetting) or not if it is too low ($K_t \rightarrow 0$, fast forgetting). Based on our previous work in the auditory modality [71] we defined models for $\tau_t \in [5, 10, 20, 30, 50, 100]$.

$$f(x_i, u_i) : \begin{cases} x_{1,i+1} = u_i + x_{1,i}K_t \\ x_{2,i+1} = (1 - u_i) + x_{2,i}K_t \end{cases} \quad \text{with } K_t = e^{-1/\tau_t} \quad (6)$$

Multiple hypotheses exist in the predictive processing literature about which internal mechanism is reflected by the MMN [73]. Accordingly, we derived three BL “outputs” which differ in their observation function that predicts the mismatch response.

- **1. Novelty detection or Surprise (BL^{surp}).** A first way to describe the MMN response in the Bayesian leaning framework is as a neuronal process encoding surprise about the category of the stimulus. It corresponds to the novelty detection model about sensory inputs described by Lieder et al. [73] and can be formulated as the Shannon surprise on the current input category: $-\ln p(u_i|x_i)$, with $p(u_i|x_i)$ being the probability of observing input u at trial i given prior belief on stimuli probability x_i .

$$g(x_i, u_i, \boldsymbol{\lambda}) = \lambda_0 - \ln p(u_i|x_i)\lambda_1 \quad (7)$$

$$\text{with } p(u_i|x_i) = \begin{cases} \frac{x_{1,i}}{x_{1,i} + x_{2,i}} & \text{if } u_i = 1 \text{ (deviant)} \\ \frac{x_{2,i}}{x_{1,i} + x_{2,i}} & \text{if } u_i = 0 \text{ (standard)} \end{cases}$$

- **2. Precision-Weighted Prediction Error (BL^{pwpe}).** In this view, the brain would implement approximated Bayesian inference by doing predictive coding : hierarchical minimization of precision-weighted prediction errors [42]. In our modelling framework we could simulate this process by computing the difference between the prediction about incoming sensory data (i.e. the expected value of the prior Beta distribution) and sensory input u_i , weighted by the precision of the model (i.e. the inverse variance

of the prior Beta distribution).

$$g(x_i, u_i, \boldsymbol{\lambda}) = \lambda_0 + \frac{u_i - \mathbb{E}[B(x_{1,i}, x_{2,i})]}{\text{var}[B(x_{1,i}, x_{2,i})]} \lambda_1 \quad (8)$$

- **3. Model adjustment (BL^{madj}).** Finally Bayesian learning can be conceptualized as the amount of belief updating about model parameters. It was described as “Bayesian surprise” in Ostwald et al. and as “model adjustment” in Lieder et al. We adopted this last formulation and quantified updating as the divergence between prior and posterior beliefs about parameter μ . In practice it can be computed as the Kullback-Leibler divergence D_{KL} (or relative entropy) between posterior and prior Beta distributions over μ , leading to the following observation function :

$$g(x_i, u_i, \boldsymbol{\lambda}) = \lambda_0 + D_{KL}(\mu_{prior}, \mu_{posterior}) \lambda_1 \quad (9)$$

with $D_{KL} = \int p(\mu_{prior}) \ln \left(\frac{p(\mu_{prior})}{p(\mu_{posterior})} \right) d\mu = \int p(\mu|x_{1,i}, x_{2,i}) \cdot \ln \left(\frac{p(\mu|x_{1,i}, x_{2,i})}{p(\mu|x_{1,i+1}, x_{2,i+1})} \right) d\mu$

For Beta distributions D_{KL} is simplified and can be solved analytically:

$$D_{KL} = \log \left(\frac{\Gamma(x_{1,i} + x_{2,i})}{\Gamma(x_{1,i+1} + x_{2,i+1})} \right) + \log \left(\frac{\Gamma(x_{1,i+1})}{\Gamma(x_{1,i})} \right) + \log \left(\frac{\Gamma(x_{2,i+1})}{\Gamma(x_{2,i})} \right) \\ (x_{1,i} - x_{1,i+1})[\psi(x_{1,i}) - \psi(x_{1,i} + x_{2,i})] + (x_{2,i} - x_{2,i+1})[\psi(x_{2,i}) - \psi(x_{1,i} + x_{2,i})] \quad (10)$$

with Γ and ψ the Gamma Euler and digamma Euler functions.

Model fitting

283

We ran VBA model inversion scheme on the 27 models defined before (null, DD, SSA with seven different τ_a and three BL outputs with six different τ_t each), independently for each subject and time window of epoched EEG data. We defined 21 non overlapping 10ms time windows from 45 to 255 ms around tone onset (45 to 55ms, 55 to 65ms, etc). The data given to the VBA model inversion function was the EEG activity averaged temporally over each 10ms window, and spatially over the ROI defined before, trial by trial.

284

285

286

287

288

Priors over models’ parameters and initial hidden states need to be set before model inversion and were identical for both studies. All priors and initial hidden states were defined as multivariate normal distributions. No evolution parameters $\boldsymbol{\theta}$ were used as for the SSA and BL model families, τ_a and τ_t were fixed for each different model and not fitted.

289

290

291

292

All models had two observation parameters ϕ (except the null model which only had one): a factor ϕ_1

293

scaling the model response and an offset ϕ_0 . We set them empirically in the same manner for all models with
prior means:

- $\phi_0 = \text{mean}(\text{standards})$: the mean amplitude over all standard tones (in the given time window)
- $\phi_1 = \text{mean}(\text{deviants}) - \text{mean}(\text{standards})$: averaged deviant minus standard amplitude, hence a
measure of the MMN response (again in the time window of interest)

and a diagonal prior covariance matrix with variance 10.

The prior for the measurement noise precision σ was a Gamma distribution of shape 0.1 and rate 1, consistent
with a standard deviation (inverse precision) for amplitude measurement of 10 μV .

Initial hidden states (X_0) are necessary to initiate the inversion and depended on the models. The null
model did not have any hidden state. For all other models, a null covariance matrix was specified, encoding
absolute certainty about the initial values of hidden states. The DD model’s single hidden state was initiated
to 0, while SSA models’ initial hidden states were set to 1, reflecting the likely assumption that both standard
and deviant specific neuronal populations are not adapted before hearing any tone and their responsiveness
is therefore maximal. BL models’ initial hidden states were also conventionally set to 1, even if no tones
were ever heard, to satisfy computational constraint (the Kullback-Leibler divergence is not defined for null
hidden states).

In study 2, all parameters were reinitialized and fitted independently for the two sessions, as these were
acquired several hours apart. For the whole experiment of Study 1, and within a session of Study 2, conditions
and blocks followed one after the other, separated only by short transitions of a few seconds, which we simply
ignored in the modeling scheme. Therefore the inputs from all conditions and blocks were used and treated as
a continuous and uninterrupted sequence for the evolution function, but only the EEG data from the control
conditions were used to inform model inversion. Similarly trials marked for rejection during pre-processing, as
well as outliers defined as any trial whose EEG amplitude exceeded four times the standard deviation over all
trials, were included in the sequence of inputs to modeled—informing the evolution function dynamics—but
not in the computation of goodness-of-fit measures.

VBA inversion outputs the summary statistics (mean and covariance) of the posterior distributions of all
model variables, as well as goodness-of-fit metrics of which two will be of particular interest to us:

- the model’s variational free energy F —an approximation of model log-evidence that favors accuracy
while penalizing model complexity [84]. It was shown to outperform other information-theoretic criteria
for model comparison [85] and so will be used in this sense (see below).

- the percentage of variance explained by the model R_{obs}^2 , which will allow us to derive an absolute goodness-of-fit measure.

Inference

We carried out inference in two hierarchical steps. First we compared models at the level of families, i.e. sets of models sharing a similar structure or computational feature. In the present study the families are SSA and the three outputs of BL, and each of them aggregates models that differ only in the value of the temporal parameter τ (τ_a for SSA and τ_t for BL). This first step allowed us to select the most plausible computational mechanism for each (10ms averaged) post-stimulus time windows of the auditory evoked response. Second, we estimated the value of τ for the winning family through Bayesian model averaging (BMA). The two steps are described below in more detail.

Bayesian Model Comparison (BMC). There are two different approaches to BMC, depending on the assumption made about the structure of the underlying population. In fixed-effect BMC (ffx-BMC), the population is assumed to be *homogeneous* in the sense that all subjects' data are best described by the same unique model. Alternatively, random-effect BMC (rfx-BMC) treats the optimal generative model as a random effect that may vary across subjects, with a fixed but unknown *heterogeneous* population distribution.

Stephan et al. (2010) recommends to select the population structure (and, therefore, the BMC procedure) based on characteristics of the effect of interest, with the homogeneous assumption being “warranted when studying a basic physiological mechanism that is unlikely to vary across the subjects” [86]. The mechanisms underlying the MMN arguably falls under this definition, justifying the use of ffx-BMC. Alternatively, the choice between inference procedures may be data-driven as described in .

In fixed-effect BMC group-level inference starts out by calculating the evidence for each model given the whole sample using the product rule for independent events:

$$p(y|m) = p(y^{(1)}, y^{(2)}, \dots, y^{(N)}|m) = \prod_{s=1}^N p(y^{(s)}|m) \quad (11)$$

where $y^{(s)}$ is the observed data for the participant s and N is the number of participants. Group model evidence is often calculated in log form:

$$\log(p(y|m)) = \sum_{s=1}^N \log p(y^{(s)}|m) \quad (12)$$

Then, the *posterior model probability* (PMP) of model m_j can be derived from all models evidence using the Bayes theorem:

$$p(m_j|y) = \frac{p(y|m_j)p(m_j)}{\sum_k p(y|m_k)p(m_k)} \quad (13)$$

In the context of a uniform distribution over model space (i.e. the prior belief that all models are equally likely), which we will assume in the present study, the previous formula can be further simplified to:

$$p(m_j|y) = \frac{p(y|m_j)}{\sum_k p(y|m_k)} \quad (14)$$

The model with the highest posterior model probability may then be selected as the most plausible at the population level. Alternatively, one may focus on the Bayes Factor for a model j (compared to all other models) which can be written as the ratio of posterior odds to prior odds:

$$BF_{m_j,all} = \frac{p(m_j|y)}{1 - p(m_j|y)} \bigg/ \frac{p(m_j)}{1 - p(m_j)} = \frac{p(m_j|y)}{1 - p(m_j|y)} (K - 1) \quad (15)$$

with K the number of compared models, all with equal prior probability $p(m_k) = 1/K$.

345

346

Bayesian Model Averaging (BMA). Once a model family has been selected, the best value for the τ parameter can be found by using a BMC procedure again, this time restricted to a single family. Alternatively it is possible to take into account the model uncertainty in the estimation of τ . BMA does so by averaging all possible (modeled) values of the parameter, weighted by the posterior probability of the corresponding model:

$$\hat{\tau} = \sum_{\substack{m \in M \\ \tau \in T}} \frac{\tau \cdot p(m_\tau|y)}{|T|} \quad (16)$$

where M designates a family of model, T the set of τ values that were modeled, and $|T|$ the size of T .

347

348

Interpretation of Bayes Factors. We will adopt the conventional heuristics proposed by Kass and Raftery (1995) [87] whereby a BF higher than 100 is interpreted as decisive evidence, a BF between 10 and 100 indicates strong evidence and a BF between 3 and 10 suggests substantial evidence. BF below 3 are deemed inconclusive.

349

350

351

352

Model recovery

353

Before taking the results of model comparison at face value, it is essential to validate that the BMC procedure 354
delivers reliable results using simulations [88]. Model recovery analysis ascertains whether and under which 355
experimental conditions (design, sample size, signal-to-noise ratio) we can conclusively arbitrate between 356
different models. The overall approach involves generating synthetic data from each of the candidate models 357
and fitting all the models to determine whether we can *recover* the true one. The main purpose of the model 358
recovery analysis for the present study is to assess the separability of SSA from BL models, and the possibility 359
to disambiguate BL outputs. Therefore we restricted the simulated model space to one specific value of τ_a 360
(for SSA models) and one for τ_t (for BL models), making the analysis more tractable. Based on previous 361
modeling results obtained from an oddball experiment by some of us [71], we chose $\tau_t = 20$. The three BL 362
output functions are highly correlated at this value of τ_t (0.93 on average), more so than at most other 363
parameter values, ensuring that the results of model recovery will be conservative. Following the same logic 364
we chose the value of τ_a that maximized the correlation between SSA and BL₂₀ models, that is 10 (average 365
SSA₁₀–BL₂₀ correlation = 0.94). 366

The details of our model recovery analysis are depicted in Fig 1. For each participant, the input sequence
that was presented to her is fed to a candidate model to generate a series of predicted responses (Fig 1A),
which is then declined in multiple versions varying in amplitude of the predicted deviant minus standard
difference response, as well as in level of Gaussian noise (Fig 1B). Amplitude and noise levels were manipulated
factorially, resulting in 30 pairs chosen to span the range of SNR values observed in the electrophysiological
data during the MMN time window. Besides, 5 datasets were generated for each (amplitude, noise) pair—these
datasets shared the same statistics and differed only in the actual realization of noise. Therefore, a total of
 $150 \times K \times N$ datasets were generated, where K denotes the number of candidate models and N the number
of participants. The candidate models were then fitted on these synthetic datasets, yielding free-energy
approximations of model evidence that were used to select the best model as described before, i.e. the
model with the highest posterior probability (Fig 1C). In practice we carried out BMC on 1000 bootstrapped
samples. Each of these samples was composed of N participants' data simulated at the appropriate levels of
amplitude and noise, each randomly draw from one of the 5 realizations of noise—allowing for 5^N possible
combinations. The results from the 1000 simulations can be conveniently summarized and displayed in a
confusion matrix $\mathbf{A} = (a_{ij})$ where the element a_{ij} denotes the frequency with which each candidate model j
has been selected for data simulated under model i (Fig 1D). The confusion matrix is of size $K \times K$ but may
be summarized in fewer pieces of information (Fig 1E). Of particular interest are two performance indices

that are defined model-wise: the *true positive rate* (TPR) which is the chance we have to select the true model as the best one assuming that it is included in the model space, and the *positive predictive value* (PPV) which is the chance that a model selected as the best is indeed the true model (again, assuming that the latter is in the model space). TPR and PPV are measures of sensitivity and specificity, respectively, with TPR amounting to statistical power in a 2-class problem. Ideally if the BMC procedure worked perfectly the confusion matrix would be the $K \times K$ identity matrix and we would have TPR=PPV= 1. Thus, another useful summary statistic of a confusion matrix is its distance from the optimal, identity matrix \mathbf{I}_K —which we computed using the L_1 norm:

$$|\mathbf{A} - \mathbf{I}_K|_1 = \sum_{i=1}^K \sum_{j=1}^K |a_{ij} - \delta_{ij}| \text{ with } \delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} \quad (17)$$

Conversely, if models were not distinguishable and the model selection produced random results under the experimental design and other relevant characteristics of the data, the confusion matrix would be a matrix \mathbf{U}_K with all elements equal to $1/K$ (hereafter called the *normalized unit matrix*) and we would have TPR=PPV= $1/K$. The randomness of the confusion matrix may be summarized by its distance from the normalized unit matrix.

Confusion matrices, as well as their distance to identity and normalized unit matrices are displayed as a function of SNR (magnitude/noise pairs) in Fig 4 and interpreted in the results section. In order to get a hint of our ability to decide between different models in our real data along the timecourse of the MMN response, we selected confusion matrices matching the level of SNR in EEG data for each participant, at each post-stimulus time window (50-250ms). Time-resolved TPR and PPV metrics were then computed based on “real SNR” and compared to conventional thresholds (Fig 6).

Absolute goodness-of-fit

BMC procedures pick a model out of a set of candidates based only on their *relative* performance, i.e. their comparative ability to predict the observed data while being parsimonious. Said otherwise, model comparison can single out a model as being better at predicting the data than other examined models, but it does not follow that the selected model is good in an *absolute* sense. It might actually be missing important features of the data, suggesting that the true generative model (or a useful approximation of it) was not in the set of candidate models and remains to be discovered.

A simple and straightforward measure of *absolute* fit is the coefficient of determination R^2 , which in the